

# **Biologieunterricht planen können**

**Empirische Untersuchung von Staatsexamensentwürfen  
angehender Biologielehrkräfte**

**Dissertation**

zur Erlangung des akademischen Grades  
des Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie  
der Freien Universität Berlin

vorgelegt von

Leroy Großmann

Berlin 2023

Diese Arbeit entstand zwischen Juni 2019 und September 2023 unter der Leitung von Prof. Dr. Dirk Krüger in der Arbeitsgruppe für Didaktik der Biologie des Instituts Biologie im Fachbereich Biologie, Chemie und Pharmazie an der Freien Universität Berlin.

Erstgutachter: Prof. Dr. Dirk Krüger, Freie Universität Berlin  
Zweitgutachterin: Prof. Dr. Daniela Mahler, Freie Universität Berlin

Tag der Disputation: 13.12.2023

## Danksagung

Ich danke all denjenigen, die mich während dieses Dissertationsvorhabens begleitet und unterstützt haben. Ein besonderer Dank gilt Prof. Dr. Dirk Krüger, der diese Arbeit in der ihm eigenen Mischung aus nüchternem Ernst und heiterer Gelassenheit betreut hat und mich eigene Wege und Irrwege hat beschreiten lassen. Für die Offenheit, dieses Thema bearbeiten zu dürfen, und das Zutrauen, dass ich dies werde bewältigen können, bin ich ihm sehr dankbar. Ich danke darüber hinaus Prof. Dr. Daniela Mahler, ohne die ich im Dickicht der Professionsforschung die Orientierung zu verlieren drohte.

Den aktuellen und ehemaligen Mitarbeiterinnen und Mitarbeitern der Didaktik der Biologie an der Freien Universität Berlin danke ich nicht zuletzt dafür, dass sie mir in Momenten der Verzweiflung Mut zugesprochen und mich in Momenten des Überschwangs auf den Boden der Tatsachen zurückgeholt haben. Für die inspirierende und freudvolle Zusammenarbeit danke ich insbesondere Christine Eßmann-Stern, Kristina Fricke, Maximilian Göhner, Dr. Sarah Huch, Prof. Dr. Moritz Krell, Sophie-Luise Müller, Dr. Stefan Nessler, Dr. Bianca Reinisch und Nubia Vogt.

Darüber hinaus haben an mehreren Beiträgen dieser Dissertation einige Kolleginnen und Kollegen mitgewirkt, denen ich für ihre Unterstützung und den wertvollen und konstruktiven Austausch danken möchte. Neben Denise Bock, der ich für kritisch-konstruktives Korrekturlesen herzlich danke, danke ich Maren Koberstein-Schwarz dafür, dass sie sich mit mir in das noch weitgehend unbestellte Feld der biologiedidaktischen Unterrichtsplanungsforschung begeben hat und von der ich viel gelernt habe. Ein großer Dank gilt zudem Stefan Mulke, der an der Auswertung der Staatsexamensentwürfe beteiligt war und dessen Perspektive als Fachseminarleiter diese Arbeit zweifelsohne sehr bereichert hat.

Weiterer Dank gilt unseren (ehemaligen) studentischen Hilfskräften Rebecca Bartl, Maike Barnebeck, und insbesondere Anthony Benzig und Sophia Mihm, die mich bei der Kodierung der Daten unterstützt haben.

Zuletzt danke ich der Senatsverwaltung für Bildung, Jugend und Familie des Landes Berlin für die Kooperation und die Bereitstellung der anonymisierten Staatsexamensentwürfe sowie dem Bundesministerium für Bildung und Forschung: Das Projekt K2Teach (Know how to teach) wurde im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1802 gefördert.

## **Selbstständigkeitserklärung**

Hierdurch versichere ich, dass ich meine Dissertation selbstständig verfasst und keine anderen als die von mir angegeben Quellen und Hilfsmittel verwendet habe. Geistiges Eigentum anderer Autorinnen und Autoren wurde entsprechend gekennzeichnet. Ebenso versichere ich, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt bzw. die Dissertation in dieser oder anderer Form an keiner anderen Fakultät als Dissertation vorgelegt habe

## Inhaltsverzeichnis

<b>Abkürzungsverzeichnis .....</b>	<b>III</b>
<b>Abbildungsverzeichnis .....</b>	<b>IV</b>
<b>Tabellenverzeichnis.....</b>	<b>IV</b>
<b>Zusammenfassung .....</b>	<b>1</b>
<b>Abstract .....</b>	<b>3</b>
<b>1 Einleitung .....</b>	<b>5</b>
<b>2 Unterrichtsplanung – eine „Jedermannskompetenz“?.....</b>	<b>15</b>
2.1. Biologieunterricht planen können .....	16
2.1.1. Biologieunterricht – zwischen Generik und Fachspezifik.....	16
2.1.2. Planen – ein Modellierprozess.....	19
2.1.3. Können – zum Konstrukt Unterrichtsplanungskompetenz .....	25
2.2. Beitrag 1: Empirical Research on Science Teachers’ Lesson Planning: A Scoping Review.....	39
2.3. Zusammenfassende Diskussion.....	42
<b>3 Untersuchung von Staatsexamensentwürfen.....</b>	<b>44</b>
3.1. Zusammenhang zwischen fachdidaktischem Wissen und Unterrichtsplanung.....	45
3.1.1. Beitrag 2: Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen.....	46
3.1.2. Beitrag 3: Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar*innen für die Qualität ihrer Unterrichtsentwürfe?.....	48
3.2. Besonderheiten der Unterrichtsplanung im Fach Biologie .....	50
3.2.1. Beitrag 4: Students’ conceptions as a neglected perspective in trainee teachers’ biology lesson plans ....	52
3.2.2. Beitrag 5: Erkenntnisgewinnung (v)erklärt? Wie plant man hypothesengeleiteten Biologieunterricht? ..	53
3.3. Zusammenfassende Diskussion.....	55
<b>4 Entwicklung und Überprüfung eines Kriterienrasters .....</b>	<b>69</b>
4.1. Entwicklung.....	71
4.1.1. Qualitätskriterien von Kriterienrastern.....	71
4.1.2. Beitrag 6: Identifying performance levels of enacted pedagogical content knowledge in trainee biology teachers’ lesson plans.....	72
4.2. Überprüfung.....	74
4.2.1. Validität.....	75
4.2.2. Beitrag 7: Assessing the Quality of Science Teachers’ Lesson Plans: Evaluation and Application of a Novel Instrument.....	80
4.3. Erträge für die Lehrkräftebildung.....	82
4.3.1. Beitrag 8: Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen.....	82
4.3.2. Beitrag 9: Probleme beim Planen? Acht Vorschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie .....	83
4.4. Zusammenfassende Diskussion.....	84

<b>5</b>	<b>Vorschläge für die Unterrichtspraxis .....</b>	<b>94</b>
5.1.	Mysterys .....	97
5.1.1.	Beitrag 10: Bergmann'sche Regel.....	98
5.1.2.	Beitrag 11: Darwin und Wallace.....	99
5.1.3.	Beitrag 12: Beobachten.....	100
5.2.	Beitrag 13: Simulation von Fischschwärmen.....	101
5.3.	Beitrag 14: Historisch-evolutionäres Schlussfolgern .....	101
5.4.	Zusammenfassende Diskussion.....	102
<b>6</b>	<b>Schluss.....</b>	<b>106</b>
6.1.	Fazit.....	106
6.2.	Implikationen und Desiderate.....	107
6.2.1.	Implikationen für die Forschung zur Unterrichtsplanungskompetenz.....	107
6.2.2.	Implikationen für die Lehrkräftebildung.....	114
6.2.3.	Desiderate .....	117
	<b>Literaturverzeichnis .....</b>	<b>118</b>
	<b>Anhang .....</b>	<b>141</b>
	<b>Publikationen dieser Arbeit .....</b>	<b>191</b>
	Beitrag 1 .....	192
	Beitrag 2 .....	291
	Beitrag 3 .....	312
	Beitrag 4 .....	333
	Beitrag 5 .....	347
	Beitrag 6 .....	362
	Beitrag 7 .....	385
	Beitrag 8 .....	427
	Beitrag 9 .....	448
	Beitrag 10 .....	461
	Beitrag 11 .....	469
	Beitrag 12 .....	477
	Beitrag 13 .....	485
	Beitrag 14 .....	493
	<b>Lebenslauf .....</b>	<b>504</b>

## Abkürzungsverzeichnis

$\alpha$	Cronbach's Alpha
AITSL	Australian Institute for Teaching and School Leadership
$\chi^2$	Chi-Quadrat
CK	Content Knowledge
Cohen's $\kappa$	Cohens' Kappa
cPCK	Collective Pedagogical Content Knowledge
d	Effektstärke
ePCK	Enacted Pedagogical Content Knowledge
ePCK <sub>p</sub>	Enacted Pedagogical Content Knowledge during Planning
ePCK <sub>r</sub>	Enacted Pedagogical Content Knowledge during Reflecting
ePCK <sub>t</sub>	Enacted Pedagogical Content Knowledge during Teaching
KAs	Knowledge of Assessment Strategies
KISR	Knowledge of Instructional Strategies
KMK	Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland
KSC	Knowledge of the Science Curriculum
KSU	Knowledge of Students' Understanding in Science
M	Mittelwert
Mdn	Median
MEN	Ministère de l'Éducation Nationale (Frankreich)
N	Gesamtstichprobengröße
n	Teilstichprobengröße
NRLU	National Council for Teacher Education (Norwegen)
OTS	Orientation to Teaching Science
p	Signifikanzwert
PCK	Pedagogical Content Knowledge
PK	Pedagogical knowledge
pPCK	Persönliches Pedagogical Content Knowledge
r	Pearson-Korrelationskoeffizient
$r_{i(t-i)}$	Itemtrennschärfe
SenBJF	Senatsverwaltung für Bildung, Jugend und Familie des Landes Berlin
U	Teststatistik des Mann-Whitney-U-Tests
UPK	Unterrichtsplanungskompetenz
z	Quotient aus Differenz von Scores und Mittelwert aller Scores und Standardabweichung

## Abbildungsverzeichnis

Abbildung 1: Unterrichtsplanung als Modellierprozess.....	21
Abbildung 2: <i>Refined Consensus Model</i> (RCM) of PCK.....	343
Abbildung 3: Heuristisches Unterrichtsplanungskompetenzmodell .....	36
Abbildung 4: Summierte <i>PCK map</i> der untersuchten Staatsexamensentwürfe.....	49

## Tabellenverzeichnis

Tabelle 1: Kompetenz 1 aus dem Kompetenzbereich „Unterrichten“ .....	25
Tabelle 2: Definitionen des Begriffs „Unterrichtsplanungskompetenz“ .....	27
Tabelle 3: Kriterien für die Entwicklung von Kriterienrastern .....	72
Tabelle 4: Übersicht über das entwickelte Validitätsargument.....	81
Tabelle 5: In der Gesamtstichprobe mindestens zwei Mal zitierte PCK-Quellen.....	95
Tabelle 6: Stufenmodell des Expertiseerwerbs .....	109



### Zusammenfassung

Obwohl die Unterrichtsplanung eine Kernaufgabe von Lehrkräften darstellt und als Prädiktor für die Unterrichtsqualität und damit den Lernerfolg von Schülerinnen und Schülern gilt, ist sie von wenigen Ausnahmen abgesehen bislang kein intensiv untersuchter Forschungsgegenstand in der empirischen biologiedidaktischen Forschung. Dabei spielen insbesondere schriftliche Unterrichtsentwürfe in der deutschen Lehrkräftebildung eine wichtige Rolle und sind genuiner Teil von Prüfungsleistungen in der 1. Phase der Lehrkräftebildung (Studium) sowie in der 2. Phase der Lehrkräftebildung (Referendariat). Trotz vermehrter Kritik an einer intransparenten und inkonsistenten Beurteilungspraxis vor allem in der 2. Phase der Lehrkräftebildung liegen nach wie vor keine empirisch geprüften Kriterien vor, mit denen die Qualität schriftlicher Unterrichtsentwürfe valide eingeschätzt werden kann. Im Rahmen des vorliegenden Dissertationsvorhabens wurde daher ein Kriterienraster entwickelt, mit dem die Qualität schriftlicher Unterrichtsentwürfe angehender Lehrkräfte valide eingeschätzt werden kann. Als theoretische Basis wurde zu diesem Zwecke das *pedagogical content knowledge* (PCK) genutzt, so dass das Kriterienraster neben generischen Aspekten der Unterrichtsplanung auch naturwissenschaftsdidaktische Befunde reflektiert. Das Vorgehen umfasste drei Schritte.

Erstens wurde nach dem Verfassen einzelner Beiträge dieser Dissertation und angesichts der zunehmenden Forschungsaktivität in der Unterrichtsplanungsforschung ein Scoping Review durchgeführt, um einen Überblick über das derzeit noch diffuse Feld der empirischen Unterrichtsplanungsforschung in den Naturwissenschaftsdidaktiken zu gewinnen. Da die zum Planen von Unterricht im Allgemeinen sowie zum Verfassen eines schriftlichen Unterrichtsentwurfs im Besonderen notwendige Unterrichtsplanungskompetenz als Konstrukt bislang nur vage konzeptualisiert ist, konnte auf dem Wege des Scoping Reviews und mit den empirischen Befunden aus den Beiträgen dieser Dissertation ein heuristisches Modell entwickelt werden, das die aktuelle empirische Forschungslage zusammenfasst und aus dem sich Desiderate für die weitere Forschung ableiten lassen. Das Modell basiert auf dem *Refined Consensus Model of PCK* und erweitert es insofern, als es neben der in der empirischen Forschung verbreiteten Identifikation konzeptuellen Wissens die Bedeutung prozeduralen Wissens in der Unterrichtsplanung betont und erstmals konkrete Variablen benennt, die den Transformationsprozess von PCK zwischen den verschiedenen *realms* im Modell erklären.

Zweitens wurde auf der Basis des heuristischen Unterrichtsplanungskompetenzmodells eine qualitative Inhaltsanalyse von Staatsexamensentwürfen ( $N = 46$ ) durchgeführt. Hierbei wurde untersucht, inwiefern der Vernetzungsgrad von PCK-Facetten mit der holistisch eingeschätzten Qualität der Staatsexamensentwürfe korreliert. Die aus der PCK-Forschung abgeleitete Annahme,

dass sich Qualität in der Anzahl der Stellen niederschlägt, an denen PCK-Facetten miteinander vernetzt werden, konnte in dieser Untersuchung nicht bestätigt werden. Unter den von einem erfahrenen Fachseminarleiter holistisch bewerteten Entwürfen fanden sich sowohl als sehr gut bewertete Entwürfe mit schwach vernetztem PCK als auch als schwach bewertete Entwürfe mit vergleichsweise stark vernetztem PCK. Daraus kann geschlossen werden, dass zur Beurteilung schriftlicher Unterrichtsentwürfe weniger die Quantität an PCK-Verbindungen als eher die Qualität der Verbindungen untersucht werden müsste.

Drittens wurden die Ergebnisse der Analyse von Staatsexamensentwürfen dazu genutzt, ein Kriterienraster zu entwickeln, dessen Kriterien überwiegend die Vernetzung von mindestens zwei PCK-Facetten adressieren. Zu diesem Zwecke wurden zu Aspekten, die laut theoretischer und empirischer Unterrichtsplanungsliteratur planungsrelevant sind, Niveaustufen formuliert, die aus den zuvor analysierten Staatsexamensentwürfen entnommen werden konnten. Auf diese Weise konnte deduktiv-induktiv ein Kriterienraster entwickelt werden, das 24 Kriterien mit jeweils drei Niveaustufen umfasst. Die Niveaustufen beschreiben Qualitätsunterschiede in der Berücksichtigung der 24 Kriterien, beispielsweise dergestalt, dass ein Aspekt gar nicht berücksichtigt wird (Level 0), dass er beschrieben wird (Level 1) oder dass er darüber hinaus auch fachdidaktisch begründet wird (Level 2). Um sicherzustellen, dass das Kriterienraster geeignet ist, um die Qualität von Unterrichtsentwürfen valide zu beurteilen, wurde ein umfassendes Validitätsargument entwickelt. Neben einem Feedback durch Dozierende (1. Phase der Lehrkräftebildung) sowie Fachseminarleiterinnen und Fachseminarleiter (2. Phase der Lehrkräftebildung), die die Relevanz der Kriterien bestätigten, wurde das Kriterienraster auf weitere Unterrichtsentwürfe ( $N = 100$ ) angewendet. Die Analysen zeigen, dass das Kriterienraster stark mit den holistischen Bewertungen der Qualität dieser Unterrichtsentwürfe korreliert und dass Referendarinnen und Referendare im Staatsexamen ( $n = 64$ ) im Durchschnitt einen signifikant höheren Score erreichen als Master-Studierende nach dem Praxissemester ( $n = 36$ ). Insgesamt lässt sich schlussfolgern, dass die Scores, die mit dem Kriterienraster gewonnen werden können, valide als Maß für die Qualität schriftlicher Unterrichtsentwürfe im Fach Biologie zu interpretieren sind. Aus den empirischen Untersuchungen, mit denen das Kriterienraster entwickelt und überprüft worden ist, werden abschließend Implikationen für die empirische naturwissenschaftsdidaktische Unterrichtsplanungsforschung sowie für die Lehrkräftebildung abgeleitet.

### Abstract

Although lesson planning is a core task of teachers, it has not been a subject of intensive research in empirical research, with few exceptions. In particular, written lesson plans play an important role in German teacher education and are a genuine part of examination performance in the 1st as well as in the 2nd phase of teacher education. Despite increased criticism of an intransparent and inconsistent assessment practice, especially in the 2nd phase of teacher education, there are still no empirically tested criteria with which the quality of written lesson plans can be validly assessed. Therefore, in this dissertation project, a scoring rubric was developed with which the quality of written lesson plans of prospective biology teachers can be validly assessed. The procedure comprises three steps.

First, a scoping review was conducted to gain an overview of the currently still diffuse field of empirical lesson planning research in science education. Since lesson planning competence has so far only been vaguely conceptualized as a construct, a first heuristic lesson planning competence model could be developed by means of the scoping review. According to this model, particularly pedagogical content knowledge (PCK) has an effect on the development of lesson planning competence.

Second, based on the heuristic lesson planning competence, a qualitative content analysis of lesson plans written for the second state examination ( $N = 46$ ) was conducted. The focus lied on the extent to which the degree of interconnectedness of PCK facets correlated with the holistically assessed quality of these lesson plans. The assumption derived from PCK research that quality is reflected in the number of instances where PCK facets are interconnected was not confirmed in this study. Among the lesson plans holistically evaluated by an experienced teacher educator in the 2<sup>nd</sup> phase of teacher training, there were both high-quality lesson plans with poorly interconnected PCK and low-quality lesson plans with relatively highly interconnected PCK. It can be concluded that for the assessment of written lesson plans, less the quantity of PCK connections but rather the quality of the PCK connections should be examined.

Third, the results of the analysis of written lesson plans were used to develop a scoring rubric whose criteria predominantly address the interconnectedness of at least two PCK facets. For this purpose, levels were formulated for aspects that are relevant for planning according to theoretical and empirical literature. In doing so, it was possible to deductively-inductively develop a scoring rubric comprising 24 criteria, each with three performance level descriptions. To ensure that the scoring rubric is suitable for validly assessing the quality of lesson plans, a comprehensive validity

argument was developed. In addition to feedback from teacher educators in the 1<sup>st</sup> and the 2<sup>nd</sup> phase of teacher education who confirmed the relevance of the criteria, the scoring rubric was applied to additional lesson plans ( $N = 100$ ). The analyses show that the scores correlate strongly with the holistic quality assessment of these lesson plans and that trainee teachers in the state examination ( $n = 64$ ) achieve a significantly higher score on average than master's students after the practical semester ( $n = 36$ ). Overall, it can be concluded that the scores that can be collected with the scoring rubric can be validly interpreted as a measure of the quality of written lesson plans in biology.

Finally, implications for empirical research in biology education and teacher education are derived from the empirical studies used to develop and test the criteria grid.

„[...] und allen Plänen gegenüber begleitet mich die Frage: ‚Was soll der Unsinn‘, eine Frage, die überhaupt ganz und gar von mir Besitz zu nehmen droht.“ (Theodor Fontane)

## 1 Einleitung

Lehrkräfte gehen einer eigentümlichen Profession nach. Künzli (1996) kennzeichnet diese Eigentümlichkeit mit dem Ausdruck „pädagogische Unschärferelation“ (S. 293):

*Ich spreche von der Unschärferelation, um damit anzuzeigen, dass sich auch professionelles Lehrerhandeln in mancher Hinsicht nicht sehr scharf gegenüber alltäglichen Erziehungs- und Unterrichtsbehandlungen abgrenzen lässt. [...] Erziehen und Unterrichten sind nämlich Tätigkeiten, die im Lebensvollzug des Miteinanderlebens eingelassen sind. In der einen oder andern [sic!] Weise können wir alle nämlich immer schon erziehen und unterrichten, freilich mehr oder weniger gut. [...] Das [...] gibt dem Beruf des Lehrers etwas elementar Unspezifisches: Lehrer und Lehrerinnen können etwas, was andere auch können, sie verfügen gleichsam über eine Jedermannskompetenz. (Künzli, 1996, S. 293)*

Unlängst nahm Neuweg (2022) im Rahmen seiner kritischen Analyse des Spannungsfeldes von Theorie und Praxis in der Lehrkräftebildung auf Künzlis polemische Bemerkung zustimmend Bezug. Seine Feststellung, dass angehende Lehrkräfte keine „*tabula rasa [seien], die mangels Wissen zunächst mehr oder weniger handlungsunfähig und daher durch Beschulung und Übung erst handlungsfähig zu machen [sind]*“ (Neuweg, 2022, S. 75), gilt auch für die Unterrichtsplanung im Besonderen. Wenn man unter Planen im Allgemeinen das Imaginieren von Handlungsschritten versteht, die zu einem in der Zukunft liegenden Ziel führen sollen (Dörner, 2021, S. 235f.), dann handelt es sich beim Planen zweifelsohne um eine generische Tätigkeit des Menschen, die keineswegs ein Spezifikum von Lehrkräften darstellt. Gleichzeitig erwerben Lehrkräfte im Zuge ihrer Professionalisierung professionelle Kompetenz, so dass erfahrenere Lehrkräfte elaborierter planen können als Novizinnen und Novizen (Backfisch et al., 2020; König et al., 2022; König et al., 2021a; Livingston & Borko, 1989; Mutton et al., 2011; Vogelsang et al., 2022; Westerman, 1991). Unterrichtsplanungskompetenz ist also erlernbar, so dass der Begriff „*Jedermannskompetenz*“ hier offenbar zu kurz greift. Es stellt sich nur die Frage, inwiefern sich das Planen von Lehrkräften von jedermanns Planen unterscheidet und welches Wissen Lehrkräfte zum Zwecke der Unterrichtsplanung benötigen<sup>1</sup>.

Die Lehrkräftebildung geht heutzutage von zwei wesentlichen Prämissen aus – erstens, dass die Kompetenzen, die Lehrkräfte zum Unterrichten benötigen, grundsätzlich erlernbar sind, und zweitens, dass es einen positiven Zusammenhang zwischen der Kompetenz der Lehrkräfte, der Qualität des von ihnen durchgeführten Unterrichts und dem Lernerfolg der Schülerinnen und

---

<sup>1</sup> Die dieser Arbeit zugrundeliegende Annahme, dass die Voraussetzung für kompetentes Handeln zunächst der Erwerb von Wissen ist, würde Neuweg (2022) dem Integrationstheorem zuordnen, das er dem Differenztheorem gegenüberstellt. Letzterem zufolge wären Wissen/Theorie und Können/Praxis zwei voneinander getrennte Entitäten, die in keiner unmittelbaren Wechselwirkung stehen. Eine genauere Betrachtung erfolgt in Kapitel 6.

Schüler gibt (z. B. Baumert et al., 2010; Blömeke et al., 2015; König et al., 2021b; Rothland et al., 2018; Sadler et al., 2013). Unterrichten „*vollzieht sich im permanenten Zyklus der Planung, Durchführung und Analyse*“ (Scholl & Plöger, 2020, S. 22; siehe auch Alonzo et al., 2019; Carlson et al., 2019) und stellt die „*Kernaufgabe*“ (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2022, S. 3) von Lehrkräften dar. Terhart (2009) zufolge zeichnet sich Unterricht vor allem dadurch aus, dass Lehrkräfte in pädagogischer Absicht planmäßig vorgehen:

*[Durch das Planen] verliert Unterrichtung den Charakter des Unsystematischen und Zufälligen, den sie im Alltag hat, und wird zu ‚Unterricht‘. [...] Die Planmäßigkeit des unterrichtlichen Handelns [...] erhöht die Effektivität des Unterrichtsvorgangs im Vergleich zu den [...] alltäglichen Formen von Unterrichtung und Information. Sie schafft andererseits jedoch zugleich mit dem höheren Grad an Verbindlichkeit und Regelmäßigkeit ein gewisses Maß an Standardisierung [...]. (Terhart, 2019, S. 122f)*

Für die plausible Annahme, dass die Qualität der Unterrichtsplanung einen Prädiktor für die Qualität des Unterrichts darstellt, gibt es inzwischen – wenn auch überschaubare – empirische Evidenz: Sowohl für Unterrichtsqualität gemessen in Form des Lernzuwachses von Schülerinnen und Schülern (Darling-Hammond et al., 2013), in Form von Schülereinschätzungen (König et al., 2021b) als auch in Form von Noten im Rahmen von Staatsexamensprüfungen (Weingarten, 2019) lassen sich positive Korrelation zwischen der Qualität der Unterrichtsplanung und der -durchführung finden. Vor diesem Hintergrund ist es schlüssig, in der Förderung von Unterrichtsplanungskompetenz eine der zentralen Aufgaben der Lehrkräftebildung zu sehen (Rothland, 2021), was sich in nationalen (KMK, 2022) wie internationalen Steuerungsdokumenten der Lehrkräftebildung (z. B. Australien: Australian Institute for Teaching and School Leadership [AITSL], 2018; Frankreich: Ministère de l'Éducation Nationale [MEN], 2013; Norwegen: National Council for Teacher Education [NRLU], 2016), in der Vielzahl allein in Deutschland verfügbarer Planungsratgeber (für eine Übersicht, siehe Vogelsang & Riese, 2017) sowie in den mehr als 100 allgemeindidaktischen Planungsmodellen und -konzepten niederschlägt, die seit Mitte des 20. Jahrhunderts publiziert worden sind (Scholl, 2018). Neuweg (2022) bezeichnet die Unterrichtsplanungskompetenz gar als „*heiligen Grad*“ (S. 24) der Lehrkräftebildung.

Wenn der Unterrichtsplanung also eine solch immense Bedeutung in bildungspolitischen, ausbildungsdidaktischen und allgemeindidaktisch-theoretischen Zugängen zugeschrieben wird, ließe sich annehmen, dass sie sowohl in der allgemein- als auch in der fachdidaktischen empirischen Forschung intensiv untersucht wird. Wernke und Zierer (2017) allerdings sprechen angesichts der geringen Publikationsaktivität zur Unterrichtsplanung in der empirischen Professionsforschung von einem „*in Vergessenheit geratenen Kompetenzbereich*“. Diese pointierte Formulierung macht einerseits zutreffend auf die vergleichsweise geringe Bedeutung der Unterrichtsplanung als Forschungsgegenstand aufmerksam, die in den wenigen aktuellen Studien wiederholt beklagt wird (z. B. Koberstein-Schwarz & Meisert, 2022, S. 2; König et al., 2020a, S. 802; Schröder et al., 2020,

S. 106; Weingarten, 2019, S. 22), andererseits ist sie in dieser apodiktischen Form inzwischen nicht mehr gänzlich zutreffend. Im aktuellsten derzeit verfügbaren Literaturreview zur Unterrichtsplanung zeigen König und Rothland (2022), dass sich in den letzten zehn Jahren vor allem im deutschsprachigen Raum ein Forschungsfeld entwickelt hat, das sowohl generische Aspekte wie Kompetenzorientierung (Weingarten, 2019), didaktische Adaptivität (Baer et al., 2011; König et al., 2015), Strukturierung (Krepf & König, 2022a; Krepf & König, 2022b) oder psychoemotionale Entscheidungstendenzen (Küth et al., 2021) in den Blick nimmt, das aber auch fachspezifische Perspektiven berücksichtigt (Deutschdidaktik: König et al., 2022; Geschichtsdidaktik: Wolf, 2022; Didaktik des Sachunterrichts: Kantreiter, 2022; Physikdidaktik: Riese et al., 2022; Schröder et al., 2020). Eine systematische Übersicht über den aktuellen Forschungsstand zur Unterrichtsplanung liegt für die Naturwissenschaftsdidaktiken derzeit noch nicht vor, woraus sich ein erstes, professionstheoretisches Desiderat ergibt:

*Desiderat 1.* Fasst man – wie es im aktuellen Diskurs zur Unterrichtsplanung weitgehend üblich ist – Unterrichtsplanungskompetenz als wissensbasiert auf (z. B. Bromme & Seeger, 1979; Calderhead, 1996; König et al., 2017; Mutton et al., 2011; Scholl & Plöger, 2020; Wolf, 2022) und verortet die eigene Forschung im Modell professioneller Handlungskompetenz (Baumert & Kunter, 2006), stellt sich die Frage, „welche Bedeutung [...] welchen Teilen des Professionswissens genau zukommt und in welcher Beziehung die Wissensfacetten zueinander stehen“ (Rothland, 2021, S. 366). Insbesondere angesichts der vielfältigen Forschungsaktivitäten, die unterschiedliche Aspekte professioneller Kompetenz von Lehrkräften in Bezug auf die Unterrichtsplanung fokussieren (z. B. *beliefs*: Bahçivan, 2016; pädagogisches Wissen: König et al., 2020a; Fachwissen: Vogelsang et al., 2022; Fachdidaktisches Wissen: Weitzel & Blank, 2020), wäre eine Konsolidierung des aktuell emergierenden Forschungsfeldes (König & Rothland, 2022) angezeigt.

In Beitrag 1 werden daher in Form eines Scoping Reviews diejenigen englischsprachigen naturwissenschaftsdidaktischen Studien zusammengefasst, in denen Unterrichtsplanungskompetenz als abhängige Variable operationalisiert wird, so dass die vorhandenen Evidenzen zu den Einflussfaktoren auf Unterrichtsplanungskompetenz identifiziert und systematisch dargestellt werden können.

Diejenige Facette von Professionswissen, die im Rahmen dieser Arbeit untersucht wird, ist das fachdidaktische Wissen (*pedagogical content knowledge*, PCK). Während Fachwissen (*content knowledge*, CK) keinen Einfluss auf Unterrichtsplanungskompetenz zu haben scheint (z. B. Vogelsang et al., 2022) scheinen pädagogisches Wissen (*pedagogical knowledge*, PK; König et al., 2020a) und vor allem PCK als Prädiktoren von Unterrichtsplanungskompetenz gelten zu können (z. B. Schröder et al.,

2020). Da es sich bei der vorliegenden Dissertation nicht um eine pädagogische oder allgemeindidaktische, sondern eine fachdidaktische Arbeit handelt, wird im Folgenden nicht die Rolle des PK, sondern die des PCK in der Unterrichtsplanung in den Blick genommen. PCK ist definiert als „*the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes*“ (Gess-Newsome, 2015, S. 36). An dieser Stelle sei bereits darauf hingewiesen, dass unter PCK im Rahmen dieser Arbeit nicht bloß das Wissen verstanden wird, mit welchen Instruktionsstrategien Fachinhalte an Schülerinnen und Schüler vermittelt werden können (Grossmann, 1990; Shulman, 1986) und welche Schülervorstellungen dabei zu berücksichtigen sind (Shulman, 1986) – auch wenn sich zahlreiche aktuellere Studien auf diese beiden Facetten – *Knowledge of Instructional Strategies for Science* und *Knowledge of Students' Understanding* (z. B. Bergqvist & Chang Rundgren, 2017; Großschedl et al., 2019; Mahler et al., 2017; Schmelzing et al., 2013) beschränken. Angesichts der Komplexität von Planungsprozessen werden auch die zwei weiteren in aktuellen Modellen berücksichtigten PCK-Facetten *Knowledge of Science Curriculum* (d. h. Wissen über die Lernziele laut Lehrplan) und *Knowledge of Assessment of Science Learning* (d. h. Wissen über Feedback und die Diagnose von Lernerfolgen) Gegenstand dieser Arbeit sein, da es sich zweifelsohne um planungsrelevante Aspekte handelt und die beiden erstgenannten PCK-Facetten die für Planungsprozesse notwendigen Wissensressourcen nur unvollständig abbilden würden. Darüber hinaus wird hinsichtlich des Körnungsgrades des untersuchten PCK nicht themen- oder konzeptspezifisches PCK (z. B. Chan & Hume, 2019) adressiert, sondern „general PCK“ (Veal & MaKinster, 1999, S. 8), das zwar auch ans jeweilige Fach (hier: Biologie) gebunden ist, aber noch nicht themenspezifisch ausdifferenziert wird.

Die von Weingarten (2019) angeregte „*umfangreiche[ ] Exploration des Forschungsfeldes im Sinne einer kritischen Bestandsaufnahme der gegenwärtigen Planungspraxen von Lehrkräften*“ (S. 28) lässt sich als Auftrag an die Fachdidaktiken im Allgemeinen und im Falle der vorliegenden Arbeit an die Biologiedidaktik verstehen: Von Haas (1998) abgesehen, der erstmalig die Unterrichtsplanung von Biologielehrkräften empirisch untersucht hat, wird die Unterrichtsplanung in der Biologiedidaktik derzeit allein hinsichtlich der Planungsprozesse bei der Auswahl von Material und Medien (Koberstein-Schwarz & Meisert, 2022) sowie in Bezug auf besondere Anwendungsfelder wie den Umgang mit digitalen Medien (z. B. Kotzebue, 2022) berücksichtigt. Als Indikator für die Vernachlässigung der Unterrichtsplanung aus biologiedidaktischer Perspektive lassen sich auch die etablierten Einführungswerke in die Biologiedidaktik (z. B. Gropengießer et al., 2013; Spörhase, 2021) heranziehen. Diese nehmen in den entsprechenden Kapiteln zur Unterrichtsplanung im Wesentlichen immer noch ausschließlich auf allgemeindidaktische Modelle Bezug (Gropengießer, 2013) oder bieten zwar eine anschauliche und gewiss hilfreiche Anleitung zum Planen von



Biologieunterricht an (Meisert, 2021), berücksichtigen dabei jedoch aktuelle empirische Befunde oder neuere Theorien nicht explizit.<sup>2</sup> Da allgemeindidaktische Modelle für Lehrkräfte zumindest explizit<sup>3</sup> keine handlungsleitende Rolle beim Planen spielen (Haas, 1998), wäre eine stärker biologiedidaktisch und damit möglicherweise näher am konkreten Planungsprozess ausgerichtete Perspektive auf Unterrichtsplanung notwendig. Dies wird inzwischen in einigen Publikationen durch die Verwendung des Kompetenz-Begriffs betont (z. B. Beckmann & Ehmke, 2023; König et al., 2022; Rothland, 2021), der empirisch in Bezug auf die Unterrichtsplanung aber bislang noch unterbestimmt ist. Daraus ergibt sich ein zweites, kompetenztheoretisches Desiderat:

*Desiderat 2.* Laut jüngeren kompetenztheoretischen Verortungen des Verhältnisses von Wissen und Können lässt sich professionelle Kompetenz von Lehrkräften als Kontinuum verstehen (Blömeke et al., 2015): Ausgehend von Dispositionen der Lehrkraft in Form von Professionswissen (z. B. PCK), Einstellungen, Motivation und Selbstregulationsfähigkeiten (Baumert & Kunter, 2006) wird situationsspezifisch gehandelt. In Bezug auf die Planung umfasst dieses Handeln die Wahrnehmung und Analyse (*Perception, Interpretation*; Blömeke et al., 2015) der Lernvoraussetzungen der Schülerinnen und Schüler sowie der curricularen Standards und der Sachstruktur des Themas, vor deren Hintergrund in einem Wechselspiel aus Kreation und Legitimation (Vogelsang & Riese, 2017) Entscheidungen getroffen werden (*Decision-Making*; Blömeke et al., 2015) und somit Unterricht geplant wird. Diese situationsspezifischen Fähigkeiten in der Unterrichtsplanung sind bislang empirisch kaum berücksichtigt worden (Campbell et al., 2022; König et al., 2021a; Scholl et al., 2019).

Daher wird in Beitrag 3 der Zusammenhang zwischen diesen Fähigkeiten und der holistischen Qualitätseinschätzung von Fachseminarleiterinnen und Fachseminarleitern untersucht. In den Beiträgen 4 und 5 werden dann exemplarisch zwei biologiespezifische Aspekte untersucht: In Beitrag 4 wird analysiert, inwiefern angehende Biologielehrkräfte Schülervorstellungen, einen besonders wichtigen Aspekt konstruktivistisch-geprägten naturwissenschaftlichen Unterrichts, in ihrer Unterrichtsplanung berücksichtigen. Beitrag 5 adressiert den Kompetenzbereich Erkenntnisgewinnung (KMK, 2005; KMK, 2020) und dabei insbesondere die Fähigkeit, hypothesengeleiteten Unterricht zur Förderung wissenschaftsmethodischer Kompetenzen zu planen. Zwar sind reichhaltige Forschungsaktivitäten sowohl zum Umgang mit Schülervorstellungen (z. B. Halim & Meerah, 2002; Hammann & Asshoff, 2014; Lin, 2016; Otero

---

<sup>2</sup> Das gilt in gleicher Weise für die entsprechenden Kapitel in den Einführungswerken der Naturwissenschaftsdidaktik (Nerdel, 2017), der Physikdidaktik (Kircher 2015) und der Chemiedidaktik (Barke et al. 2018).

<sup>3</sup> Implizit hingegen scheinen allgemeindidaktische Modelle insofern eine Orientierungsfunktion zu haben, als Lehrkräfte einzelne Aussagen der Modelle verinnerlichen und dies – offenbar eher unbewusst – in ihr Planungs Handeln einfließen lassen (Causton-Theoharis et al. 2008).

& Nathan, 2008) als auch zur Förderung von Erkenntnisgewinnungskompetenzen (z. B. Capps & Crawford, 2013; Gyllenpalm & Wickman, 2011; Krell & Krüger, 2022; Rönnebeck et al., 2016) und damit einhergehende Schwierigkeiten von Lehrkräften in der unterrichtlichen Umsetzung verfügbar, allerdings nicht aus Perspektive der Unterrichtsplanung. Dies ist erstaunlich, denn empirische Befunde zu den Herausforderungen in der situationsspezifischen Unterrichtsplanung könnten aufgegriffen und zum Zwecke der passgenauen Professionalisierung angehender Lehrkräfte in der 1. und in der 2. Phase der Lehrkräftebildung genutzt werden (z. B. Karlström & Hamza, 2021; Zaragoza et al., 2023). Dies würde beispielsweise erfordern, Stichproben zu verschiedenen Zeitpunkten der Ausbildung hinsichtlich ihrer Stärken und Schwächen in der Unterrichtsplanung zu untersuchen und daran anknüpfende Fördermaßnahmen zu entwickeln. Daraus ergibt sich ein drittes, forschungsmethodisches Desiderat:

*Desiderat 3.* Obwohl der Erwerb von Unterrichtsplanungskompetenz als langjähriger berufsbiographischer Prozess aufgefasst wird (Munthe & Conway, 2017) und sich Unterrichtsplanungskompetenz im Verlauf der Ausbildung mit zunehmender Praxiserfahrung weiterentwickelt (Backfisch et al., 2020; König et al., 2022; Mutton et al., 2011; Riese et al., 2022; Vogelsang et al., 2022), werden zumeist Studierende (1. Phase der Lehrkräftebildung) und selten Referendarinnen und Referendare (2. Phase der Lehrkräftebildung) untersucht. Dies wird auch jenseits der Unterrichtsplanungsforschung als Limitation der empirischen Lehr-Lern-Forschung im Allgemeinen verzeichnet (Anderson-Park & Abs, 2020; Kunz & Uhl, 2022)<sup>4</sup>. Ausnahmen bilden derzeit vor allem die zahlreichen Arbeiten aus der Deutschdidaktik aus dem Projekt PlanvoLL-D (z. B. Fladung, 2022; König et al., 2015; Krepf & König, 2022b). Eine Prüfung, inwiefern angehende Biologielehrkräfte in der 2. Phase bereits über Unterrichtsplanungskompetenz verfügen und worin möglicherweise Schwierigkeiten bestehen, würde die von Weingarten (2019) geforderte Exploration leisten, aus der wiederum Förderpotentiale im Rahmen der Lehrkräftebildung gewonnen werden könnten. Zugleich würde es dem schwerwiegenden Einwand von Rothland (2021) Rechnung tragen, der zurecht fragt, ob in Studien zur Unterrichtsplanungskompetenz bei Studierenden nicht vielleicht oft etwas gemessen wird, was angesichts ihres Ausbildungsstandes noch gar nicht entwickelt sein kann.

---

<sup>4</sup> Dies gilt im Besonderen auch für die Naturwissenschaftsdidaktiken im deutschsprachigen Raum: Prüft man alle Artikel ( $N = 54$ ), die in der *Zeitschrift für Didaktik der Naturwissenschaften* im Zeitraum 2012-2022 veröffentlicht worden sind und in denen (angehende) Lehrkräfte untersucht wurden, fällt auf, dass Referendarinnen und Referendare nur in insgesamt drei Studien, nämlich in Mahler und Arnold (2022), in Meinhardt et al. (2018) sowie in Beitrag 3 dieser kumulativen Arbeit, Großmann und Krüger (2022c), in die Stichproben miteinbezogen worden sind.

Daher werden in den Beiträgen 2-7 Staatsexamensentwürfe im Fach Biologie untersucht, die also von Referendarinnen und Referendaren am Ende der 2. Phase der Lehrkräftebildung verfasst und im Rahmen der unterrichtspraktischen Prüfung durchgeführt und reflektiert worden sind. Auf diese Weise wird zugleich dem Problem fehlender ökologischer Validität begegnet, das Kang (2017) beschreibt:

*[R]esearch on teachers' planning was often conducted in a controlled setting to uncover cognitive decisions made by teachers. Researchers tend to rely on either teacher self-reported data or data generated from think-aloud methods [...]. There is little empirical evidence that focuses on teachers' effective planning in the natural teaching environments. (Kang, 2017, S. 56)*

Die Analyse von Staatsexamensentwürfen kann in diesem Sinne auch dazu dienen, den state-of-the-art im Bereich der Unterrichtsplanung zu dokumentieren, zumindest im Land Berlin. Angesichts der Bedeutung insbesondere schriftlicher Unterrichtsentwürfe im Rahmen der zweiten Staatsexamensprüfung (z. B. in Berlin: VSLVO, 2022/05.08.2022, §22 (4) und (5)) ergibt sich daraus ein viertes, ausbildungsdidaktisches Desiderat:

*Desiderat 4.* In Ergänzung zur deskriptiv-analytischen Aufklärung der Zusammenhänge zwischen den Variablen, die Unterrichtsplanungskompetenz beeinflussen, ist die Frage zu klären, „wie fachübergreifend eine ‚gute‘ Unterrichtsplanung, wie eine qualitative Planungsperformanz empirisch zu identifizieren und bestimmen ist, die auf die Ausprägung der ihr zugrunde liegenden Planungskompetenz schließen lässt“ (Rothland, 2021, S. 365). Die Vielzahl an (zumeist allgemeindidaktischen) Planungsratgebern scheint angehenden Lehrkräften nur in begrenztem Maße eine Hilfe beim Planen von Unterricht zu sein. Zumindest wird in den von Vogelsang und Riese (2017) untersuchten 15 Planungsratgebern vor allem das Begründen von Planungsentscheidungen (*legitimierendes Planungs Handeln*), weniger das Entwickeln einer funktionalen Prozessstruktur des Unterrichts (*kreierendes Planungs Handeln*) fokussiert. Dass in etwa der Hälfte dieser Planungsratgeber zentrale kognitive Herausforderungen beim Planen von Unterricht wie die fachliche Klärung des Inhalts, die sachlogische Strukturierung oder die Berücksichtigung der Lernvoraussetzungen durch Maßnahmen der Binnendifferenzierung (König et al., 2021a) gar nicht berücksichtigt werden, lässt sich als ein Defizit deuten, das durch ein geeigneteres Instrument gelöst werden kann:

*Up to now, there are hardly any scientifically proven criteria for the analysis and evaluation of written lesson plans. We therefore consider it an important task of empirical teacher education research to present suitable procedures for the assessment of written lesson plans. (Krepf & König, 2022b, S. 14)*

Zu diesem Zwecke wurde auf der Basis theoretischer und empirischer Befunde in Zusammenarbeit mit Biologie-Fachseminarleiterinnen und -Fachseminarleitern aus der 2. Phase der Lehrkräftebildung ein Kriterienraster entwickelt, das in den Beiträgen 6, 7 und 8 vorgestellt wird.

Zwar sind bereits einige wenige Kriterienraster zur Unterrichtsplanung publiziert, die jedoch – wenn überhaupt – allenfalls mithilfe von Experteneinschätzungen auf Validität geprüft worden sind. Umfassende Validitätsevidenzen sind für keines dieser Kriterienraster gesammelt worden. Ein solches Instrument könnte beispielsweise in der 2. Phase der Lehrkräftebildung eingesetzt werden, um im Rahmen der Staatsexamensprüfung Unterrichtsentwürfe vergleichbar bewerten zu können. Dies würde der Kritik an der deutschen Lehrkräftebildung entgegenwirken, die sich im Vergleich zur Lehrkräftebildung beispielsweise in den USA, Neuseeland und Japan durch eine „*subjective nature of the assessment of new teachers and the lack of clearly stated criteria*“ (Howe, 2006, S. 294) auszeichnet. Die Relevanz transparenter und valider Beurteilungen zeigt sich beispielsweise in Bezug auf die Frage, inwiefern derartige Beurteilungen auf der Höhe des bildungswissenschaftlichen und fachdidaktischen Professionalitätsdiskurses sind (Kruse, 2022). Erfassbar wird dies zum Beispiel konkret an den Beurteilungsinstrumenten, die von den Ausbildungsinstitutionen in der 2. Phase der Lehrkräftebildung in Deutschland genutzt werden. Strietholt und Terhart (2009) analysierten zu diesem Zweck bundesweit 201 Instrumente und kommen zu folgendem Resümee:

*Die Gruppe derjenigen Instrumente [...], bei denen nicht spezifiziert wird, wie sich die inhaltlichen Kriterien in der Beurteilung nieder schlagen [sic!], ist die mit Abstand größte: In jedem zweiten Instrument bleibt unklar, wie die Referendare letztendlich bewertet werden. Es kann vermutet werden, dass die Beurteilungskriterien unverbindlich sind und nicht systematisch zur Beurteilung herangezogen werden. (Strietholt & Terhart, 2009, S. 641)*

Da Referendarinnen und Referendare die Beurteilung auch selbst als nur mittelmäßig transparent einschätzen (Kärner et al., 2019), könnte das im Rahmen dieser Dissertation entwickelte Kriterienraster einen Beitrag zur Vereinheitlichung der Bewertungsmaßstäbe und zur Transparenz der Beurteilung von Unterrichtsentwürfen im Referendariat leisten – vorausgesetzt es erlaubt valide Interpretationen bezüglich der Planungsqualität. Daraus ergibt sich ein fünftes, testtheoretisches Desiderat:

*Desiderat 5.* Bevor das entwickelte Kriterienraster zur Messung von Unterrichtsplanungskompetenz eingesetzt werden kann, ist es erforderlich, umfassend zu prüfen, ob die mit dem Kriterienraster gewonnen Testscores tatsächlich die Qualität von Unterrichtsentwürfen widerspiegeln. Dies ist ein oft vernachlässigter Schritt in der Entwicklung von Kriterienrastern im Allgemeinen (Reddy & Andrade, 2010) und in Bezug auf die Unterrichtsplanung im Besonderen: Zwar liegen derzeit bereits einige Kriterienraster zur Analyse von Unterrichtsentwürfen vor, für die allerdings entweder keinerlei Belege für die Validität der Testscore-Interpretationen vorgelegt wurden (z. B. Enugu & Hokayem, 2017; Forbes & Davis, 2010; Kademian & Davis, 2018; Kotzebue, 2022) oder für die

allenfalls Inhaltsvalidität mittels Expertenbefragungen sichergestellt wurden (Backfisch et al., 2020; Ndiokubwayo et al., 2022; Ruys et al., 2012).

Daher werden in Beitrag 7 Validitätsevidenzen innerhalb des so genannten „*argument-based approach to validity*“ (Kane, 1992; Kane, 2013) basierend auf den *Standards for educational and psychological testing* (AERA et al., 2014) gesammelt und die Qualität des Kriterienrasters auf diese Weise evaluiert.

Die somit aus fünf Perspektiven umrissene Forschungslücke bietet reichhaltiges Potential für die empirische biologiedidaktische Forschung. Zur Relativierung des Anspruchs der vorliegenden Arbeit und zur Konkretisierung des Forschungsvorhabens sei daher Folgendes zu berücksichtigen: Innerhalb des Modells der professionellen Kompetenz (Baumert & Kunter, 2006) wird hier ausschließlich das fachdidaktische Wissen (*pedagogical content knowledge*, PCK) im Fach Biologie und sein Verhältnis zur Planungsqualität untersucht. Untersuchungsgegenstand ist nicht der Planungsprozess, sondern ausschließlich das Planungsprodukt in Form von Staatsexamensentwürfen im Fach Biologie.

Neben einem Scoping Review, das den aktuellen Stand der empirischen naturwissenschaftsdidaktischen Unterrichtsplanungsforschung zusammenfasst und Desiderate aufzeigt (Beitrag 1), wird mithilfe eines deduktiv-induktiv entwickelten Kategoriensystems (Beitrag 2) der Zusammenhang zwischen der Vernetzung fachdidaktischen Wissens und der Qualität von Staatsexamensentwürfen untersucht (Beitrag 3). Exemplarisch werden die Berücksichtigung von Schülervorstellungen (Beitrag 4) und die Strukturierung von Stunden zum Kompetenzbereich Erkenntnisgewinnung (Beitrag 5) als Beispiele situationsspezifischer Fähigkeiten innerhalb von Unterrichtsplanungskompetenz jeweils in zwei vergleichenden Fallstudien untersucht. Aus den Staatsexamensentwürfen sowie aus der theoretischen und empirischen Literatur werden darauf aufbauend anhand eines heuristischen Modells qualitative Niveaustufen formuliert und die Entwicklung des Kriterienrasters anhand von drei Beispielkriterien dargestellt (Beitrag 6). Das vollständige Kriterienraster wird anschließend auf Unterrichtsentwürfe von Staatsexamenskandidatinnen und -kandidaten sowie von Master-Studierenden angewendet und umfassend auf Validität der Testscore-Interpretationen geprüft (Beitrag 7). Das Kriterienraster wird danach den Zielgruppen innerhalb der Lehrkräftebildung sowohl der 1. als auch der 2. Phase der Lehrkräftebildung präsentiert – zunächst Dozierenden bzw. Fachseminarleitungen (Beitrag 8), anschließend Studierenden bzw. Referendarinnen und Referendaren mit praktischen Beispielen und Hilfestellungen (Beitrag 9). Jenseits des Kriterienrasters werden abschließend Unterrichtsideen vorgestellt (Beiträge 10-14), die Lehrkräfte in der Planung guten Biologieunterrichts unterstützen können.

Zunächst jedoch ist die Frage zu klären, inwiefern Unterrichtsplanungskompetenz mehr als eine

„Jedermannskompetenz“ darstellt, und auf diesem Wege die Relevanz dieses Forschungsvorhabens theoretisch zu legitimieren.

„Ja mach nur einen Plan, sei nur ein großes Licht, mach gleich noch einen zweiten Plan, geben tun sie beide nicht.“ (Bertolt Brecht)

## 2      Unterrichtsplanung – eine „Jedermannskompetenz“?

Nähert man sich der Unterrichtsplanung als empirisch zu untersuchendem Konstrukt an, so fällt auf, dass derzeit noch keine einheitlichen Begrifflichkeiten im Diskurs genutzt werden: Zum einen wird nicht immer klar zwischen *Unterrichtsvorbereitung* und *Unterrichtsplanung* unterschieden (z. B. Haas, 1998, S. 10; van der Valk & Broekman, 1999). Im Rahmen dieser Arbeit wird *Unterrichtsplanung* als übergeordnete und vor allem kognitive Fähigkeit aufgefasst, die auf die zielgerichtete Entwicklung einer Unterrichtsstunde für eine Lerngruppe abzielt. *Unterrichtsvorbereitung* hingegen wird hier als die sich an die Planung anschließende, konkrete Vorbereitung der Unterrichtsdurchführung verstanden (z. B. Kopieren der Arbeitsblätter, Herstellen der gewünschten Sitzordnung). Diese Unterscheidung ginge verloren, wenn man dem im Diskurs gängigen Vorgehen folgen und für eine synonyme Begriffsverwendung plädieren würde (z. B. Gassmann, 2013, S. 106; Seel, 2011, S. 32).

Schwerwiegender für die Forschung ist jedoch die Vielfalt an vermeintlichen Synonymen in Bezug auf die empirisch zu untersuchenden Konstrukte: So werden innerhalb des Diskurses die Begriffe *Planungsfähigkeiten* (z. B. Schröder et al., 2020), *Planungskompetenz* (z. B. König et al., 2015; Schnebel et al., 2017), *unterrichtliche Planungskompetenz* (z. B. Rothland, 2021) und *Unterrichtsplanungskompetenz* (z. B. Brodhäcker, 2014; König & Rothland, 2022) gleichermaßen genutzt, offenbar in der Annahme, dasselbe zu meinen. Auch innerhalb einzelner Publikationen werden vermeintliche Synonyme genutzt, so sprechen König und Rothland (2022) sowohl von *Planungsfähigkeiten* als auch von *Unterrichtsplanungskompetenz*, während Wolf (2022) präzise das *Planungswissen* als Teil von *Planungskompetenz* unterscheidet. Koberstein-Schwarz und Meisert (2022) wiederum nutzen mit *planning competence*, *planning skills* und *planning capacities* gleich drei verschiedene Begriffe für dasselbe Konstrukt. Rothland (2021) problematisiert den „inflationäre[n] Gebrauch des Kompetenzbegriffs auf verschiedenen Ebenen und eine fehlende Differenzierung von (Teil-)Aspekten professioneller Kompetenz, Kompetenzbereichen und Kompetenzfacetten [...] oder Kompetenzklassen, Fähigkeitsdimensionen und Fertigkeiten [...] sowie generell von Kompetenz und Performanz [...]“ (Rothland, 2021, S. 351).

Dass dies jedoch „kein Spezifikum der Forschung zur Unterrichtsplanungskompetenz ist“ (Rothland, 2021, S. 351), kann in Bezug auf fachdidaktische Forschung und den international geprägten PCK-Diskurs auch damit zusammenhängen, dass PCK zugleich als „*knowledge and skill*“ definiert wird (z. B. Gess-Newsome, 2015, S. 36), so dass die Irritation bereits im Begriff angelegt ist. Entsprechend definieren Carlson et al. (2019) im *Refined Consensus Model* (RCM) das persönliche

PCK (*personal PCK*; pPCK) von Lehrkräften als „*reservoir of knowledge and skills that teachers can draw upon during the practice of teaching*“ (S. 85) und unterscheiden somit nicht zwischen der Kognition, d. h. dem expliziten Wissen (*knowledge*), das im Rahmen der Lehrkräftebildung vermittelt wird, und der Performanz, d. h. dem Resultat, das sich nach der Anwendung von Fähigkeiten und Fertigkeiten, die Lehrkräfte im Planen, Durchführen und Reflektieren von Unterricht nutzen (*skills*), beobachten lässt.

Da sich diese Arbeit als Beitrag zur Professionsforschung versteht und diese wiederum den Begriff der *professionellen Kompetenz* (Baumert & Kunter, 2006) als Fluchtpunkt in den Blick nimmt, ist eine kritische Analyse des *Kompetenz*-Begriffs notwendig. Zum Zwecke der terminologischen und damit auch konzeptionellen Klarheit werden daher in diesem Kapitel die drei titelgebenden Begriffe dieser Arbeit – *Biologieunterricht*, *Planen* und *Können* – aus Perspektive der Unterrichtsplanungsforschung theoretisch fundiert, konzeptionell verortet und in ein auf Blömeke et al. (2015) basierendes heuristisches Unterrichtsplanungskompetenzmodell überführt. Damit ist vorweggenommen, dass das zu untersuchende Konstrukt im Rahmen dieser Arbeit *Unterrichtsplanungskompetenz* genannt wird.<sup>5</sup>

### 2.1. Biologieunterricht planen können

#### 2.1.1. Biologieunterricht – zwischen Generik und Fachspezifik

Zwischen dem akademischen Diskurs zur Unterrichtsplanung auf der einen und der Planungspraxis von Lehrkräften auf der anderen Seite besteht ein offensichtlicher Widerspruch: Während in der Unterrichtsplanungsforschung ein Primat allgemeindidaktischer Ansätze verzeichnet werden kann (Scholl, 2018), planen Lehrkräfte naturgemäß Unterricht in ihren jeweiligen Fächern. Unterrichtsplanung steht also in einem Spannungsfeld von Generik und Fachspezifik (König et al., 2020b): Einerseits sind generische Planungsentscheidungen zu treffen, die sich zwischen den verschiedenen Fächern nicht allzu sehr unterscheiden (z. B. Phasenstruktur mit Einstieg, Erarbeitung, Sicherung), und andererseits sind fachspezifische Entscheidungen zu treffen, für die in den jeweiligen Fächern spezifisches Fachwissen (*content knowledge*, CK) und fachdidaktisches Wissen (*pedagogical content knowledge*, PCK) benötigt werden (z. B. kognitive Aktivierung zu einem bestimmten Thema; Gestaltung von Lernaufgaben). Während Baumert und

---

<sup>5</sup> Es sei bereits an dieser Stelle darauf hingewiesen, dass auch in den dieser kumulativen Arbeit zugrunde liegenden Beiträgen die das zu untersuchende Konstrukt beschreibenden Begriffe inkonsistent verwendet wurden. Die Inkonsistenzen werden vor dem Hintergrund der in diesem Kapitel vorgenommenen Definitionen in den zusammenfassenden Diskussionen der einzelnen Beiträge jeweils transparent gemacht und kritisch diskutiert.



Kunter (2006) die Unterrichtsplanung in ihrem wegweisenden Stichwort-Artikel zur professionellen Kompetenz von Lehrkräften einzig dem generischen pädagogischen Wissen (*pedagogical knowledge*, PK) zuordnen und damit explizit “[a]llgemeindidaktisches Konzeptions- und Planungswissen”, “[m]etatheoretische Modelle der Unterrichtsplanung”, “[f]achübergreifende Prinzipien der Unterrichtsplanung” sowie “Unterrichtsmethoden im weiten Sinne” (S. 485) meinen – ohne jedoch auszuführen, was das jeweils genau bedeutet – werden inzwischen auch fachdidaktische Perspektiven auf die Unterrichtsplanung entwickelt (z. B. Fladung, 2022; Riese et al., 2022; Wolf, 2022). König et al. (2020b) betonen die Bedeutung fachspezifischer Planungsentscheidungen:

*Only subject-specific planning decisions may enable in-depth elaborations, as student learning of a school subject takes place in a particular domain [...]. Pre-service teachers are required to reflect on subject-specific issues if they want to make planning decisions on selecting adequate learning tasks and assigning them to students.* (König et al., 2020b, S. 135)

Sie zeigen am Beispiel des Faches Deutsch, dass es Referendarinnen und Referendaren signifikant schwerer fällt, fachspezifische Planungsentscheidungen in Unterrichtsentwürfen darzulegen als generische. Dabei verweisen sie in ihrer Argumentation auf Shulman (1986), der mit der Idee, PCK als charakteristische Wissensform von Lehrkräften zu postulieren, die Bedeutung der Fachspezifität für die Gestaltung von Lehr-Lern-Prozessen erstmals betonte und die empirische Professionsforschung damit bis heute prägt. Unter dem Einfluss der in den 1970er und frühen 1980er Jahre vorwiegend kognitionspsychologisch geprägten Forschung zur Unterrichtsplanung<sup>6</sup> (z. B. Morine-Dershimer, 1979; Peterson et al., 1978; Yinger, 1980; Zahorik, 1975) fasst er den damaligen Forschungsstand folgendermaßen zusammen:

*In reading the literature of research on teaching, it is clear that central questions are unasked. The emphasis is on how teachers manage their classrooms, organize activities, allocate time and turns, structure assignments, ascribe praise and blame, formulate levels of their questions, plan lessons and judge general student understanding. What we miss are questions about the content of the lessons taught, the questions asked, and the explanations offered.* (Shulman, 1986, S. 8)

Dies ist bemerkenswert, denn Shulman scheint die Unterrichtsplanung als eine von vielen pädagogischen Tätigkeiten von Lehrkräften zu betrachten. Dass er zwar das PCK, nicht aber die Unterrichtsplanung als fachspezifisch aufzufassen scheint und diese mit seinem Konzept von PCK offenbar nicht in Verbindung bringt, ist angesichts der Gebundenheit von Unterrichtsplanung an das jeweilige Fach verblüffend und spiegelt sich bis auf wenige Ausnahmen (Alonzo et al., 2019; Koberstein-Schwarz & Meisert, 2022; Stender & Brückmann, 2020; Weitzel & Blank, 2020) in der naturwissenschaftsdidaktischen Forschung bis heute wider. Da „das Fach den eigentlichen Handlungsrahmen von Lehrkräften darstellt [und] [d]as Fach bis in Einzelheiten hinein die Textur des Unterrichts

<sup>6</sup> Da ausführliche Synopsen dieser noch nicht fachdidaktisch ausgerichteten Forschungsarbeiten bereits in anderen Dissertationen geliefert wurden (z. B. Gassmann, 2013; Weingarten, 2019) und für diese Arbeit nicht zielführend sind, wird an dieser Stelle auf eine vertiefende Betrachtung verzichtet.

[bestimmt]“ (Baumert & Kunter, 2006, S. 492), stellen das CK und das PCK unabdingbare Ressourcen für die Planung von Unterricht dar. Scholl et al. (2019) konstatieren:

*Weder die allgemein- und fachdidaktischen Planungsansätze selbst noch deren Vergleiche sind darauf ausgerichtet, über die Spezifikation der Planungsbereiche hinaus genauere Aussagen zu den jeweils angenommenen Planungsoperationen oder dem für die Planung notwendigen fachlichen, fachdidaktischen und allgemeindidaktischen Wissen herauszustellen und damit das Konstrukt der Planungskompetenz näher zu bestimmen. [Es] lässt sich deshalb feststellen, dass trotz aller bisherigen Bemühungen das für die Planung notwendige Wissen und die darin gründende situationspezifische Kompetenz der Unterrichtsplanung sowohl konzeptionell als auch empirisch nahezu ungeklärt sind [...] und das Verhältnis der vorliegenden Planungsansätze unter dem Aspekt der Modellierung dieser Kompetenz immer noch unbestimmt ist. (Scholl et al., 2019, S. 76)*

Der Vergleich zweier ausgewählter allgemeindidaktischer Planungsmodelle (Klafki, 2007; Heimann, 1962) mit fachspezifischen Planungsmodellen (in den Fächern Mathematik, Sport, Geographie) führt Scholl et al. (2019) zu der Schlussfolgerung, dass Planungsentscheidungen ein spezifisches Planungswissen erfordern, das stets sowohl generische als auch fachspezifische Teile umfasst. Aus diesem Grund werden sowohl in der qualitativen Inhaltsanalyse von Staatsexamensentwürfen (Beiträge 2 und 3) als auch im Kriterienraster (Beiträge 6-9) keineswegs ausschließlich biologiespezifische Kategorien genutzt, sondern überwiegend generische, die auf die Biologie angewendet werden. Damit wird dem Vorschlag von Şen (2023) gefolgt, auf fachspezifische Aspekte bezogene Entscheidungen als PCK zu kodieren, und nur solche, die ohne jedweden Fachbezug allgemein auf das Unterrichten bezogen sind, als PK zu kodieren.

Die Dichotomie zwischen Generik und Fachspezifik mittels einer Synthese aufzulösen, erscheint zudem insofern tragfähig, als auch im aktuellen Diskurs um Unterrichtsqualität inzwischen nicht mehr dichotom gedacht wird, sondern Unterrichtsqualitätsmerkmale entlang eines Kontinuums aufgefasst werden – sowohl in der Allgemeindidaktik (Begrich et al., 2023; Praetorius et al., 2020) als auch in der Biologiedidaktik (Neuhaus, 2021). Dort schlägt Neuhaus (2021) vor, als eher biologiespezifische Unterrichtsqualitätsmerkmale den Einsatz realer Objekte, den reflektierten Umgang mit Modellen, den Umgang mit Schülervorstellungen und die Fachsprache dem PCK zuzuordnen, und die fachliche Richtigkeit, die für Schülerinnen und Schüler angemessene Komplexität der Aufgaben, das Lernen im Kontext und die inhaltliche Strukturierung der Inhalte dem CK zuzuordnen. Zweifelsohne sind dies wichtige Merkmale, die in der Planung von Biologieunterricht zu beachten sind. Allerdings wird im Folgenden zu zeigen sein, dass Unterrichtsplanung im Fach Biologie ohne generische Planungsoperationen nicht auskommt.

Es ist nun zu klären, wie Unterrichtsplanungskompetenz zu operationalisieren ist. Zunächst werden daher relevante Befunde zur *Unterrichtsplanung* dargelegt, ehe anschließend das zu untersuchende Konstrukt *Unterrichtsplanungskompetenz* präzise definiert wird.

### 2.1.2. Planen – ein Modellierprozess

Dieser Arbeit liegt das Planungsverständnis von Dörner (2021) zugrunde, auf das sich bereits Weingarten (2019) bezieht:

*Beim Planen tut man nicht, man überlegt, was man tun könnte. Planen besteht wohl im Kern darin, dass man sich die Konsequenzen bestimmter Aktionen vor Augen führt und prüft, ob sie eine Annäherung an das gewünschte Ziel erbringen. Wenn einzelne Aktionen dies nicht tun, bildet man wohl auch längere Aktionsketten. (Dörner, 2021, S. 235f.)*

Demnach besteht Planen darin, Handlungsschritte zu imaginieren, die zu einem in der Zukunft liegenden Ziel führen sollen. Auf die Planung von Unterricht angewendet bedeutet dies, dass eine Prozessstruktur des Unterrichts erdacht werden muss, mit der die Schülerinnen und Schüler zu einem Ziel – dem angestrebten Kompetenzzuwachs – geführt werden können. In diesem Sinne lässt sich Planen als Problemlöseprozess auffassen (Anderson & Kratwohl, 2014, S. 87), denn Problemlösen beschreibt den Vorgang, bei dem ausgehend von einem defizitäreren Ausgangszustand mithilfe geeigneter Operationen ein angestrebter Endzustand hergestellt wird (Funke, 2003). Welche Operationen sind es aber, die im Zuge der Unterrichtsplanung durchgeführt werden müssen und deren Bewältigung offenbar mehr als „Jedermannskompetenz“ verlangt? Statt den theoretischen und empirischen Diskurs zu rekapitulieren, der in anderen aktuellen Dissertationen zur Unterrichtsplanung (Gassmann, 2013; Kantreiter, 2022; Stender, 2014; Weingarten, 2019) bereits ausführlich dargestellt wurde, wird an dieser Stelle auf Basis ausgewählter Literatur ein Modell vorgeschlagen, das das Planen von Unterricht als Modellierprozess auffasst (Abb. 1). Da Planen grundsätzlich nichts anderes darstellt als die Modellierung von Operationen (hier: einer Prozessstruktur von Unterricht) zur Lösung eines Problems (hier: Kompetenzförderung), erscheint es plausibel, die Unterrichtsplanung als Modellierprozess zu betrachten und die dabei notwendigen gedanklichen Vernetzungen, die Gegenstand der empirischen Beiträge dieser Arbeit sein werden, zu visualisieren. Dabei wird die Einsicht, Unterrichtsqualitätsmerkmale entlang eines Kontinuums aufzufassen (Neuhaus, 2021), insofern berücksichtigt, als das Modell auf den ersten Blick vorwiegend generische Aspekte beschreibt, die dann in den unterschiedlichen Fächern zu unterschiedlichen Nuancierungen führen können (v. a. hinsichtlich der zu fördernden Kompetenzen und der Lerngegenstände). Das Modell lässt sich anhand von fünf für diese Arbeit relevanten Aspekten charakterisieren:

Erstens wird in Anlehnung an Krüger und Upmeyer zu Belzen (2021) zunächst zwischen der Erfahrungswelt (hier: *Unterrichtswelt*, linke Hälfte) und der Modellwelt (hier: *Planungswelt*, rechte Hälfte) unterschieden. Da Unterrichtsplanung nicht isoliert stattfindet und keinem Selbstzweck dient, sondern in einen ständigen Zyklus aus Planung, Durchführung und Reflexion von Unterricht

eingebunden ist (Alonzo et al., 2019; Carlson et al., 2019; Scholl & Plöger, 2020), wird zwischen dem vorunterrichtlichen *Planungsprozess* (obere Hälfte), d. h. der Planung selbst, und der sich daran anschließenden *Planungsdurchführung* (untere Hälfte) unterschieden (John, 2006), bei der der mentale oder schriftlich niedergeschriebene Unterrichtsentwurf als Planungsprodukt im Unterricht durchgeführt und reflektiert wird. Auf diese Weise wird die etablierte Unterscheidung in den Planungsprozess (obere Hälfte) und das Planungsprodukt (gedankliches Modell/Unterrichtsentwurf) berücksichtigt.

Zweitens trägt es der Bedeutung vernetzten Denkens Rechnung, das zweifelsohne eines der wichtigsten Kennzeichen von Unterrichtsplanung darstellt (Scholl et al., 2022; Shavelson & Stern, 1981; Stender & Brückmann, 2020; Vogelsang & Riese, 2017), das aber zugleich auch als besondere Schwierigkeit insbesondere von Planungsanfängerinnen und -anfängern gilt (Bullough, 1987; Gassmann, 2013; Schrader & Schöb, 2016; Westerman, 1991). Unterrichtsplanung ist kein linearer Prozess, auch wenn dies von einigen Planungsmodellen suggeriert wird (z. B. Klafki, 1958). Dass dies keineswegs eine neue Erkenntnis ist, zeigen bereits die Ausführungen von Schulz (1975), der sich gegen Wolfgang Klafkis „*Primat der Didaktik [...] im Verhältnis zur Methodik*“ (Klafki, 1963, S. 23) wandte und die Bedeutung der Vernetzung verschiedener Planungsaspekte betonte:

*Die Auffassung der Didaktik als eine Strukturtheorie des Lehrens und Unterrichtens führt zum Prinzip der Interdependenz der Unterricht konstituierenden Momente. Die anthropologischen Bedingungen didaktischen Denkens in der Gegenwart fordern Variabilität des Unterrichts, die Möglichkeit der Mitsteuerung des Unterrichtsprozesses durch die Lernenden. Die empirische Selbstkontrolle didaktischen Denkens drückt sich im Planungsprinzip größtmöglicher Kontrollierbarkeit des Unterrichtserfolges aus. (Schulz, 1975, S. 44 f.)*

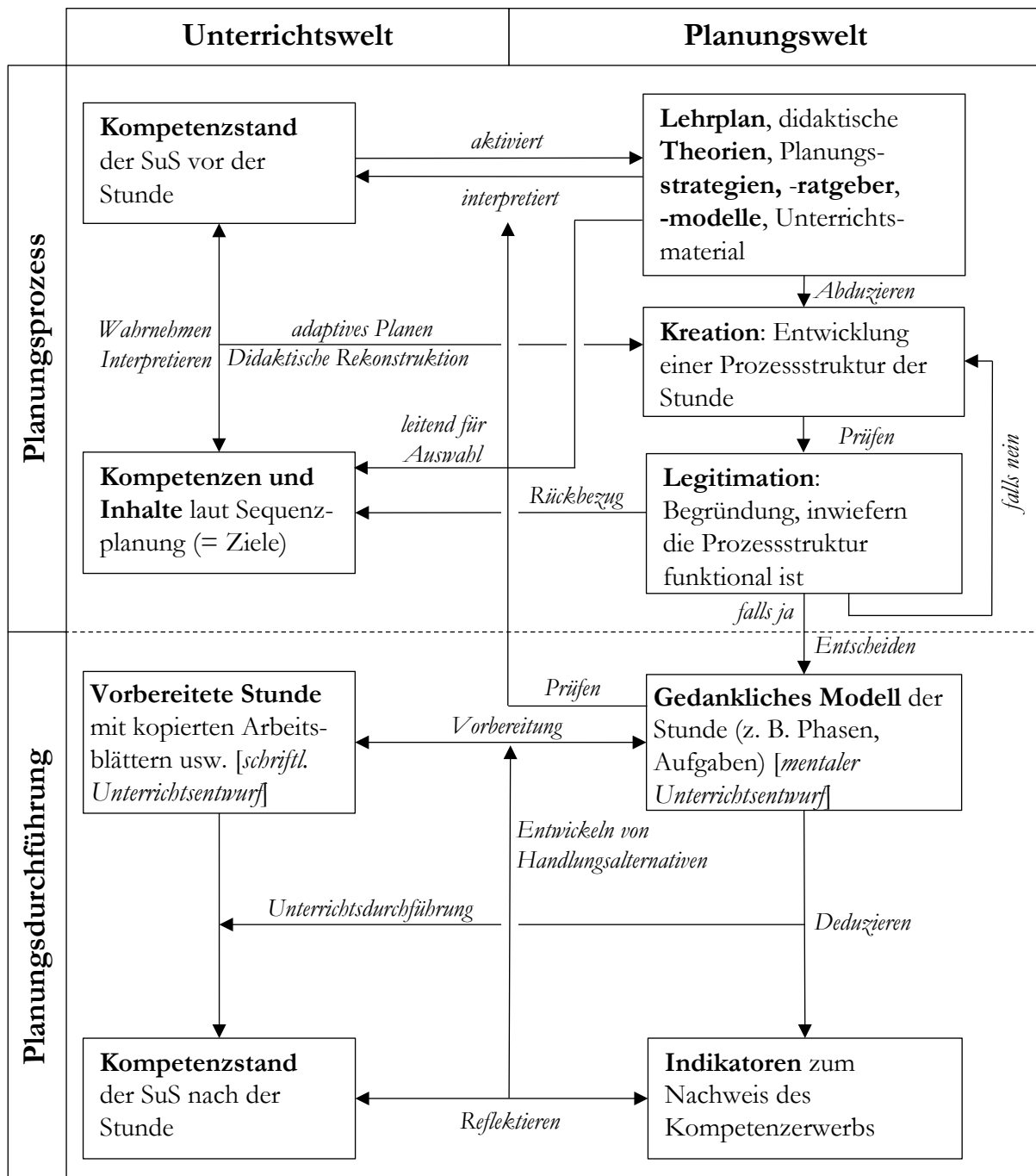
Die Arbeiten von Paul Heimann, Gunter Otto und Wolfgang Schulz, die gegen Klafkis bildungstheoretische Didaktik opponierten und in ihren lehr-lern-theoretischen Modellen (Berliner Modell, Hamburger Modell) die Vernetzung von Unterrichtszielen, Inhalten, Methoden und Medien betonen, lösen sich somit vom Primat der Didaktik (d. h. der Bildungsziele) gegenüber der Methodik.<sup>7</sup> Allerdings wurde wiederholt eingewendet, dass diese Modelle empirisch nicht fundiert seien. Scholl et al. (2022) weisen darauf hin, dass die interdependente Entscheidungsfindung vor allem im Zuge des Decision-Making-Paradigmas der 1970er und 1980er Jahre größere Aufmerksamkeit erlangte (z. B. Shavelson & Stern, 1981), dass sie aber bislang weder in einem eigenen Modell charakterisiert noch in einem Test operationalisiert wurde. Das von ihnen vorgeschlagene Rahmenmodell interdependenten Entscheidens und das darauf basierende Instrument (Scholl et al., 2022) bieten die Möglichkeit, den Zusammenhang zwischen der Fähigkeit,

---

<sup>7</sup> Für eine ausführliche Darstellung der Entwicklung allgemeindidaktischer Planungsmodelle, siehe z. B. Scholl (2018) und Weingarten (2019).

Abbildung 1

Unterrichtsplanning als Modellierprozess



*Anmerkung.* In der Planung einer Unterrichtsstunde (Planungsprozess, obere Hälfte) wird ausgehend vom derzeitigen Kompetenzstand der Schülerinnen und Schüler (Unterrichtswelt, linke Seite) in einem iterativen Prozess aus Kreation und Legitimation gedanklich eine Unterrichtsstunde modelliert (Planungswelt, rechte Seite; vgl. Krüger und Upmeyer zu Belzen, 2021, S. 132). Dieser mentale Unterrichtsentwurf als Modell wird für die Durchführung der Stunde konkretisiert (hier: in Form eines schriftlichen Unterrichtsentwurfs), der Unterricht wird also vorbereitet (Planungsdurchführung, untere Hälfte). Aus dem Modell als gedankliches Planungsprodukt werden Indikatoren abgeleitet, die im Planungsprozess im Idealfall bereits mitgedacht worden sind und anhand derer nachgewiesen werden kann, ob die Schülerinnen und Schüler die angestrebte Kompetenzentwicklung in der durchgeführten Stunde erfahren haben. In der Reflexion werden Handlungsalternativen entwickelt, das gedankliche Modell erneut geprüft und in einem neuen Planungsprozess das gedankliche Modell der Stunde verändert – sofern die angestrebte Kompetenzentwicklung nicht bzw. nur teilweise erreicht wurde. SuS = Schülerinnen und Schüler.

Interdependenzen herzustellen, und der Qualität von Planungsprozessen empirisch zu untersuchen. Die Annahme eines solchen Zusammenhangs erscheint auch insofern plausibel, als es in der PCK-Forschung vermehrt Hinweise darauf gibt, dass die Vernetzung des PCK positiv mit Unterrichtsqualität im Allgemeinen (z. B. Park et al., 2011) und ihren Teilaspekten wie der kognitiven Aktivierung (z. B. Keller et al., 2017) oder der individuellen Unterstützung von Lernenden (z. B. Kunter et al., 2013) korreliert. Zum anderen gibt es Hinweise auf einen Zusammenhang zwischen dem Umfang und der Qualität des PCK von Lehrkräften und den Leistungen ihrer Schülerinnen und Schüler (Förtsch et al., 2016; Jin et al., 2015; Mahler et al., 2017).

Drittens – mit heutigem lerntheoretischem Vokabular gesprochen – nehmen Otto, Heimann und Schulz bereits eine konstruktivistische Perspektivumkehr vorweg: Neben den ohnehin miteinander zu vernetzenden Aspekten wie dem Lernziel, dem zu unterrichtenden Fachinhalt, den Medien und Methoden rücken die Schülerinnen und Schüler in den Mittelpunkt des Interesses, die ihren Lernprozess eigenaktiv mitgestalten können und die Lehrkraft von der aktiven in eine reaktive Position verweisen. In diesem Zeichen steht das Prinzip didaktischer Adaptivität (König et al., 2015): Adaptives Planen bedeutet, die kognitive und motivationale Heterogenität innerhalb der eigenen Lerngruppe in der Unterrichtswelt zu diagnostizieren (*Wahrnehmen* und *Interpretieren*, Abb. 1) und für einzelne Schülerinnen und Schüler differenzierende Aufgabenstellungen anzubieten, die möglichst allen Schülerinnen und Schülern ermöglichen, ihren jeweiligen nächsten Lernschritt gehen zu können. Bereits in diesem ersten Teil des Modellierprozesses müssen Lehrkräfte vier zentrale *cognitive demands* (König et al., 2021a) bewältigen, nämlich die Einbettung der Unterrichtsstunde in eine Unterrichtssequenz und damit ein langfristiges Konzept zur Kompetenzförderung (*unit contextualization* → Abb. 1: *Sequenzplanung*), die Sachstrukturanalyse bzw. fachliche Klärung einschließlich einer didaktischen Reduktion (*content transformation* → Abb. 1: *Didaktische Rekonstruktion*) sowie die passgenaue Ausrichtung des Lehr-Lern-Prozesses an den aktuellen Kompetenzstand der Schülerinnen und Schüler, der zunächst zu diagnostizieren ist (*adaptation to students' needs* → Abb. 1: *adaptives Planen*). Diese Analyse wiederum sollte sich auf die Kompetenz beschränken, die in der jeweiligen Stunde gefördert werden soll und die auf einem Standard aus dem Lehrplan basieren sollte (*clarity of learning objectives* → Abb. 1: *Kompetenzen und Inhalte; Lehrplan*).

Viertens liegt dem Modellierprozess in der Planungswelt (rechte Hälfte) die von Vogelsang und Riese (2017) vorgeschlagene Heuristik zugrunde, der zufolge Unterrichtsplanung einen zyklischen Prozess von Kreation und Legitimation umfasst. Ideen zur Gestaltung der Unterrichtsstunde werden entweder direkt aus dem Lehrplan oder auf dem Wege der *Aktivierung* und *Interpretation* des aktuellen Kompetenzstandes der Schülerinnen und Schüler induktiv entwickelt. Der aktuelle

Kompetenzstand der Schülerinnen und Schüler wird dabei vor dem Hintergrund von Theorie- oder Erfahrungswissen analysiert. Mit dessen Hilfe werden Ideen entwickelt (*Abduktion*), wie der nächste Schritt der Kompetenzförderung gestaltet sein müsste. Dabei werden in einem rekursiven, iterativen Prozess (John, 2006; Karlström & Hamza, 2021; Yinger, 1980) solange Ideen entwickelt und wieder verworfen, bis die Lehrkraft die eigenen Planungsentscheidungen als funktional bewertet. Eine Planungsentscheidung ist dann funktional, wenn sie dazu beiträgt, dass die Schülerinnen und Schüler die angestrebte Kompetenzentwicklung erfahren können (Großmann et al., 2022). Insbesondere die Gestaltung von Aufgaben und die Entwicklung einer zielführenden Stundenstruktur stellen zwei weitere kognitive Herausforderungen dar (König et al., 2021a). Sobald die Passung der einzelnen Elemente (z. B. Methoden, Aufgaben) zum angestrebten Ziel gegeben ist, wird der Kurations- und Legitimationsprozess beendet (*Entscheiden*) und ein vorläufig vollendetes, gedankliches Modell (*mentaler Unterrichtsentwurf*) der geplanten Stunde für die Planungsdurchführung (untere Hälfte) genutzt.

Fünftens stärkt das Modell eine Perspektive, die vor allem in der angloamerikanischen Didaktik lange Tradition hat (z. B. Wiggins & McTighe, 2005), die im deutschsprachigen Raum aber noch wenig verbreitet ist – wenngleich sie bereits in Klafkis (1996) breit rezipierten *Perspektivenschema* oder in Schulz' (1980) *Hamburger Modell* angelegt ist: Analog zur Ableitung von Hypothesen aus einem Modell (Krüger & Upmeier zu Belzen, 2021) werden aus dem Planungsprodukt Indikatoren abgeleitet (*Deduktion*), mit denen am Ende der Stunde nachgewiesen werden kann, inwiefern die Schülerinnen und Schüler das angestrebte Lernziel erreicht haben. Wenn also das kreierte und legitimierte Modell, d. h. die Planung der Unterrichtsstunde, gelungen ist, müsste sich der angestrebte Kompetenzzuwachs anhand der Indikatoren nachweisen lassen.<sup>8</sup> Insofern knüpft das Modell zur Unterrichtsplanung als Modellierprozess (Abb. 1) an die Forschung zur Reflexionskompetenz an, da es die Unterrichtsplanung nicht als in sich geschlossenen, distinkten Prozess betrachtet, sondern die Planung als Teil eines ständigen Zyklus sieht, in dem die Durchführung des Unterrichts („*reflection-in-action*“; Schön, 1983, S. 49) reflektiert wird („*reflection on action*“; Schön, 1983, S. 276) und daraus Rückschlüsse für die Planung der darauffolgenden Stunde gewonnen werden können („*reflecion-for-action*“; Körkkö et al., 2016, S. 200).

Welcher Zweck wird nun im Rahmen der vorliegenden Arbeit mit dem Modell zur Unterrichtsplanung als Modellierprozess (Abb. 1) verfolgt? Zum einen werden in dem Modell wesentliche theoretische Annahmen und empirische Befunde zur Unterrichtsplanung in Beziehung

---

<sup>8</sup> Natürlich muss es nicht allein an der Planung liegen, falls der angestrebte Kompetenzzuwachs nicht erreicht werden kann. Es kommen zahlreiche Faktoren infrage, die in der Durchführung des Unterrichts die Unterrichtsqualität beeinflussen können. Insgesamt ist der Zusammenhang zwischen Unterrichtsplanung und -durchführung bislang empirisch noch kaum untersucht (Rothland, 2021).

zueinander gesetzt, so dass die Planung von Unterricht als Prozess veranschaulicht werden kann. Zum anderen lässt sich daraus eine Limitation für die im Folgenden zu berichtenden Beiträge ableiten: Dem Modell zur Unterrichtsplanung als Modellierprozess (Abb. 1) liegt die etablierte Unterscheidung in Planungsprozesse und Planungsprodukte zugrunde (John, 2006). Im Rahmen dieser Arbeit wurden als Planungsprodukte keine mentalen Modelle von Unterrichtsstunden untersucht, sondern ausschließlich schriftliche Langentwürfe (*schriftlicher Unterrichtsentwurf*, Abb. 1) analysiert, die im Rahmen der Staatsexamensprüfung verfasst worden sind. Wie einleitend bemerkt spielen solche schriftlichen Unterrichtsentwürfe eine wichtige Rolle im Rahmen der deutschen Lehrkräftebildung (VSLVO, 2022/05.08.2022), allerdings liegt noch kein Instrument vor, mit dem die Qualität von Entwürfen valide eingeschätzt werden kann (Krepf & König, 2022b). Es ließe sich zurecht einwenden, dass Unterrichtsentwürfe nicht mehr als ein Artefakt darstellen (Seel, 2011), das nur Auskünfte über die letztlich niedergeschriebenen Planungsgedanken zulässt, die Abwägungsprozesse im Kreieren und Legitimieren, die angewandten Planungsstrategien und die dabei genutzten Wissensressourcen aber außer Acht lässt. Wohl auch deshalb gibt es innerhalb der PCK-Forschung keine Studien, die ausschließlich Unterrichtsentwürfe als Datenmaterial nutzen (Chan & Hume, 2019). Da das Ziel dieser Arbeit jedoch vor allem darin besteht, empirisch geprüfte Qualitätskriterien (Krepf & König, 2022b) zu entwickeln, erscheint eine Fokussierung auf schriftliche Unterrichtsentwürfe insofern einen geeigneten methodischen Weg darzustellen, als somit nicht in einem kontrollierten und folglich unauthentischen Design, sondern ökologisch valide (Kang, 2017) ein Beitrag zur Unterrichtsplanungsforschung geleistet werden könnte.

Es ist anzunehmen, dass das Modell für nahezu sämtliche Kontexte, in denen geplant wird, Gültigkeit besitzt, wenn man von der schulischen Unterrichtssituation abstrahieren würde. Bislang wurde gezeigt, dass die Planung von Unterricht eine komplexe Aufgabe darstellt, die über alltägliches Planen deutlich hinausgeht und die damit keineswegs als „Jedermannskompetenz“ gelten kann: Die zielgerichtete Vernetzung verschiedenster allgemein- und fachdidaktischer Wissensbestände ist eine herausfordernde Aufgabe, auf die angehende Lehrkräfte im Zuge ihrer Ausbildung vorbereitet werden sollen.

Um es der empirischen Forschung zugänglich zu machen, muss das Konstrukt Unterrichtsplanungskompetenz abschließend präzise definiert werden. Daher wird das soeben dargelegte Modell nun im dritten und letzten Abschnitt dieses Theorieteils in ein auf Blömeke et al. (2015) basierendes Kompetenzmodell überführt. Zu diesem Zwecke ist zu klären, in welchem Verhältnis die in der Forschung verwendeten Begriffe *Wissen*, *Fähigkeiten* und *Können* zum *Kompetenz*-Begriff und damit zum in den Einzelbeiträgen dieser kumulativen Arbeit untersuchten Konstrukt *Unterrichtsplanungskompetenz* stehen.



**2.1.3. Können – Zum Konstrukt *Unterrichtsplanungskompetenz***

Die in Abbildung 1 dargestellten Prozesse zeigen, dass Unterrichtsplanung eine komplexe Herausforderung darstellt. Aspekte wie die didaktische Adaptivität und die Herstellung von Interdependenzen werden in den *Standards für die Lehrerbildung* (KMK, 2022, S. 7) explizit als Ziele des theoretischen (1. Phase der Lehrkräftebildung, links) und des praktischen Ausbildungsabschnitts (2. Phase der Lehrkräftebildung, rechts) genannt (Tab. 1).

**Tabelle 1**

*Kompetenz 1 aus dem Kompetenzbereich „Unterrichten“ (KMK, 2022, S. 7)*

Lehrkräfte planen Unterricht unter Berücksichtigung unterschiedlicher Lernvoraussetzungen und Entwicklungsprozesse fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.	
<i>Standards für die theoretischen Ausbildungsabschnitte</i> (1. Phase der Lehrkräftebildung)	<i>Standards für die praktischen Ausbildungsabschnitte</i> (2. Phase der Lehrkräftebildung)
Die Absolventinnen und Absolventen ...	Die Absolventinnen und Absolventen ...
... kennen die einschlägigen Erziehungs- und Bildungstheorien, verstehen bildungs- und erziehungstheoretische Ziele sowie die daraus abzuleitenden Standards und reflektieren diese kritisch.	... können aus den einschlägigen Erziehungs- und Bildungstheorien Zielperspektiven und Handlungsprinzipien ableiten.
... kennen allgemeine und fachbezogene Didaktiken und wissen, was bei der Planung von Unterrichtseinheiten auch in leistungsheterogenen Gruppen beachtet werden muss.	... verknüpfen fachwissenschaftliche und fachdidaktische Argumente und planen und gestalten Unterricht auch unter Berücksichtigung der Leistungsheterogenität.
... kennen unterschiedliche Unterrichtsmethoden, Aufgabenformate bzw. Aufgabenformen und wissen, wie man sie anforderungs- und situationsgerecht einsetzt.	... wählen Inhalte, Medien und Methoden, Arbeits- und Kommunikationsformen anhand entsprechender Qualitätskriterien unter Bezug auf Curricula und ggf. individuelle Förderpläne aus.
... kennen Konzepte der Mediendidaktik sowie Möglichkeiten und Grenzen eines anforderungs- und situationsgerechten Einsatzes von analogen und digitalen Medien in Schule und Unterricht.	... integrieren analoge und digitale Medien didaktisch sinnvoll und reflektieren den eigenen Medieneinsatz.
... kennen Verfahren für die Beurteilung von Lehrleistung und Unterrichtsqualität.	... überprüfen die Qualität des eigenen Lehrens und reflektieren die Passung zu den Lernvoraussetzungen und Lernbedürfnissen der Schülerinnen und Schüler.

Es fällt jedoch auf, dass zahlreiche der oben genannten, planungsrelevanten Aspekte (z. B. Gestaltung von Aufgaben; König et al., 2021) nicht konkret benannt werden, was möglicherweise auch damit erklärt werden kann, dass der Gestaltung der *Standards für die Lehrerbildung* (KMK, 2022) kein etabliertes, theoretisch konsistentes und empirisch gesichertes Unterrichtsplanungskompetenzmodell zugrunde gelegt werden konnte. Stützte man sich allein auf die *Standards für die Lehrerbildung* (KMK, 2022), würden viele wichtige Aspekte in den noch zu entwickelnden Kriterien zur Beurteilung schriftlicher Unterrichtsentwürfe nicht berücksichtigt werden. Da es sich in den Standards für die Lehrkräftebildung um zu fördernde *Kompetenzen* handelt, ist klärungsbedürftig, wie Unterrichtsplanungskompetenz in der Forschung bislang definiert und modelliert wird.

Statt eine Synopse der kaum überschaubaren theoretischen und empirischen Forschungsarbeiten zum Kompetenzbegriff zu liefern, die seit dem sogenannten PISA-Schock zu Beginn der 2000er Jahre vor allem in den deutschsprachigen Erziehungswissenschaften und Fachdidaktiken veröffentlicht worden sind, werden in diesem Kapitel ausgewählte theoretische Überlegungen konkret auf die Unterrichtsplanung angewendet und das Konstrukt *Unterrichtsplankompetenz* auf diese Weise definiert. Die bislang inkonsistente Nutzung verschiedener Begriffe wie *Planungswissen* (Wolf, 2022), *Planungsfähigkeiten* (z. B. Schröder et al., 2020), *Planungskompetenz* (z. B. König et al., 2015; Schnebel et al., 2017) oder *Unterrichtsplankompetenz* (König & Rothland, 2022) erfordert es, den Kompetenzbegriff zunächst allgemein zu definieren, um ihn anschließend auf die Unterrichtsplanung anwenden zu können und auf diesem Wege konzeptionelle und terminologische Klarheit zu schaffen.

Auf den ersten Blick scheint in den Naturwissenschaftsdidaktiken Konsens darüber zu bestehen, dass Wissen Bestandteil von Kompetenz ist. So führen Vorholzer und Aufschnaiter (2020) beispielsweise aus:

*Es herrscht große Einigkeit, dass es sich bei Kompetenzen um erlernbare, domänenspezifische kognitive sowie motivationale, volitionale und soziale Fähigkeiten, Fertigkeiten und Bereitschaften handelt (vgl. Kompetenzbegriff nach Weinert, 2001; diskutiert in z. B. [...]Klieme et al., 2007). [...] Obwohl in schulischen Kontexten gelegentlich missverstanden, scheint zudem unstrittig, dass „Wissen“ im Sinne einer verständnisvollen Nutzung von Fakten, Definitionen, Regeln, Gesetzen und Theorien, ebenfalls Bestandteil von Kompetenz ist (Klieme et al. 2007 Klieme et al., 2007 [...]). (Vorholzer & Aufschnaiter, 2020, S. 3)*

Mehrere ist an dieser Aussage bemerkenswert: Erstens ist die „große Einigkeit“ höchst fraglich, wenn man bedenkt, dass der zweite Teil der Kompetenz-Definition von Weinert (2001) in der Fortführung bei Klieme et al. (2007) herausgelöst wurde und diese gar nicht mehr von motivationalen, volitionalen und sozialen Fähigkeiten sprechen. Vergleiche:

*Kompetenzen [sind] die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können. (Weinert, 2001, S. 27f.)*

*Kompetenzen sind kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen. (Klieme et al., 2007, S. 7)*

Entsprechend wenden Forschende in der Naturwissenschaftsdidaktik die Kompetenzdefinition auf die jeweils von ihnen untersuchten Konstrukte an, so dass z. B. auch in der Definition von Modellierkompetenz (Krüger & Upmeyer zu Belzen, 2021, S. 134) nicht mehr von motivationalen, volitionalen und sozialen Bereitschaften gesprochen wird und stattdessen einzig die kognitive Dimension von Kompetenz fokussiert wird. Dies ist in der naturwissenschaftsdidaktischen Forschung nicht unüblich (Krell, 2020). Zweitens wird das bei Weinert (2001) noch erwähnte Problemlösen von Vorholzer und Aufschnaiter (2020) unterschlagen, das im Rahmen dieser Arbeit

jedoch als zentrales Wesensmerkmal von Kompetenz verstanden wird. Und drittens wird zwar in der Tat Wissen als Teil von Kompetenz definiert, doch das daraus erwachsende Missverständnis „in schulischen Kontexten“, wie Vorholzer und Aufschnaiter (2020) oben ausführen, verdient eine genauere Betrachtung der Ursachen. So spricht beispielsweise Neuweg (2022) nicht von *Kompetenz*, sondern von *Könnerschaft*,

*weil ja erst zu zeigen wäre, welches explizite Wissen für welches Können wann und für wen hinreichend, notwendig oder zumindest hilfreich ist. Dies bleibt undeutlich, wenn der Wissensbegriff vom Kompetenzbegriff nicht unterschieden wird und in einer schwer durchschaubaren Weise alles umfasst: ‚theoretische Elemente, und [...] Faustregeln und praktische Erfahrungen‘ (Bromme, 1992, S. 9). Was auch immer Könner wo gelernt haben – in der formalen Ausbildung, in der Kindheit und Schulzeit, in der eigenen Schulpraxis –, man nennt es ‚Wissen‘ und geht davon aus, dass es ‚herangezogen wird, um zu handeln [...]. Können wird dann zum bloßen Ausdruck eines inneren Wissens – mit ‚Null Rest‘ gleichsam. Damit werden die Brüche zwischen Wissen und Können und mit ihnen auch das Interesse an der Könnerschaft entsorgt, um sich dann umso energischer dem Wissen zuwenden zu können und vielleicht über Umwegen [sic!] doch auch wieder den Ausbildungswissenschaften. (Neuweg, 2022, S. 23)*

Inwiefern trifft diese Kritik an der mangelnden Trennschärfe der untersuchten Konstrukte auch auf die Unterrichtsplanungsforschung zu und wie lässt sich ihr im Rahmen der vorliegenden Dissertation begegnen?

Auch wenn es noch keinen Konsens zu *einem* Unterrichtsplanungskompetenzmodell gibt, so liegen inzwischen doch mehrere Kompetenzmodelle und Definitionen von Unterrichtsplanungskompetenz vor, die umfassend und chronologisch beispielsweise von Rothland (2021) und Wolf (2022) kritisch gewürdigt wurden. Schon hinsichtlich einer Definition von Unterrichtsplanungskompetenz besteht allerdings keine Einigkeit (Tab. 2).

## Tabelle 2

Definitionen des Begriffs „*Unterrichtsplanungskompetenz*“

Quelle	Definition
Nauck (1996)	„Wer [...] Planungshilfen selbständig und souverän anwendet, gilt als kompetent.“ (S. 183)
Greiten (2014)	„Als Unterrichtsplanungskompetenz wird die Fähigkeit zur Planung von Unterrichtsstunden, Unterrichtsreihen und offenen Lernumgebungen unter Berücksichtigung fachlichen, fachdidaktischen, methodischen, pädagogischen, lernpsychologischen Wissens und systemischer Bedingungen verstanden. Die Unterrichtsplanungskompetenz besteht aus einem System von Unterkompetenzen.“ (S. 110)
Wernke und Zierer (2017)	„Fachkompetenz, didaktische Kompetenz und pädagogische Kompetenz stehen in einem Wechselwirkungsverhältnis, das an einer Stelle besonders zum Tragen kommen kann, ja muss, wenn Unterricht professionell gestaltet wird. Die Rede ist von der Unterrichtsplanung. In ihr manifestieren sich Fachkompetenz, didaktische Kompetenz und pädagogische Kompetenz. Sie stellen die Pfeiler der Planungskompetenz einer Lehrkraft dar [...].“ (S. 10)

- König et al. (2017) „*Folgt man aktuellen kompetenztheoretischen Überlegungen [...] können die mit Tests aus TEDS-LT („Teacher Education and Development Study: Learning to Teach“, L.G.) erfassten Wissensbestände als kognitive Dispositionen definiert werden, die situationsspezifische Fähigkeiten bedingen, d. h. die erfolgreiche unterrichtliche Wahrnehmung und Interpretation (z. B. von Lernvoraussetzungen der Schülerinnen und Schüler) sowie erfolgreiche unterrichtliche Entscheidungen (z. B. die adaptiven Planungsentscheidungen) wahrscheinlich machen. [...] [W]ir [folgen] dieser allgemeinen Einteilung von Wissen in kognitive Dispositionen und der Planungskompetenz als situationspezifische des Wahrnehmens, Interpretierens und Entscheidens [...]. Planungskompetenz entfaltet sich in einer Planungssituation, welche wiederum durch verschiedene Bedingungen (z. B. die zu unterrichtende Lerngruppe, das Fachcurriculum, die zeitliche Reichweite der Planung) geprägt ist [...], gleichzeitig aber immer Mittel zum Zweck (Unterrichtshandeln, Unterrichtsprozess, Schülerlernen usw.) ist.“ (S. 124f.)*
- Weingarten (2019) „*Im Folgenden sollen in Anlehnung an Funke und Glodowski (1990, 144ff.) für den Bereich der Planerstellung und Planausführung wesentliche Basiskompetenzen im Sinne von erforderlichen Teilleistungen ausdifferenziert und für den Bereich der Unterrichtsplanung näher veranschaulicht werden. Für die Planerstellung, verstanden als gedanklicher Entwurf eines auf Kompetenzerwerb hin ausgerichteten Lernangebots, sind demnach folgende Basiskompetenzen kennzeichnend: 1. Erkennen von zeitlichen Ereignisabfolgen und sinnhafte Sequenzierung [...] 2. Berücksichtigung der zur Verfügung stehenden zeitlichen, materiellen und personenbezogenen Ressourcen [...]. 3. Alternativen einplanen [...]. 4. Angemessenes Auflösungslevel [...].“ (S. 16f.)*
- König und Rothland (2022) „*UPK (Unterrichtsplanungskompetenz, L.G.) [bezeichnet] erlernbare kognitive Fähig- und Fertigkeiten (cognitive abilities and skills) sowie damit verbundene affektiv-motivationale Dispositionen von Lehrkräften, die sich inhaltlich und funktional auf Anforderungen der Planung von Unterricht beziehen und die erfolgreiche Bewältigung solcher Anforderungen wahrscheinlich machen. [...] [D]iese Fähig- und Fertigkeiten [beinhalten] das Treffen von planungsbezogenen Entscheidungen, möglicherweise einhergehend mit oder aufbauend auf kognitiven Prozessen der Wahrnehmung und Interpretationen von Situationen, wie sie von aktuellen situationspezifischen Ansätzen der Kompetenzforschung hervorgehoben werden; sie sind jedenfalls spezifisch für Planungssituationen und unterscheiden sich dadurch von Konzeptionen professionellen Wissens in bisherigen empirischen Studien, die diese Kontextualisierung nicht oder nicht in diesem Auflösungsgrad aufweisen. Gleichwohl sind grundsätzlich Zusammenhänge oder sogar Überschneidungen zu den bislang fokussierten Facetten professionellen Wissens denkbar.“ (S. 781)*
- 

In diesen Definitionen fallen mehrere Dinge auf: Erstens spricht Nauck (1996) noch vor der Implementierung der Kompetenzorientierung in der Lehrkräftebildung zu Beginn der 2000er Jahre von Kompetenz und bezieht sich dabei als einziger auf Planungsstrategien – einen bis heute insgesamt unterbelichteten Aspekt in der Unterrichtsplanungsforschung, wie in Beitrag 1 zu zeigen sein wird. Die anderen Definitionen fokussieren hingegen das Professionswissen als einen Aspekt von professioneller Kompetenz (Baumert & Kunter, 2006). Greiten (2014) bietet dabei einen Zirkelschluss an: Wenn sie Unterrichtsplanungskompetenz als Fähigkeit, Unterricht zu planen definiert, bleibt unklar, worin diese Kompetenz genau besteht. Und während fachliches, fachdidaktisches und pädagogisches Wissen noch leicht mit der etablierten Topologie des Professionswissen (z. B. Baumert & Kunter, 2006) aus CK, PK und PCK gleichgesetzt werden können, bleibt unklar, warum nicht methodisches Wissen dem PCK (vgl. Park & Oliver, 2008) und

lernpsychologisches Wissen dem PK<sup>9</sup> (vgl. Voss et al., 2011) untergeordnet worden sind. Auch Wernke und Zierer (2017) beziehen sich zwar explizit auf CK, PK und PCK, bezeichnen die drei Konstrukte jedoch nicht als Formen von Professionswissen, sondern nennen sie *Kompetenzen*, die dann wiederum „Pfeiler“ von Planungskompetenz seien. Demnach sei „Fachkompetenz als bestimmende Größe des Professionswissens [von Lehrkräften zu] nennen“ (Wernke & Zierer, 2017), so dass Kompetenz Teil von Wissen wäre. Die Annahme, dass Unterrichtsplanung eine wissensbasierte Tätigkeit ist (Calderhead, 1996; Mutton et al., 2011) und dass Lehrkräfte dabei explizit auf die drei Formen von Professionswissen zurückgreifen (Scholl et al., 2019; Wolf, 2022), ist plausibel, doch gibt es bislang beispielsweise für die von Wernke und Zierer (2017) vorgeschlagene Verbindung von PK, CK und PCK zu Unterrichtsplanungskompetenz kaum empirischen Belege (Rothland, 2021). Dies gilt ebenso für die Annahme, dass die Ausprägung von CK, PK und PCK positiv mit der Unterrichtsplanungskompetenz korreliert (Zierer et al., 2015). Für PCK wird dieser mögliche Zusammenhang in Beitrag 3 dieser Arbeit untersucht.

Auch Weingarten (2019) nimmt in seiner Definition einen Ebenenwechsel vor, indem er aus anderer theoretischer Richtung vier Basiskompetenzen für Planungskompetenz benennt und somit ebenfalls einen Bruch zur generischen Kompetenzdefinition (Klieme et al., 2007; Weinert, 2001) schafft: Demnach ist der Kompetenzbegriff klar übergeordnet und Wissen sowie Fähigkeiten werden darunter subsumiert, so dass eine Kompetenz eben nicht als Kombination weiterer Kompetenzen zu verstehen ist. Aus professions- und kompetenztheoretischer Perspektive fruchtbarer scheint der Ansatz von König et al. (2017), die ihr Kompetenzmodell auf dem innovativen kompetenztheoretischen Ansatz von Blömeke et al. (2015) aufbauen. Innovativ ist dieser Ansatz dergestalt, dass er zwei bis dahin verfolgte Kompetenzauffassungen vereint (Wolf, 2022, S. 52ff.): *Holistische* Kompetenzauffassungen wie das Standard-Modell von Oser (1997) beschreiben das angestrebte, von außen beobachtbare Verhalten von Lehrkräften, also ihre Performanz. *Analytische* Kompetenzauffassungen wie beispielsweise das Modell professioneller Handlungskompetenz von Baumert und Kunter (2006) fokussieren die Identifikation von Variablen, die als Voraussetzungen für Kompetenz angesehen werden, also die Dispositionen von Lehrkräften. Blömeke et al. (2015) entwickeln eine Synthese beider Auffassungen, indem sie Kompetenz als Kontinuum verstehen. Ausgehend von kognitiven und affektiv-motivationalen Dispositionen beschreiben sie situationsspezifische Fähigkeiten, die letztlich zu einem beobachtbaren Verhalten – der Performanz – führen. Aus der Performanz wiederum wird auf die

---

<sup>9</sup> So wird das entsprechende Konstrukt beispielsweise sowohl im Projekt COACTIV-R (*Cognitive Activation in the Classroom: The Orchestration of Learning Opportunities for the Enhancement of Insightful Learning in Mathematics–Referendariat*; Voss et al., 2011) als auch im Projekt ProwiN (*Professionswissen von Lehrkräften in den Naturwissenschaften*; Lenske et al., 2015) als *general pedagogical/psychological knowledge* bzw. als *pädagogisch-psychologisches Professionswissen* bezeichnet und lernpsychologisches Wissen somit dem PK zugeordnet.

Kompetenz geschlossen. Dieser Ansatz hat der empirischen Forschung neue Perspektiven eröffnet:

*It is obvious that knowledge and personality tests as well as performance assessments have important functions to fulfill in a competence assessment. The limitation to either-or should be of concern because they each only tap into parts of the construct definition. Using combinations of approaches, we may also be able to cover the processes mediating the transformation of dispositions into performance. (Blömeke et al., 2015, S. 9)*

Genau an dieser Stelle setzen König et al. (2017) und darauf aufbauend König und Rothland (2022) in ihren Definitionen von Unterrichtsplanungskompetenz an. Insbesondere die situationsspezifischen Fähigkeiten werden dabei betont. Diese werden erstaunlicherweise in der Forschung bislang kaum berücksichtigt (König et al., 2021a), obwohl sich Unterrichtsplanung immer auf spezifische Lerngruppen (*adaptives Planen*, Abb. 1) bezieht und es somit naheliegt, konkrete Planungsprozesse bzw. -produkte aus authentischen Kontexten (z. B. realem Unterricht an Schulen, der Staatsexamensprüfung<sup>10</sup>) zu untersuchen. Ein Kategorienfehler scheint jedoch bei König et al. (2017) insofern vorzuliegen, als sie den Begriff *Planungskompetenz* allein auf die situationsspezifischen Fähigkeiten anwenden (S. 125). Dann jedoch wäre Planungskompetenz nicht mehr als Kontinuum von Dispositionen über situationsspezifische Fähigkeiten hin zur Performanz zu verstehen, aus der auf Kompetenz geschlossen werden kann (Blömeke et al., 2015). Stattdessen wäre Planungskompetenz Teil dann von Kompetenz. In der aktuellsten Variante haben König und Rothland (2022) diese Inkonsistenz behoben. Ihre heuristische Arbeitsdefinition von Unterrichtsplanungskompetenz (S. 781) liegt in modifizierter Form dieser Dissertation zugrunde:

*Unterrichtsplanungskompetenz bezeichnet allgemeindidaktisches sowie fachdidaktisches Wissen, erlernbare kognitive Fähig- und Fertigkeiten sowie damit verbundene affektiv-motivationale Dispositionen von Lehrkräften, die genutzt werden, um im Zuge der Unterrichtsplanung zwei situationsspezifische Probleme zu lösen – einerseits die Kompetenzen von Schülerinnen und Schülern didaktisch adaptiv zu fördern und zu diesem Zweck eine Passung zwischen den zu treffenden Planungsentscheidungen herzustellen, und andererseits diese Entscheidungen im Hinblick auf ihr Potential zur Kompetenzförderung zu begründen.*

Drei Änderungen wurden im Vergleich zur Definition von König und Rothland (2022) vorgenommen: (1) Die Fach- und Themenspezifität wurde im Sinne des Begriffs „*kontextspezifisch*“ aus der Kompetenz-Definition von Klieme et al. (2007) ergänzt, so dass fachdidaktisches Wissen (PCK) explizit als Teil von Unterrichtsplanungskompetenz genannt wird. Damit wird der

---

<sup>10</sup> Natürlich stellt auch die Staatsexamensprüfung keinen authentischen Kontext im Vergleich zur Unterrichtsplanung, -durchführung und -reflexion von ausgebildeten Lehrkräften im Schulalltag dar. Abgesehen von Verbeamtungen verfassen diese Lehrkräfte keine schriftlichen Unterrichtsentwürfe mehr. Authentisch ist die Staatsexamensprüfung jedoch insofern, als die dort entwickelten Unterrichtsentwürfe nicht so verfasst worden sind, wie es der Autor dieser Arbeit vorgegeben hat, sondern so, wie es im Land Berlin unter den Gegebenheiten der 2. Phase der Lehrkräftebildung üblich ist. Sie spiegeln somit die Situation am Ende der Lehramtsausbildung in Berlin authentisch wider.

Überlegung Rechnung getragen, dass die Unterrichtsplanung keine rein generische Tätigkeit darstellt, sondern fachspezifische Überlegungen notwendig sind (König et al., 2020b; Scholl et al., 2019). Auf diese Weise wird markiert, dass das im Rahmen dieser Arbeit untersuchte Konstrukt sinnvoll aus fachdidaktischer Perspektive untersucht werden kann. (2) Das Problemlösen wird eingeführt, das Klieme et al. (2007) nicht aus der ursprünglichen Definition von Weinert (2001) übernommen haben. Auf diese Weise wird verdeutlicht, dass die kognitiven Herausforderungen in der Unterrichtsplanung (König et al., 2021) vor allem darin bestehen, eine Passung zwischen den verschiedenen in der Planung zu berücksichtigenden Variablen (z. B. der zu fördernden Kompetenz, dem Fachinhalt, der Lerngruppe) und damit didaktische Adaptivität (König et al., 2015) herzustellen, um Schülerinnen und Schülern einen Kompetenzzuwachs zu ermöglichen (Abb. 1). (3) Um der von Vogelsang und Riese (2017) vorgeschlagenen Heuristik gerecht zu werden und die plausible Annahme, dass Unterrichtsplanung einen iterativen Prozess aus Kreation und Legitimation von Planungsentscheidungen darstellt, aufzugreifen, wurde die Legitimation von Planungsentscheidungen als Teil von Unterrichtsplanungskompetenz ergänzt.

Im Rahmen dieser Arbeit wird Unterrichtsplanungskompetenz aus biologiedidaktischer Perspektive untersucht. Es ist unstrittig, dass (angehende) Lehrkräfte für erfolgreiches Unterrichten fachdidaktisches Wissen benötigen (z. B. Kulgemeyer & Riese, 2018; Rollnick, 2017). Als empirisch untersuchbares Konstrukt wurde das fachdidaktische Wissen als *pedagogical content knowledge* (PCK) von Lee Shulman in die didaktische Forschung eingeführt (Shulman, 1986, Shulman, 1987) und hat sich in den vergangenen drei Jahrzehnten neben dem pädagogischen Wissen (*pedagogical knowledge*, PK) und dem inhaltsbezogenen Fachwissen (*content knowledge*, CK) empirisch als die wichtigste Wissensart von Lehrkräften herausgestellt (Chan & Hume, 2019). PCK stellt die wichtigste Wissensressource von Lehrkräften dar, weil es positiv mit Unterrichtsqualität im Allgemeinen (z. B. Park et al., 2011) und ihren Teilaspekten wie der kognitiven Aktivierung (z. B. Keller et al., 2017), der individuellen Unterstützung von Lernenden (z. B. Kunter et al., 2013) oder auch der Erklärbarkeit von Lehrkräften (Kulgemeyer & Riese, 2018) korreliert. Infolgedessen sind Befunde plausibel, die einen positiven Zusammenhang zwischen dem PCK-Niveau von Lehrkräften und den Lernzuwächsen von Schülerinnen und Schülern beschreiben (z. B. Baumert et al., 2010; Förtsch et al., 2016; Mahler et al., 2017; Sadler et al., 2013).

Die ursprünglichen Annahmen, denen zufolge PCK entweder als distinkte Wissensform gar nicht existiere, sondern in der Anwendung von PK bzw. CK enthalten sei (*integrative model*; Gess-Newsome, 1999) oder dass PCK eine Synthese aus PK und CK darstelle (*transformative model*; Gess-Newsome, 1999), wurden inzwischen empirisch widerlegt: PK, CK und PCK lassen sich als distinkte Wissensdomänen unterscheiden (z. B. Großschedl et al., 2015; Jüttner & Neuhaus, 2013;

König et al., 2018). Gleichwohl wird PCK nach wie vor als *amalgam* aus PK und CK betrachtet (Kind & Chan, 2019; Krepf et al., 2018; Neumann et al., 2019), geht aber nicht in ihnen auf, sondern umfasst neben Wissen über fachspezifische Instruktionsstrategien, das Curriculum und Feedback vor allem auch Wissen über Schülerkognition (Magnusson et al., 1999; Park & Oliver, 2008).

Anders als Shulman (1986, 1987), der PCK als Wissensform betrachtet, wird PCK demnach sowohl als Wissen als auch als Können aufgefasst (Gess-Newsome, 2015). Diese Erweiterung ist für die vorliegende Arbeit deswegen von Bedeutung, weil der oben hergeleiteten Definition von Unterrichtsplanungskompetenz folgend sowohl Wissen als auch Können maßgeblich den Unterrichtsplanungsprozess bestimmen<sup>11</sup>. Deutlich wird dies im *Refined Consensus Model* (RCM; Abb. 2; Carlson et al., 2019), in dem drei sogenannte *realms* von PCK unterschieden werden, also Ebenen, auf denen man PCK betrachten kann:

- *Collective PCK* (cPCK) meint fachdidaktisches Wissen, das in einer Community (z. B. unter Forscherinnen und Forschern in der Biologiedidaktik) als gesichert gilt und akzeptiert wird. Es manifestiert sich beispielsweise in Forschungspublikationen oder fachdidaktischen Lehrbüchern und entspricht am ehesten der Konzeption von PCK, wie sie Shulman (1986, 1987) vorgeschlagen hat.
- *Personal PCK* (pPCK) meint das fachdidaktische Wissen einer individuellen (angehenden) Lehrkraft. Es ist das *„reservoir of knowledge and skills that the teacher can draw upon during the practice of teaching“* (Carlson et al., 2019, S. 85) und das Ergebnis der individuellen Erfahrungen einer (angehenden) Lehrkraft. Auch wenn hier von der *„practice of teaching“* gesprochen wird, ist es plausibel anzunehmen, dass pPCK auch in der Planung und der Reflexion genutzt wird.
- *Enacted PCK* (ePCK) meint den Teil des pPCK einer individuellen (angehenden) Lehrkraft, der im Prozess des Planens, Durchführens oder Reflektierens von Unterricht genutzt wird. Es ist *„the specific knowledge and skills utilised by an individual teacher in a particular setting, with a particular student or group of students, with a goal for those students to learn a particular concept, collection of concepts, or a particular aspect of the discipline“* (Carlson et al., 2019, S. 83f.).

Im Prozess des enactments wird also ein Teil des pPCK als kognitive Wissensressource in Handeln überführt. Dabei wird unterschieden, ob dieses Handeln in der Planung vor dem Unterricht (*plan*; ePCK<sub>p</sub>; obere Hälfte in Abb. 1), während der Durchführung des Unterrichts (*teach*; ePCK<sub>t</sub>) oder in

---

<sup>11</sup> Wie später noch zu diskutieren sein wird, ist die oft fehlende begriffliche Unterscheidung zwischen Wissen und Können unter anderem auch insofern problematisch, als dadurch Dissonanzen zwischen dem Wissen einer (angehenden) Lehrkraft und ihrem Können definitorisch verschleiert werden (Neuweg, 2022, S. 22f.). Aus diesem Grunde und zum Zwecke der besseren Anschlussfähigkeit an den internationalen Diskurs wird in Beitrag 1 der Kompetenz-Begriff vermieden und von *„knowledge and skills for lesson planning“* gesprochen.

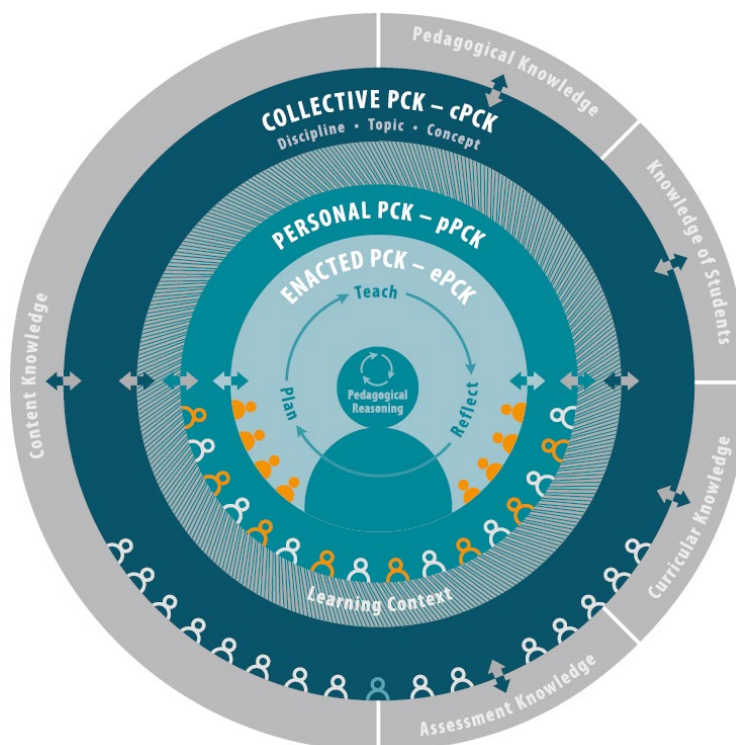


der Reflexion nach dem Unterricht (*reflect*; ePCK<sub>r</sub>; beides untere Hälfte in Abb. 1) stattfindet (Alonzo et al., 2019). Die in der Definition von Unterrichtsplanungskompetenz benannten situationspezifischen Probleme sind es, die durch die Nutzung von ePCK<sub>p</sub> gelöst werden müssen. Voraussetzung dafür wiederum ist, dass man neben PK und CK über hinreichendes pPCK verfügt. Als Konzeptualisierung des ePCK wird in dieser Arbeit das Pentagon-Modell (Park & Oliver, 2008) genutzt, das die Vernetzung verschiedener Wissensfacetten beschreibt: *Orientation to Teaching Science* (OTS), *Knowledge of the Science Curriculum* (KSC), *Knowledge of Students' Understanding in Science* (KSU), *Knowledge of Instructional Strategies* (KISR) und *Knowledge of Assessment Strategies* (KAs). Park und Oliver (2008) betonen, dass diese Aspekte im Zuge des Planens, Durchführens und Reflektierens von Unterricht in Beziehung zueinander gesetzt werden sollen, also integriert werden:

*This integration is accomplished through the complementary and ongoing readjustment by both reflection-in-action and reflection-on-action, resulting in strengthened coherence among the components [...]. This model suggests that the development of one component within PCK will, in turn, influence the development of others, and ultimately enhance this holistic PCK. Because PCK, which comprises effective teaching, requires the integration of the components in highly complex ways, lack of coherence among components can be problematic in developing PCK and increased knowledge of a single component may not be sufficient to stimulate significant change in practice. (Park & Oliver, 2008, S. 814)*

## Abbildung 2

*Refined Consensus Model (RCM) of PCK (Carlson et al., 2019, S. 83)*



*Anmerkung.* Die im Zentrum des Modells dargestellten Tätigkeiten entsprechen den im Fließtext benannten Konstrukten, d. h. *plan* = ePCK<sub>p</sub>, *teach* = ePCK<sub>t</sub> und *reflect* = ePCK<sub>r</sub>.

In einem Wechselspiel aus *reflection-in-action* (während des Unterrichts) und *reflection-on-action* (in der Planung vor oder der Reflexion nach dem Unterricht) werden die fünf PCK-Facetten fortwährend aufeinander bezogen. Dieser Prozess wird als *pedagogical reasoning* bezeichnet (Park & Suh, 2019) und bildet den Kern fachdidaktischen Handelns von Lehrkräften, so dass er im RCM zentral verortet wird (Abb. 2; Carlson et al., 2019). Auch wenn Park und Oliver (2008) die Bedeutung der Vernetzung von ePCK<sub>p</sub> einzig aus der PCK-Forschung herleiten, lässt sie sich auch vor dem Hintergrund der am Experten-Novizen-Paradigma ausgerichteten kognitionspsychologischen Studien der 1970er, 1980er und 1990er Jahre begründen: Anders als Novizinnen und Novizen greifen Expertinnen und Experten auf hierarchisch strukturierte, differenzierte und vernetzte Wissensnetzwerke zurück, wenn sie mit problemhaltigen Situationen konfrontiert sind (z. B. Glaser & Bassok, 1989; Gobbo & Chi, 1986; Shavelson, 1972). In diesem Sinne stellen Pearsall et al. (1997) fest:

*As new knowledge is constructed, concept meanings are clarified and the knowledge base takes on a more complex form that includes multiple levels of hierarchy and a branching, dendritic structure. This process that results in increasing structural complexity is referred to as progressive differentiation; hence, both subsumption and superordinate learning result in the progressive differentiation of cognitive structure. Additionally, the process of progressive differentiation is accompanied by the explicit delineation of similarities and differences among closely related concepts—an event called integrative reconciliation. When individuals learn about differences and similarities in the atmospheres of Earth, Venus, and Mars, for example, their knowledge structures become more interconnected, integrated, and cohesive. This kind of integration is especially characteristic of expertise in a knowledge domain [...]. (Pearsall et al., 1997, S. 196)*

Auf das ePCK<sub>p</sub> angewendet bedeutet dies, dass deklaratives und konzeptuelles Wissen allein (z. B. über Schülervorstellungen, über Instrukionsstrategien) nicht genügen, um guten Unterricht zu planen. Neben diesem *know what* benötigen Lehrkräfte *know how*, d. h. prozedurales Wissen darüber, wie das verfügbare deklarative und konzeptuelle Wissen angewendet werden kann (Shavelson, 1972; vgl. Stender & Brückmann, 2020), um erfolgreiche Lehr-Lern-Prozesse zu planen. So müssten beispielsweise zur Planung einer Stunde neben den curricularen Standards (KSC) und dem daraus abzuleitenden Lernziel der aktuelle Lernstand der Schülerinnen und Schüler (KSU) diagnostiziert und daran angepasst Instrukionsstrategien (KISR) ausgewählt und gestaltet werden, die genau diesen Schülerinnen und Schülern einen Lernzuwachs ermöglichen, der wiederum auf geeignetem Wege diagnostiziert werden muss (KAs). Das deklarative Wissen über den Lehrplan, in der Literatur beschriebene Schülervorstellungen oder Unterrichtsmethoden allein genügt also nicht, um eine funktionale Prozesstruktur des Unterrichts zu entwickeln. Wenngleich nicht auf die Unterrichtsplanung im Besonderen, sondern auf Lernprozesse im Allgemeinen bezogen, beschreiben Glaser und Bassok (1989) das Verhältnis von deklarativem bzw. konzeptuellem Wissen auf der einen und prozeduralem Wissen auf der anderen Seite so:

*[...] [T]he course of knowledge acquisition proceeds from a declarative or propositional form to a compiled, procedural, condition-action form [...]. Novices can know a principle, or a rule, or a specialized vocabulary*

*without knowing the conditions of effective application. In contrast, when experts access knowledge, it is functional or bound to conditions of applicability. Moreover, experts' knowledge is closely tied to their conceptions of the goal structure of a problem space. Experts and novices may be equally competent at recalling specific items of information, but experts chunk these items in memory in cause and effect sequences that relate to the goals and subgoals of problem solution and use this information for further action. The progression from declarative knowledge to well-tuned functional knowledge is a significant dimension of developing competence. (Glaser & Bassok, 1989, S. 635)*

Die Auffassung, dass sich nicht in der Verfügbarkeit konzeptuellen Wissens, sondern in dessen Überführung in prozedurales Wissen Kompetenz ausdrückt (Stender & Brückmann, 2020), schließt unmittelbar an die oben genannte, dieser Arbeit zugrunde liegende Definition von Unterrichtsplanungskompetenz an. Die Annahme dieser Arbeit lautet folglich, dass eine Lehrkraft, die die einzelnen Facetten ihres ePCK<sub>p</sub> miteinander verknüpfen und somit auf eine vernetzte Wissensbasis zurückgreifen kann, die Probleme, mit denen sie in der Planung konfrontiert ist, erfolgreicher lösen kann und somit über ausgeprägtere Unterrichtsplanungskompetenz verfügt als eine Lehrkraft, die über weniger stark vernetztes ePCK<sub>p</sub> verfügt.

Diese Annahme basiert auf drei Befunden: Erstens wurde in der Expertiseforschung wiederholt gezeigt, dass die Vernetzung von Wissen ein wesentliches Kennzeichen von Expertinnen und Experten im Vergleich zu Novizinnen und Novizen darstellt (z. B. Berliner, 2001; Dreyfus & Dreyfus, 1987; Livingston & Borko, 1989; Westerman, 1991). Zweitens besteht innerhalb der PCK-Forschung Konsens darüber, dass weniger die Berücksichtigung einzelner PCK-Facetten beim Planen, Durchführen und Reflektieren von Unterricht als vielmehr deren Vernetzung ein wesentliches Qualitätsmerkmal von Lehrkräften darstellt (Abell, 2008; Chan et al., 2019; Hashweh, 2005; Park & Chen, 2012; Park & Oliver, 2008; Reynolds & Park, 2021; Schiering et al., 2023; Schneider & Plasman, 2011). Dies ist insofern plausibel, als Unterrichtsqualität in hohem Maße davon abhängt, inwiefern eine Lehr-Lern-Umgebung gestaltet wird, die kompetenzorientiert ist und adaptiv an den bisherigen Kompetenzstand der Schülerinnen und Schüler anknüpft sowie mit geeigneten Instruktionsstrategien einen Kompetenzzuwachs ermöglicht. König et al. (2015) zeigen dies am Beispiel von Aufgabenstellungen und Binnendifferenzierungsmaßnahmen. Zu diesem Zwecke müssen verschiedene planungsrelevante Aspekte zueinander in Beziehung gesetzt und funktionale Planungsentscheidungen getroffen werden. Wie in den Beiträgen 3 und 9 zu zeigen sein wird, meint Funktionalität in diesem Zusammenhang, dass eine Planungsentscheidung einen Beitrag dazu leistet, dass eine spezifische Lerngruppe den angestrebten Kompetenzzuwachs erfährt. Dies deckt sich mit der Auffassung von van Driel et al. (2014) zur Frage, was elaboriertes PCK überhaupt auszeichnet:

*[H]igh-quality PCK is not characterized by knowing as many strategies as possible to teach a certain topic plus all the misconceptions student may have about it but by knowing when to apply a certain strategy in recognition*

*of students' actual learning needs and understanding why a certain teaching approach may be useful in one situation.* (van Driel et al., 2014, S. 865)

Drittens führen Decristan et al. (2020) aus Perspektive der Unterrichtsqualitätsforschung aus, dass Lehr-Lernprozesse so choreographiert werden sollen, dass Elemente der Oberflächen- und Tiefenstrukturebene<sup>12</sup> so aufeinander abgestimmt sind, dass sie ihre lernförderliche Wirkung entfalten können (zur Choreographie-Metapher, siehe Oser & Baeriswyl, 2001).

Was im Rahmen dieser Arbeit also als Unterrichtsplanungskompetenz verstanden wird, wird aus der Unterrichtsplanungsforschung, der PCK-Forschung und der Unterrichtsqualitätsforschung gestützt: Es geht um die Fähigkeit, funktionale Planungsentscheidungen zu treffen (*Problem 1. Art*, Abb. 3), ihre Auswirkungen zu antizipieren und sie zu begründen (*Problem 2. Art*, Abb. 3). Im Funktionalitätsbegriff ist immanent angelegt, dass eine Planungsentscheidung nicht per se richtig oder falsch ist, sondern dass jeweils für jede zu fördernde Kompetenz in der jeweiligen Lerngruppe unter Berücksichtigung des inhaltlichen Themas zu prüfen ist, welche Instruktionsstrategien am geeignetsten sind, um das Lernziel zu erreichen. Diese Entscheidungen treffen und begründen (Vogelsang & Riese, 2017) zu können, stellt die zentrale Herausforderung in der Unterrichtsplanung dar und erfordert unter anderem PCK (z. B. Riese et al., 2022)

Führt man nun die Überlegungen zur Unterrichtsplanung als Modellierprozess (Abb. 1) und zum PCK (Abb. 2) mit der Kompetenzauffassung von Blömeke et al. (2015) zusammen und betont das in den Definitionen zur Unterrichtsplanungskompetenz (Tab. 2) vernachlässigte Problemlösen als konstitutives Kennzeichen von Kompetenz (Weinert, 2001), ergibt sich das folgende heuristische Unterrichtsplanungskompetenzmodell (Abb. 3) als konzeptueller Rahmen der vorliegenden Arbeit.

Zur terminologischen Klarheit sei abschließend anzumerken, wie der Begriff *Qualität* im Rahmen dieser Arbeit aufgefasst wird: Bislang wurde auf Unterrichtsqualität Bezug genommen, die mittels der drei Basisdimensionen definiert wird und die überwiegend in quantitativ ausgerichteten Studien den Einfluss bestimmter Variablen auf den Lernerfolg von Schülerinnen und Schüler untersucht hat (Begrich et al., 2023; Praetorius et al., 2020). In Bezug auf die Unterrichtsplanung kann der mögliche Lernerfolg zwar gut begründet, aber nur im Sinne einer probabilistischen Voraussage eingeschätzt werden. Als Fluchtpunkt des im Rahmen dieser Arbeit wichtigen Begriffs *Planungsqualität*, die mithilfe des zu entwickelnden Kriterienrasters eingeschätzt werden soll (Krepf

---

<sup>12</sup> Alternativ ist der Begriff *Sichtstruktur* geläufig. Dieser suggeriert jedoch, dass nur Merkmale auf der Sicht- bzw. Oberflächenstrukturebene sichtbar seien, obwohl die auf der Tiefenstrukturebene liegenden latenten Merkmale auf dem Wege der empirischen Prüfung durch Indikatoren ebenfalls sichtbar gemacht werden können (Decristan et al., 2020). Dieser Argumentation folgend wird im Rahmen dieser Arbeit nicht von *Sichtstrukturen*, sondern von *Oberflächenstrukturen* gesprochen.

& König, 2022b), taugt der mögliche Lernerfolg also nur mittelbar. In diesem Sinne stellt Klafki (2007) fest:

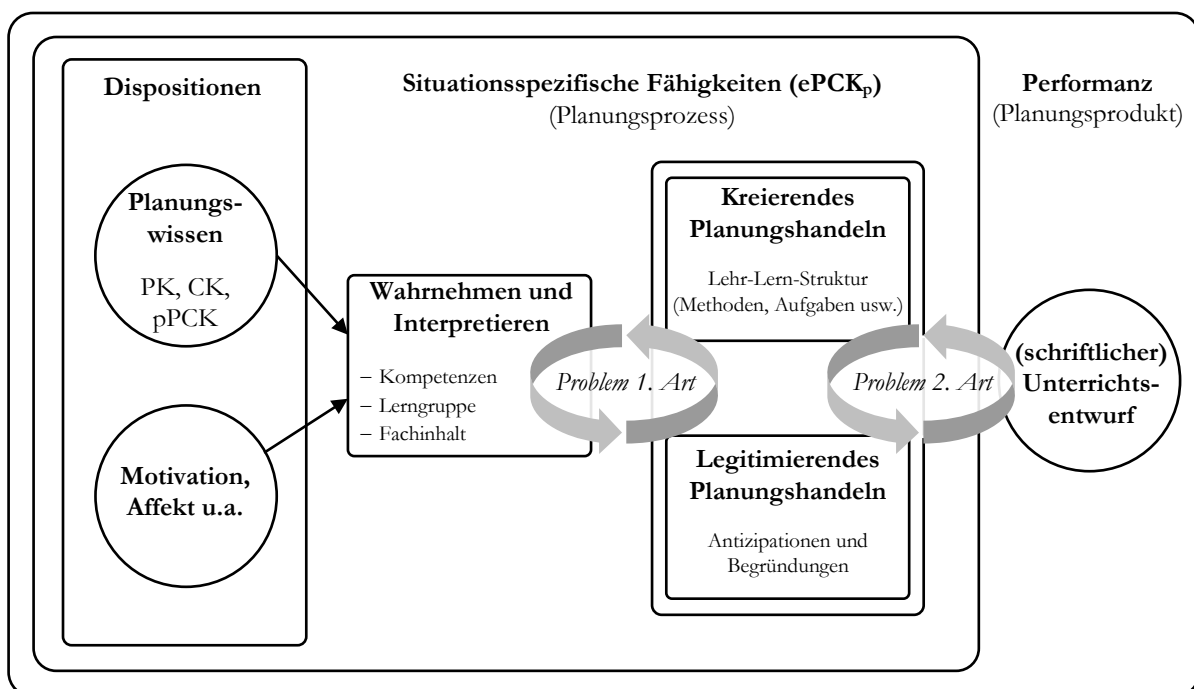
*„Unterrichtsplanung im hier vertretenen Sinn kann nie mehr als ein offener Entwurf sein, der den Lehrer zu reflektierter Organisation, Anregung, Unterstützung und Bewertung von Lernprozessen und Interaktionsprozessen, also zu flexiblem Unterrichts Handeln befähigen soll. Der Maßstab für die didaktische Qualität einer Unterrichtsplanung ist nicht, ob der tatsächlich abgelaufene Unterricht dem Plan möglichst genau entspricht, sondern ob die Planung dem Lehrer didaktisch begründbares, flexibles Handeln im Unterricht und den Schülern produktive Lernprozesse, die einen – wie auch immer begrenzten – Beitrag zur Entwicklung ihrer Selbstbestimmungs- und Solidaritätsfähigkeit darstellen, ermöglichte.“* (Klafki, 2007, S. 269)

Während Klafki (2007) hier vor allem die didaktische Begründung des geplanten Unterrichts betont und damit das Wechselspiel von Kreation und Legitimation (Vogelsang & Riese, 2017; Abb. 2) vorwegnimmt, geht für Kantreiter (2022) Planungsqualität in den Merkmalen der Unterrichtsqualität auf. Sie definiert den Begriff als

*„angemessene Berücksichtigung von fach- und themenspezifischen Qualitätsmerkmalen, welche für die Unterrichtsqualität entscheidend sind. Über die Unterrichtsqualität vermittelt sichert die Planungsqualität den kognitiven und motivationalen Entwicklungsfortschritt der Schüler\*innen [...]. Analog zur Unterrichtsqualität sind auch die Merkmale der Planungsqualität auf der Ebene der Tiefenmerkmale zu verorten [...] Dabei umfasst die Planungsqualität sowohl eine Prozess- als auch eine Produktebene [...]“* (Kantreiter 2022, S. 100)

**Abbildung 3**

*Heuristisches Unterrichtsplanungskompetenzmodell (nach Blömeke et al., 2015; König et al., 2017)*



Hilfreich ist die Unterscheidung zwischen Prozess- und Produktebene insofern, als im Rahmen dieser Arbeit schriftliche Unterrichtsentwürfe als Planungsprodukte, und nicht die vorangestellten Planungsprozesse untersucht werden. Eine stärkere Anbindung des Begriffs *Planungsqualität* an den Kompetenz-Diskurs führt im Rahmen dieser Arbeit ohnehin dazu, dass Planungsprozesse selbst nicht als Performanz (Blömeke et al., 2015) aufgefasst werden (Abb. 3), sondern bloß der Unterrichtsentwurf im Sinne eines Modells für eine Prozessstruktur der durchzuführenden Unterrichtsstunde (Abb. 1). Dieser Unterrichtsentwurf kann im Sinne von Zaragoza et al. (2023) ein geschriebenes Dokument sein – so wie im Rahmen dieser Dissertation –, oder ein vorangestellter mentaler Entwurf einer Stunde, der nicht schriftlich fixiert sein muss.

Im Rahmen dieser Arbeit wird folglich definiert: Ein Unterrichtsentwurf *von hoher Planungsqualität* zeichnet sich dadurch aus, dass das Problem 1. Art (Abb. 3) so gelöst wird, dass die wesentlichen Planungsentscheidungen (z. B. Phasenstruktur, Aufgaben) funktional sind, also einen Beitrag zum Kompetenzzuwachs der adressierten Lerngruppe leisten (Kreation; Vogelsang & Riese, 2017), und dass das Problem 2. Art (Abb. 3) so gelöst wird, dass die entsprechenden Planungsentscheidungen in ihrem Potential zur Kompetenzförderung fachdidaktisch begründet werden (Legitimation; Vogelsang & Riese, 2017). Sowohl die kreierte Planungsentscheidungen als auch ihre Begründungen sollten dabei im Einklang mit den empirisch gestützten Merkmalen von Unterrichtsqualität stehen. In diesem Sinne wird eine hohe Planungsqualität im Rahmen dieser Arbeit mit ausgeprägter Unterrichtsplanungskompetenz gleichgesetzt, denn die Planungsqualität zeigt sich in der Performanz, also dem schriftlichen Unterrichtsentwurf. Von der Performanz kann laut Blömeke et al. (2015) auf die Kompetenz geschlossen werden.

Zusammengefasst werden im Rahmen der folgenden Beiträge die fünf in der Einleitung skizzierten Desiderate auf folgende Weise untersucht: Erstens wird im Rahmen eines Scoping Reviews ein Überblick über die empirische naturwissenschaftsdidaktische Forschung gegeben (Beitrag 1). Dabei geht es unter anderem darum, die bislang untersuchten Dispositionen und ihre Bedeutung für die situationsspezifischen Fähigkeiten (Abb. 3) systematisch zu erfassen und auf der Basis des RCM ein heuristisches Modell vorzuschlagen, das die beim Planen von Unterricht notwendigen Variablen darstellt. Zweitens werden im Rahmen einer Analyse von Staatsexamensentwürfen die situationsspezifischen Fähigkeiten (Abb. 3) hinsichtlich des Zusammenhangs von ePCK<sub>p</sub> und Planungsqualität analysiert (Beitrag 2 und 3). Während diese Analyse derart generische Aspekte der Unterrichtsplanung fokussiert, dass sie in nahezu gleicher Weise auch in anderen Fächern durchgeführt werden könnte, werden in zwei weiteren Beiträgen zwei Besonderheiten des Biologieunterrichts untersucht: der Umgang mit Schülervorstellungen als einem wichtigen Aspekt der Lerngruppe (Beitrag 4) und die Förderung von Erkenntnisgewinnungskompetenzen als einem

der vier Kompetenzbereiche des Biologie-Curriculums. Als Datenmaterial werden Staatsexamensentwürfe genutzt, so dass auf diese Weise der Fokus auf der 2. Phase der Lehrkräftebildung liegt. Drittens wird auf der Grundlage der Beiträge 1-5 ein Kriterienraster entwickelt, so dass erstmals theoriebasierte und empirisch geprüfte Kriterien zur Beurteilung schriftlicher Unterrichtsentwürfe (Krepf & König, 2022b) vorliegen. Hierbei geht es also darum, die Performanz, d. h. die schriftlichen Unterrichtsentwürfe, in ihrer Qualität bewerten zu können (Abb. 3). Um sicherzustellen, dass das Kriterienraster tatsächlich geeignet ist, die Qualität schriftlicher Unterrichtsentwürfe valide zu erfassen, wurde ein umfangreiches Validitätsargument auf der Basis der *Standards for Educational and Psychological Testing* (AERA et al., 2014) entwickelt.

Zunächst also wird ein Überblick über den aktuellen Forschungsstand anhand eines Scoping Reviews geliefert.

### **2.2. Beitrag 1: *Establishing common ground in empirical research on science teachers' lesson planning – A scoping review***

Beitrag 1 wurde noch nicht publiziert, aber ist eingereicht: Großmann, L., Koberstein-Schwarz, M., Scholl, D., Krüger, D. & Meisert, A. (eingereicht). Establishing common ground in empirical research on science teachers' lesson planning – A scoping review. [geteilte Erstautorenschaft]

#### **Zusammenfassung**

In diesem Beitrag wurden die folgenden Forschungsfragen untersucht:

- F1. What is the state of science education research on lesson planning in terms of demographics, methodological approaches, and bibliographic coupling?
- F2. Which aspects of science teachers' knowledge and skills for lesson planning are investigated in science education research?
- F3. Which variables influence science teachers' knowledge and skills for lesson planning?

Ausgehend von der Beobachtung, dass es bislang noch keinen systematischen Überblick über die empirische naturwissenschaftsdidaktische Unterrichtsplanungsforschung gibt, wurde in Beitrag 1 ein Scoping Review durchgeführt. Auf diesem Wege konnte einerseits konsolidiert werden, wie das Forschungsfeld bislang methodisch arbeitet (F1, F2), und andererseits konnte zusammengefasst werden, welche Dispositionen (Blömeke et al., 2015) für das Konstrukt Unterrichtsplanungskompetenz bislang empirisch untersucht wurden. Anders als die Anordnung der Beiträge in dieser Dissertation suggeriert, war das Scoping Review nicht Ausgangspunkt des Dissertationsvorhabens. Stattdessen ergab sich im Laufe der Forschungsarbeiten und angesichts

der zunehmenden Publikationsaktivitäten in der Unterrichtsplanungsforschung die Notwendigkeit, den aktuellen Forschungsstand systematisch zusammenzufassen. Die Befunde werden im Rahmen dieser Arbeit als theoretischer Rahmen genutzt, ehe die weiteren Beiträge dann den Fokus auf die empirische Untersuchung von Unterrichtsplanungskompetenz legen. Angesichts der nach wie vor geringen Verbreitung des Kompetenz-Begriffs wird im Beitrag selbst nicht von Unterrichtsplanungskompetenz, sondern von *knowledge and skills for lesson planning* gesprochen.

Es ist zu beachten, dass sich die meisten der untersuchten  $N = 66$  Studien gar nicht *expressis verbis* in der Unterrichtsplanungsforschung verorten und das untersuchte Konstrukt auch nicht „Unterrichtsplanungskompetenz“ nennen. Stattdessen handelt es sich beispielsweise oft um PCK-Studien, die als Datenmaterial schriftliche Unterrichtsentwürfe nutzen und die eher die Analyse von PCK statt von Unterrichtsplanungskompetenz fokussieren. Aus der Fülle der in Beitrag 1 berichteten Ergebnisse sind an dieser Stelle die folgenden vier besonders relevanten Beobachtungen festzuhalten.

Erstens konnte die Annahme von König und Rothland (2022) bestätigt werden, dass die Forschung zur Unterrichtsplanung ein emergierendes Feld darstellt: Während im Zeitraum von 1987-2004 nur  $n = 3$  Publikationen gefunden wurden, die den Einschlusskriterien genügten und keines der Ausschlusskriterien erfüllten, waren es für den Zeitraum 2005-2022 insgesamt  $n = 63$  Artikel. Dabei fällt auf, dass die überwiegende Anzahl dieser Studien qualitative Zugänge wählt und anhand kleiner Stichproben ohne Experimental-Kontrollgruppen-Designs mögliche Zusammenhänge zwischen Variablen wie dem PCK, CK oder *beliefs* auf der einen und Unterrichtsplanung auf der anderen Seite untersucht. Offenbar werden weiterhin in offenen Verfahren mögliche Zusammenhänge exploriert, um Einsichten in die Unterrichtsplanung von Lehrkräften zu erhalten. Auf dieser Basis ließe sich ein noch nicht vorhandenes Kompetenzmodell entwickeln, das diese Zusammenhänge auf Niveaustufen expliziert. Erst vor dem Hintergrund einer auf qualitative Weise gefestigten empirischen Basis lassen sich dann zukünftig auch Instrumente entwickeln, mit denen Unterrichtsplanungskompetenz valide erfasst werden könnte. Dass also bislang kaum quantitative Forschung mit großen Stichproben durchgeführt wird, kann als Beleg dafür interpretiert werden, dass sich die Unterrichtsplanungsforschung konzeptionell noch in ihren Anfängen befindet.

Zweitens wurden jedoch zahlreiche empirische Befunde zu verschiedensten Variablen gefunden, die auf Unterrichtsplanungskompetenz einwirken. Auf dieser Basis konnte ausgehend vom RCM (Carlson et al., 2019) ein heuristisches Modell entwickelt werden, das die derzeitige empirische Befundlage zusammenfasst. Wenn das Feld also dennoch weiterhin vorwiegend qualitativ exploriert, stellt sich die Frage, warum die reichhaltig vorhandenen Befunde nicht bereits von anderen Wissenschaftlerinnen und Wissenschaftlern in ein derartiges Modell überführt wurden



oder auf der Basis dieser Befunde quantitative Forschung initiiert wurde. Die Analyse der bibliographischen Kopplung gibt darüber Aufschlüsse: Es wurde untersucht, inwiefern die  $N = 66$  in das Scoping Review einbezogenen Studien dieselben Publikationen zitieren, wie stark sie also bibliographisch gekoppelt sind. Es zeigte sich, dass zwar auf allgemeine Literatur zur Professionalität von Lehrkräften (z. B. Grossmann, 1990; Shulman, 1987) und zur qualitativen Forschung (z. B. Miles & Huberman, 1994) Bezug genommen wird, jedoch kaum auf Studien zur Unterrichtsplanung. Insgesamt verorten sich die gefundenen Studien offenkundig eher in derjenigen Literatur, der ihre zu untersuchenden Konstrukten entspringen (z. B. der PCK-Forschung), als der Literatur zur Unterrichtsplanung. Hierin läge auch ein Appell an Forschende, die Unterrichtsplanung empirisch untersuchen, verstärkt an diesbezügliche aktuelle theoretische und empirische Befunde anzuknüpfen, die inzwischen reichhaltig vorhanden sind (z. B. Beckmann & Ehmke, 2023; König, Bremerich-Vos, Buchholtz, Fladung & Glutsch, 2020b; König et al., 2021a; Koberstein-Schwarz & Meisert, 2022; Scholl et al., 2022; Vogelsang et al., 2022; Weitzel & Blank, 2020; Zaragoza et al., 2023).

Drittens zeigte sich, dass die Unterrichtsplanungsforschung bislang überwiegend *conceptual knowledge* untersucht und nur in wenigen Ansätzen auch *procedural knowledge* in den Blick nimmt. Dies ist insofern relevant, als *conceptual knowledge* im Sinne beispielsweise einzelner PCK-Facetten (vgl. Magnusson et al., 1999) zwar eine notwendige Bedingung für Unterrichtsplanungskompetenz darzustellen scheint, aber noch keine hinreichende: Erst die Integration der verschiedenen Wissensbestände und damit die Herstellung von Interdependenzen (Scholl et al., 2022) zwischen ihnen führt dazu, dass eine funktionale Prozessstruktur von Unterricht entwickelt werden kann. Wenngleich für die weiteren Beiträge nicht von Belang, erwächst hieraus der Auftrag, die Planungsstrategien von (angehenden) Lehrkräften genauer zu untersuchen, die in ihren Planungsprozessen diese Interdependenzen mehr oder weniger erfolgreich herstellen. So ließe sich beispielsweise prüfen, welche Planungsstrategien Lehrkräfte anwenden, die Unterricht mit hoher Planungsqualität planen können und wie genau sie ihr *conceptual knowledge* nutzen. Im Sinne des oben vorgeschlagenen heuristischen Kompetenzmodells (Abb. 3) müssten hier also die situationsspezifischen Fähigkeiten in den Blick genommen werden, die bislang empirisch noch unterbeleuchtet sind (König et al., 2021a).

Viertens wurde gezeigt, dass die weit überwiegende Anzahl der Studien nicht Planungsprozesse, sondern Planungsprodukte untersucht. Die vorliegende Arbeit reiht sich damit methodisch in eine Forschungslinie ein, die Unterrichtsentwürfe nutzt, um beispielsweise PCK von (angehenden) Lehrkräften zu analysieren. Insofern ließe sich am Innovationsgehalt der vorliegenden Arbeit zweifeln. Es zeigt sich jedoch, dass keine einzige dieser Studien Unterrichtsentwürfe mit dem Ziel

untersucht, die Planungsqualität kriterienorientiert zu ermitteln. Stattdessen werden einzelne Variablen professioneller Kompetenz (z. B. PCK, *beliefs*) in den Unterrichtsentwürfen identifiziert und beschrieben. Die ganzheitliche Betrachtung von Unterrichtsentwürfen mit dem Ziel, die Planungsqualität valide einzuschätzen, wurde von keiner der 66 Studien in diesem Scoping Review angestrebt.

### 2.3. Zusammenfassung

Ausgangspunkt der bisherigen Ausführungen war die Frage, inwiefern Unterrichtsplanungskompetenz eine „Jedermannskompetenz“ darstellt. Sowohl aus theoretischer als auch aus empirischer Perspektive wurde gezeigt, dass Unterrichtsplanungskompetenz ein komplexes Konstrukt darstellt: Aus der Literatur zur Unterrichtsplanung, zur Unterrichtsqualität und zur Lehrkräfteprofessionalität konnte basierend auf Blömeke et al. (2015) ein heuristisches Unterrichtsplanungskompetenzmodell (Abb. 3) vorgeschlagen werden. Sowohl die verschiedenen Dispositionen als auch die Herstellung von Interdependenzen zwischen den planungsrelevanten Aspekten als zentrale Problemlösefähigkeit zeigen, dass die Unterrichtsplanung eine komplexe Herausforderung darstellt. Legitimiert wird der Ansatz, ein derartiges Kompetenzmodell vorzuschlagen, durch die empirischen Befunde des Scoping Reviews: Insbesondere hinsichtlich der zahlreichen Dispositionen (d. h. PCK, CK, *beliefs*) als auch hinsichtlich der situationsspezifischen Fähigkeiten (z. B. Planungsstrategien) liegen inzwischen Befunde vor, die zeigen, dass Unterrichtsplanung eine vor allem kognitiv herausfordernde Aufgabe für Lehrkräfte darstellt (König et al., 2021), die ein systematisierendes Modell erforderlich macht.

Einen verbreiteten Ansatzpunkt, um diese Herausforderungen zu untersuchen, bieten schriftliche Unterrichtsentwürfe. Sie stellen die Performanz dar, anhand derer auf die zugrunde liegende Unterrichtsplanungskompetenz geschlossen werden kann. Dass von den 66 im Scoping Review untersuchten Studien keine einzige den Versuch unternommen hat, jenseits der Identifikation beispielsweise von PCK-Facetten ganzheitlich die Planungsqualität der Unterrichtsentwürfe valide zu erfassen, wird als Desiderat in den folgenden Beiträgen aufgegriffen, um ein Kriterienraster zu entwickeln. Dass ausführliche schriftliche Unterrichtsentwürfe möglicherweise aber auch eine Besonderheit der deutschen Lehrkräftebildung darstellen und daher in den englischsprachigen Artikeln ein solcher Zugang gar nicht gefunden werden konnte, sollte zumindest dergestalt berücksichtigt werden, dass nicht auszuschließen ist, dass es in nicht-englischsprachigen Publikationen möglicherweise doch bereits entsprechende Forschungsvorhaben gegeben haben könnte.

Im Folgenden wird ein solches Kriterienraster entwickelt, indem zunächst qualitativ Staatsexamensentwürfe analysiert und aus dieser Analyse Niveaustufen planungsrelevanter Kategorien entwickelt werden. Das so entwickelte Kriterienraster wird umfassend auf Validität der Testscore-Interpretationen (AERA et al., 2014) geprüft.

*„Pläne machen ist mehrmalen eine üppige, prahlerische Geistesbeschäftigung, dadurch man sich ein Ansehen von schöpferischem Genie gibt, indem man fordert, was man selbst nicht leisten, tadelt, was man doch nicht besser machen kann, und vorschlägt, wovon man selbst nicht weiß, wo es zu finden ist [...]“ (Immanuel Kant)*

### 3 Untersuchung von Staatsexamensentwürfen

Gewiss muss man kein schöpferisches Genie sein, wie Kant polemisch suggeriert, um planen zu können. Angesichts der Bedeutung, die insbesondere die Staatsexamensprüfung für angehende Lehrkräfte besitzt (z. B. König et al., 2015; Roloff et al., 2020), ist allerdings nicht von der Hand zu weisen, dass das Verfassen von Unterrichtsentwürfen allzu leicht zu eben jener „prahlerischen Geistesbeschäftigung“ verkommt, die den Unterrichtsentwurf vom Instrument, das im Dienste der für die Schülerinnen und Schüler viel wichtigeren Unterrichtsdurchführung steht, zum Selbstzweck erhebt. Folgt man der plausiblen und inzwischen auch empirisch belegbaren Annahme, dass es einen positiven Zusammenhang zwischen der professionellen Kompetenz von Lehrkräften, der Qualität des von ihnen gestalteten Unterrichts und dem Lernerfolg der Schülerinnen und Schüler gibt (Blömeke et al., 2015; König et al., 2021b; Rothland et al., 2018), hat sich auch die Unterrichtsplanung dergestalt unterzuordnen, dass sie zwar hochqualitativen Unterricht ermöglichen soll, diesen aufgrund der Kontingenz unterrichtlicher Prozesse aber niemals mit Gewissheit gewährleisten kann. Wie immer man auch Unterrichtsqualität fassen möchte<sup>13</sup>, ist das Verhältnis zwischen Planungsqualität und Durchführungsqualität empirisch bislang noch unterbelichtet und wird auch in den folgenden Beiträgen nicht berührt. Eher sollen mit Blick auf das Ziel dieses Dissertationsvorhabens, empirisch geprüfte Kriterien zur Beurteilung schriftlicher Unterrichtsentwürfe (Krepf & König, 2022b) zu entwickeln, allein die Planungsprodukte – in diesem Fall Staatsexamensentwürfe – berücksichtigt werden. Dazu werden auch Befunde der empirischen Unterrichtsqualitätsforschung miteinbezogen. Im Fokus steht jedoch die ökologisch valide Entwicklung solcher Kriterien, indem authentische Staatsexamensentwürfe analysiert werden. Aus der Art und Weise, wie in ihnen ePCK<sub>p</sub> genutzt wird, lassen sich in einem deduktiv-induktiven Verfahren derartige Kriterien herleiten. Staatsexamensentwürfe stellen für diese Zwecke das geeignete Datenmaterial dar, denn auch wenn Unterrichtsplanung als lebenslanger berufsbiographischer Prozess aufgefasst wird (Munthe & Conway, 2017), werden solche Unterrichtsentwürfe nur innerhalb der 1. und 2. Phase der Lehrkräftebildung und in der Regel nicht mehr von ausgebildeten Lehrkräften in der 3. Phase verlangt. Staatsexamensentwürfe sind also

---

<sup>13</sup> So ist insbesondere im deutschsprachigen Raum die aus der TIMSS-Studie hervorgegangene Unterteilung in die drei Basisdimensionen *Klassenführung*, *Kognitive Aktivierung* und *Konstruktive Unterstützung* (Klieme et al., 2001) weit verbreitet und ein positiver Zusammenhang zu effektivem Unterricht inzwischen empirisch gestützt (z. B. Kunter & Ewald, 2016, S. 15-18). Kantreiter (2022) weist ergänzend auch auf die zwölf Merkmale von Brophy und Good (1986), das QAIT-Modell von Slavin (1994) oder die zehn Kriterien guten Unterrichts von Helmke (2017) hin.

insofern geeignet, als aus ihnen diejenigen Qualitätskriterien abgeleitet werden können, die am Ende der deutschen Lehrkräftebildung im Rahmen der Staatsexamensprüfung zur Beurteilung herangezogen werden. Insbesondere in den Beiträgen 2 und 3 wird also zunächst der mögliche Zusammenhang zwischen der Vernetzung des ePCK<sub>p</sub> und der Qualität der Staatsexamensentwürfe untersucht.

#### **3.1. Zusammenhang zwischen fachdidaktischem Wissen und Unterrichtsplanung**

In zahlreichen naturwissenschaftsdidaktischen Studien wird die Vernetzung von PCK-Facetten analysiert (für eine Übersicht, siehe Chan & Hume, 2019; Şen et al., 2022), entweder um die Interaktion zwischen einzelnen PCK-Facetten untereinander zu explorieren (z. B. Aydin & Boz, 2013; Mavhunga, 2020; Park & Chen, 2012), einen möglichen Zusammenhang zwischen PCK und CK zu untersuchen (z. B. Kaya 2009) oder um die Entwicklung von PCK im Rahmen von Interventionen (z. B. Demirdöğen et al., 2016) nachzuvollziehen.

Ausgehend vom *Refined Consensus Model* (Carlson et al., 2019) betonen Alonzo et al. (2019) die Wechselwirkung zwischen ePCK<sub>p</sub>, ePCK<sub>t</sub> und ePCK<sub>r</sub> innerhalb des Zyklus aus Unterrichtsplanung, -durchführung und -reflexion (vgl. Scholl & Plöger, 2020). Von der Wechselwirkung abgesehen sind die drei Konstrukte selbst bislang kaum empirisch untersucht. Auch die Beleuchtung des Zusammenhangs zwischen cPCK/pPCK und Planung (ePCK<sub>p</sub>) steht noch am Anfang (Stender & Brückmann, 2020).

Ziel der Beiträge 2 und 3 war es, die Konstrukte *PCK* und *Unterrichtsplanungskompetenz* zusammenzuführen und über den Vergleich der analytischen, theoriegeleiteten Identifizierung von PCK-Facetten in den Entwürfen einerseits und dem holistischen Urteil von Fachseminarleiterinnen und Fachseminarleitern andererseits Merkmale von Unterrichtsentwürfen zu identifizieren, die die Qualität ausmachen. Das dazu genutzte Verfahren – der *PCK map approach* (Park & Suh, 2019) – ist in der internationalen PCK-Forschung inzwischen weit verbreitet (Chan, 2022) und wurde in der Biologiedidaktik erstmals und bislang einzig von Pohlmann (2019) im Rahmen einer Untersuchung der Wirksamkeit einer Lerngelegenheit zur Förderung von Bewertungskompetenz eingesetzt.

**3.1.1. Beitrag 2: *Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen***

Beitrag 2 wurde publiziert: Großmann, L. & Krüger, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen. *Erkenntnisweg Biologiedidaktik*, 19, 21–39.

**Zusammenfassung**

In diesem Beitrag wurden die folgenden Forschungsfragen untersucht:

- F1. Wie lässt sich auf der Grundlage des Pentagon-Modells ein Kategoriensystem entwickeln, das den Vernetzungsgrad des ePCK<sub>p</sub> von Biologie-Lehrkräften erfasst und eine valide Analyse der vorliegenden Staatsexamensentwürfe erlaubt?
- F2. Inwiefern ist das entwickelte Kategoriensystem geeignet, um das ePCK<sub>p</sub> von Lehrkräften in Staatsexamensentwürfen zu identifizieren?

Als methodischer Zugang für die Analyse der Staatsexamensentwürfe wurde die qualitative Inhaltsanalyse gewählt. Trotz des teilweisen Dissens über die verschiedenen Varianten dieser Forschungsmethodik<sup>14</sup>, lassen sich vor dem Hintergrund des Diskurses einige zentrale Wesensmerkmale der qualitativ-strukturierenden Inhaltsanalyse (Mayring, 2015) skizzieren, die bei der Entwicklung des Kategoriensystems berücksichtigt wurden:

Erstens wurde deduktiv ein Kategoriensystem entwickelt, das auf Basis des Pentagon-Modells (Park & Oliver, 2008) die fünf Oberkategorien *Orientation to Teaching Science* (OTS), *Knowledge of Science Curriculum* (KSC), *Knowledge of Students' Understanding in Science* (KSU), *Knowledge of Instructional Strategies for Teaching Science* (KISR) und *Knowledge of Assessment of Science Learning* (KAs) umfasst. Auf der Basis von theoretischer und empirischer Literatur zur Unterrichtsplanung wurden für alle fünf Facetten Subkategorien entwickelt, so dass das allgemeine Konstrukt PCK spezifisch auf die Unterrichtsplanung hin angepasst wurde (ePCK<sub>p</sub>).

Zweitens wurden induktiv nach Durchsicht der untersuchten Staatsexamensentwürfe Subkategorien ergänzt bzw. entfernt, um sicherzustellen, dass eine Passung zwischen dem Kategoriensystem und dem Datenmaterial gegeben ist. Nach der Lektüre von 16 Staatsexamensentwürfen wurde theoretische Sättigung (Krüger & Riemeier, 2014) erreicht, d. h. es wurden keine weiteren Aspekte im Datenmaterial gefunden, die als neue Subkategorien im

---

<sup>14</sup> Vergleiche zum Beispiel Stamann et al. (2016), Mayring (2019) und für eine Übersicht aus Perspektive der Naturwissenschaftsdidaktiken Göhner und Krell (2020).

Kategoriensystem hätten ergänzt werden müssen. Drei Subkategorien wurden nicht im Datenmaterial gefunden und daher entfernt.

Drittens wurden über die deduktiv-induktive Kategorienbildung hinaus weitere Maßnahmen der Qualitätssicherung ergriffen, um die Güte der qualitativen Inhaltsanalyse zu gewährleisten. Göhner und Krell (2020) nennen zahlreiche Maßnahmen, von denen die folgenden in Beitrag 2 berücksichtigt wurden: *Formale Aspekte des Kategoriensystems*: In einem Kodierleitfaden (siehe Anhang 1) wurde das Vorgehen bei der Inhaltsanalyse offengelegt und die Kodiereinheit definiert.

- *Intra-Rater-Übereinstimmung*: Der Erstkodierer hat die Daten im Abstand von 14 Tagen zwei Mal kodiert, um zu prüfen, ob das Auswertungsverfahren hinreichend stabil ist.
- *Inter-Rater-Übereinstimmung*: Die Daten wurden von einer zweiten Person kodiert, um mithilfe eines Übereinstimmungsmaßes (Cohen's  $\kappa$ ) zu prüfen, ob die Auswertung intersubjektiv nachvollziehbar ist.
- *Training der BeurteilerInnen*: Der Zweitkodierer wurde in mehreren Sitzungen geschult, das Kategoriensystem anzuwenden. Zu diesem Zwecke wurden Struktur und Inhalt von Staatsexamensentwürfen sowie die Kategorien samt Beispielen diskutiert und das Kodieren an zufällig ausgewählten Passagen geübt.
- *Dokumentation*: Während der Kodierung wurden Notizen angefertigt, in denen in Zweifelsfällen dargelegt wurde, warum eine bestimmte Textstelle wie kodiert worden ist. Diese Passagen konnten dann zwischen Erst- und Zweitkodierer diskutiert werden.
- *Diskussion der Kodierungen*: Erst- und Zweitkodierer haben unabhängig voneinander in MAXQDA2020 kodiert und Abweichungen in ihren Kodierungen anschließend diskutiert.
- *Konsensbildung*: Erst- und Zweitkodierer haben abweichende Kodierungen mit dem Ziel diskutiert, einen Konsens herzustellen und sich – falls nötig nach Re-Analyse der entsprechenden Textpassagen – gemeinsam auf eine Kodierung zu einigen.

Die Anwendbarkeit des auf diese Weise entwickelten Kategoriensystems wurde an einem exemplarischen Fall erfolgreich erprobt. Die Ergebnisse insbesondere der Inter- und Intra-Rater-Übereinstimmung weisen darauf hin, dass das Kategoriensystem zur reliablen und validen Identifikation von ePCK<sub>p</sub> in Biologie-Staatsexamensentwürfen geeignet ist.

**3.1.2. Beitrag 3: Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe?**

Beitrag 3 wurde publiziert: Großmann, L. & Krüger, D. (2022c). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, 28, 4. <https://doi.org/10.1007/s40573-022-00141-w>

**Zusammenfassung**

In diesem Beitrag wurden die folgenden Forschungsfragen untersucht:

- F1. Inwiefern unterscheidet sich die Häufigkeit von ePCK<sub>p</sub>-Facetten in elaboriert und wenig elaboriert eingeschätzten Entwürfen?
- F2. Inwiefern unterscheidet sich die Häufigkeit der Vernetzung von ePCK<sub>p</sub>-Facetten in elaboriert und wenig elaboriert eingeschätzten Entwürfen?
- F3. Inwiefern drückt sich in der Berücksichtigung von ePCK<sub>p</sub>-Facetten und in deren Vernetzung Planungsqualität aus?

Es ist anzumerken, dass sich der im Beitrag 3 und in F3 genutzte Begriff „Planungsqualität“ auf die Bewertung der Unterrichtsentwürfe durch den Fachseminarleiter in Form von Noten bezieht, während er sich oben und in den zusammenfassenden Diskussionen auf die Performanz bezieht und mit Unterrichtsplanungskompetenz gleichgesetzt wird.

Aufbauend auf dem in Beitrag 2 entwickelten Kategoriensystem wird in Beitrag 3 eine Brücke geschlagen zwischen dem theoretischen Teil der Ausbildung (1. Phase der Lehrkräftebildung), dem das zugrundeliegende PCK-Modell (Park & Oliver, 2008) entstammt, und dem praktischen Teil der Ausbildung (2. Phase der Lehrkräftebildung), in dem neben Ausbildungstraditionen an den jeweiligen Ausbildungsstandorten auch prüfungsrechtliche und andere bürokratische Gegebenheiten auf die Unterrichtspraxis der angehenden Lehrkräfte einwirken. Mithilfe des *PCK map approaches* (Park & Suh, 2019) wird der Versuch unternommen, Staatsexamensentwürfe sowohl aus theoretischer als auch aus unterrichtspraktischer Perspektive zu betrachten und den in der empirischen Forschung bislang kaum berücksichtigen Aspekt der Planungsqualität (Vogelsang & Riese, 2017) zu untersuchen. Zu diesem Zwecke wurden zwei Schritte unternommen:

1. Das Kategoriensystem wurde auf  $N = 46$  Staatsexamensentwürfe angewendet. Auf diese Weise wurde identifiziert, welche PCK-Elemente die Referendarinnen und Referendare genutzt haben, um ihre Unterrichtsplanungen zu entwickeln und zu begründen. Durch Identifikation von Textpassagen, in denen Subkategorien aus mindestens zwei der fünf



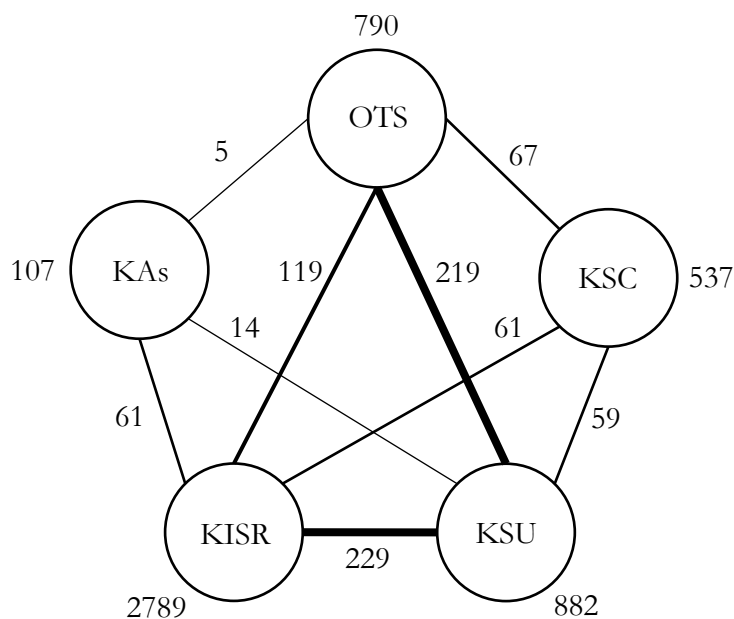
Oberkategorien (OTS, KSC, KSU, KISR, KAs) gemeinsam auftreten, konnte die Vernetzung zwischen den PCK-Elementen in Form von *PCK maps*<sup>15</sup> visualisiert werden.

- Ein erfahrener Biologie-Fachseminarleiter hat dieselben 46 Entwürfe entsprechend der Kriterien bewertet, die auch in Staatsexamensprüfungen im Land Berlin angewendet werden. Die Benotung der Entwürfe folgte dem in Deutschland üblichen universitären Benotungssystem (1.0, 1.3 – sehr gut; 1.7, 2.0, 2.3 – gut; 2.7, 3.0, 3.3 – befriedigend; 3.7, 4.0 – ausreichend; 4.3, 4.7, 5.0 – mangelhaft).

Die Kodierung der Staatsexamensentwürfe zeigt, in welchem Maße die verschiedenen ePCK<sub>p</sub>-Facetten überhaupt berücksichtigt werden. Abbildung 4 zeigt eine summierte *PCK map* für alle 46 in Beitrag 3 analysierten Staatsexamensentwürfe. Summiert bedeutet, dass auf der Ebene der fünf Oberkategorien alle Kodierungen der Einzelfacetten sowie der daraus resultierenden Verbindungen über die Gesamtstichprobe hinweg addiert wurden. Zwei Ergebnisse fallen auf: Erstens zeigen sich die stärksten Verbindungen in der Trias aus OTS, KSU und KISR, also in der

**Abbildung 4**

*Summierte PCK map der untersuchten Staatsexamensentwürfe (N = 46)*



*Anmerkung.* OTS = *Orientation to Teaching Science* (Kompetenzen); KSC = *Knowledge of Science Curriculum* (Fachliche Klärung); KSU = *Knowledge of Students' Understanding in Science* (Schülerkognition); KISR = *Knowledge of Instructional Strategies for Teaching Science* (Instruktionsstrategien); KAs = *Knowledge of Assessment of Science Learning* (Feedback). Die Liniendicke variiert je nach Anzahl der kodierten Verbindungen.

<sup>15</sup> In Ergänzung zu den in Beitrag 3 gezeigten *PCK maps* finden sich die *PCK maps* aller untersuchten Entwürfe (N = 46) in Anhang 2.

Verbindung zwischen der angestrebten Kompetenzentwicklung (OTS), dem Vorwissen und den bereits entwickelten Kompetenzen der Schülerinnen und Schüler (KSU) und den daraufhin ausgewählten Instruktionsstrategien (KISR).

Zweitens werden nur wenige Verbindungen zu KSC (Fachliche Klärung) und vor allem zu KAs (Feedback) hergestellt. Für die eigentliche Analyse in Beitrag 3 wurden die Noten des Fachseminarleiters als Kriterien genutzt, um für die Analyse die am besten bewerteten (oberes Quartil) mit den am schwächsten bewerteten Entwürfen (unteres Quartil) vergleichen zu können. Auf diese Weise wurde die Bedeutung der Einzelnoten insofern relativiert, als es unerheblich wurde, ob ein Staatsexamensentwurf mit einer 2.3 oder einer 2.7 bewertet wurde. Die beiden Teilgruppen (jeweils  $n = 11$ ) unterscheiden sich offensichtlich maßgeblich in ihrer Planungsqualität, zumindest aus Sicht des Fachseminarleiters. Prüft man nun, ob sich die beiden Gruppen hinsichtlich der Anzahl der PCK-Facetten (F1) oder der Verbindungen zwischen den PCK-Facetten (F2) unterscheiden, zeigt sich Folgendes: In den vom Fachseminarleiter als gut bewerteten Entwürfen werden im Median sämtliche PCK-Facetten sowie alle PCK-Verbindungen (abgesehen von den KAs-Verbindungen, die beide Gruppen nicht berücksichtigen) häufiger adressiert als in den Entwürfen, die der Fachseminarleiter als schwach bewertet hat. Allerdings sind die Ergebnisse des Mann-Whitney-*U*-Tests nach Anwendung des Dunn-Bonferroni-Posthoc-Tests als statistisch nicht signifikant anzusehen. Der sich daran schließende Fallvergleich (F3) zeigt folglich, dass es sowohl Entwürfe mit hoher Planungsqualität (gute Note) gibt, die eher schwach vernetztes ePCK<sub>p</sub> zeigen, als auch Entwürfe mit schwacher Planungsqualität (schlechte Note), die eher stark vernetztes ePCK<sub>p</sub> zeigen. Die Quantität der Verbindungen von ePCK<sub>p</sub>-Facetten allein scheint demzufolge nicht für die Planungsqualität in Form der Einschätzung des Fachseminarleiters entscheidend zu sein. Die dem *PCK map approach* im Besonderen und der Professionsforschung im Allgemeinen zugrunde liegende Annahme, dass die Anzahl der Vernetzungen von Wissensbeständen einen Qualitätsindikator darstellt (Abell, 2008; Chan et al., 2019; Hashweh, 2005; Park & Chen, 2012; Park & Oliver, 2008; Reynolds & Park, 2021; Schiering et al., 2023; Schneider & Plasman, 2011), konnte somit in Beitrag 3 für die Unterrichtsplanung nicht bestätigt werden. Mögliche Gründe dafür werden in der zusammenfassenden Diskussion erörtert.

### 3.2 Besonderheiten der Unterrichtsplanung im Fach Biologie

In der Analyse der 46 Staatsexamensentwürfe (Beitrag 3) konnten zwei Beobachtungen gemacht werden, die nicht Gegenstand der dort untersuchten Fragestellungen waren und daher dort nicht berichtet wurden, die aber im Rahmen einer biologiedidaktischen Untersuchung von

Unterrichtsentwürfen relevant sind und daher in zwei weitere Beiträge ausgelagert wurden. Beide Aspekte – die Berücksichtigung von Schülervorstellungen und die inhaltliche Strukturierung – werden von Neuhaus (2021) den eher biologiespezifischen Unterrichtsqualitätsmerkmalen zugeordnet. Die Analyse dieser eher fachdidaktisch ausgerichteten Merkmale ergänzt somit die eher generischen Überlegungen in den Beiträgen 2 und 3.

Erstens wurde festgestellt, dass Schülervorstellungen (als eine Subkategorie von KSU) von den Referendarinnen und Referendare in dieser Stichprobe kaum berücksichtigt worden sind. Dies erstaunt, weil es im Sinne der konstruktivistischen Lerntheorie (vgl. Riemeier, 2007) inzwischen zahlreiche theoretische (z. B. Modell der Didaktischen Rekonstruktion, Duit et al., 2012) und empirische Befunde (Pacaci et al., 2023) gibt, die das Potential der Berücksichtigung von Schülervorstellungen in Lehr-Lern-Prozessen aufzeigen. Zudem sind *expressis verbis* an Lehrkräfte gerichtete Publikationen verfügbar, die den Versuch unternehmen, die Forschungsbefunde in die Unterrichtspraxis zu überführen und Lehrkräfte dabei zu unterstützen, Schülervorstellungen in ihrem Unterricht zu berücksichtigen (z. B. Hammann & Asshoff, 2014; Kattmann, 2015). Die kognitive Herausforderung, die dabei zu bewältigen ist, ist laut dem *CODE-PLAN model* (König et al., 2021a) die *adaptation to learning dispositions*. Eine systematische Untersuchung zum Umgang mit Schülervorstellungen und zu ihrem Potential in der Planung von Biologieunterricht wird daher in Beitrag 4 dargestellt.

Zweitens zeigte sich, dass die Umsetzung des angestrebten Kompetenzzuwachses oft insofern eine Herausforderung darstellte, als eine mangelnde Passung zwischen dem Lernziel (OTS) und den zur Förderung ausgewählten Instruktionstrategien (KISR, z. B. Aufgaben, Phasenstruktur) bestand. Dies zeigte sich in Staatsexamensentwürfen für sämtliche vier Kompetenzbereiche, insbesondere aber für den Kompetenzbereich Erkenntnisgewinnung. Wiederholt wurden Lernziele auf der Basis von Standards aus dem Kompetenzbereich Erkenntnisgewinnung formuliert, die letztlich eher zum Erwerb neuen Fachwissens als zur Förderung prozessbezogener Kompetenzen im Bereich Erkenntnisgewinnung führten. Diese Schwierigkeit kann aus der irreführenden Benennung des Kompetenzbereichs Erkenntnisgewinnung (KMK, 2005; KMK, 2020) resultieren, die den Eindruck erweckt, als sollte biologisches Fachwissen mithilfe praktischer Arbeitsweisen (z. B. dem Experimentieren) gewonnen werden (Krell & Krüger, 2022). Dass die Planung von Stunden zur Förderung wissenschaftsmethodischer Kompetenzen auf Ebene der *practical skills*, des *scientific reasoning* sowie der *epistemological views* (Mayer, 2007) eine Herausforderung darstellt, ist bereits mehrfach gezeigt worden (z. B. Capps & Crawford, 2013; Gyllenpalm & Wickman, 2011). Die kognitive Herausforderung, die dabei zu bewältigen ist, ist laut dem *CODE-PLAN model* (König et al., 2021a) das *phasing* (d. h. die Entwicklung einer zielführenden Phasenstruktur). Eine explizit

biologiedidaktische Perspektive auf die Unterrichtsplanung zum Kompetenzbereich Erkenntnisgewinnung hat bislang einzig Tardent Kuster (2020) vorgenommen, die eine positive Korrelation zwischen dem PCK, gemessen mithilfe eines Fragebogens, und der Qualität schriftlicher Unterrichtsentwürfe zum Experimentieren gefunden hat. Vor diesem Hintergrund wird in Beitrag 5 der Frage nachgegangen, wie angehende Biologielehrkräfte Stunden zur Förderung von Erkenntnisgewinnungskompetenzen planen. Der Fokus liegt dabei auf dem Umgang mit Hypothesen in der Entwicklung einer Phasenstruktur und somit auf der inhaltlichen Strukturierung von Biologieunterricht.

#### **3.2.1. Beitrag 4: *Students' conceptions as a neglected perspective in trainee teachers' biology lesson plans***

Beitrag 4 wurde publiziert: Großmann, L. & Krüger, D. (2022b). Students' Conceptions as a Neglected Perspective in Trainee Teachers' Biology Lesson Plans. In K. Korfiatis & M. Grace (Hrsg.), *Current Research in Biology Education. Contributions from Biology Education Research* (S. 181-193). ERIDOB, Springer. [https://doi.org/10.1007/978-3-030-89480-1\\_14](https://doi.org/10.1007/978-3-030-89480-1_14)

#### **Zusammenfassung**

In diesem Beitrag wurden die folgenden Forschungsfragen untersucht:

- F1. Which facets of knowledge about *Students' Understanding in Science* as part of their ePCK<sub>p</sub> are most taken into account by German trainee teachers in written biology lesson plans?
- F2. To what extent do German trainee teachers align their planned instructional strategies in biology lessons to their analysis of students' conceptions?

Da in den Staatsexamensentwürfen, die in Beitrag 3 untersucht wurden ( $N = 46$ ), nur zwei Mal Schülervorstellungen kodiert worden sind, wurde die Stichprobe für Beitrag 4 erweitert: Sämtliche Staatsexamensentwürfe, die im Rahmen dieses Dissertationsprojekts von der Berliner Senatsverwaltung für Bildung, Jugend und Familie in anonymisierter Form zur Verfügung gestellt worden sind ( $N = 107$ ), wurden hinsichtlich der ePCK<sub>p</sub>-Facette KSU untersucht. Induktiv wurden sechs Subkategorien genutzt, die die Referendarinnen und Referendare in ihren Staatsexamensentwürfen zur Beschreibung ihrer Lerngruppen genutzt haben: Vorwissen, Schülervorstellungen, Lernschwierigkeiten, methodische Kenntnisse, Motivation/Interesse und Sozialverhalten. Quantitativ zeigt sich, dass in nahezu allen Entwürfen auf das Vorwissen der Schülerinnen und Schüler eingegangen wird ( $n = 106$ ), jedoch nur in sehr wenigen auf Schülervorstellungen ( $n = 6$ ). Qualitativ zeigt sich, dass drei der sechs Entwürfe, in denen Schülervorstellungen kodiert wurden, auf Spekulationen basieren, wie die

Schülerinnen und Schüler ein bestimmtes Thema verstehen werden. Zwei der sechs Entwürfe ziehen empirische Befunde zu Schülervorstellungen aus der Literatur heran, allerdings ohne dass dies erkennbare Auswirkungen auf die geplante Stunde gehabt hat. Nur in einem einzigen Entwurf werden die Vorstellungen der Schülerinnen und Schüler erhoben und mit ihnen explizit gearbeitet, allerdings auch hier nicht als klassischer *conceptual change*-Ansatz (vgl. Krüger, 2007).

#### 3.2.2. Beitrag 5: *Erkenntnisgewinnung (v)erklärt: Wie plant man hypothesengeleiteten Biologieunterricht?*

Beitrag 5 wurde noch nicht publiziert, ist aber bereits angenommen: Großmann, L. & Krüger, D. (in Druck). Erkenntnisgewinnung (v)erklärt. Wie plant man hypothesengeleiteten Biologieunterricht? In: B. Reinisch, D. Krüger & D. Mahler (Hrsg.), *Biologiedidaktische Nature-of-Science-Forschung: Zukunftsweisende Praxis*. Berlin: Springer.

#### **Zusammenfassung**

In diesem Beitrag wurden die folgenden beiden Thesen diskutiert:

T1. Der Begriff *Hypothese* wird im Biologieunterricht häufig missverständlich verwendet, denn statt als epistemisches Konzept im Sinne einer Voraussage eines Untersuchungsergebnisses (Krell & Krüger, 2022) wird er zumeist alltagssprachlich im Sinne eines vagen, initialen Erklärungsversuchs verwendet. Daher sollten die Begriffe *Hypothese* und *Erklärung* bzw. *Erklärungsversuch* voneinander unterschieden werden.

T2. Es sollen explizit prozedurale und epistemische Aspekte des Begriffs *Hypothese* reflektiert werden, um Kompetenzen im Bereich *Erkenntnisgewinnung* zu fördern (*learning about science*, Hodson, 2014).

In der Analyse der Staatsexamensentwürfe (Beitrag 3) wurde festgestellt, dass viele Unterrichtsentwürfe, in denen als Ziel die Förderung von Erkenntnisgewinnungskompetenzen angegeben ist, eher Fachwissen bzw. Sachkompetenz adressieren. In Beitrag 5 wird untersucht, welche Rolle der Begriff „Hypothese“ für die inhaltliche Strukturierung von Biologieunterricht spielt.

Ausgangspunkt der Analyse war die Beobachtung, dass in vielen der geplanten Erkenntnisgewinnungsstunden zu Beginn sogenannte Hypothesen formuliert werden, die bei näherer Betrachtung jedoch keine Hypothesen sind (T1). Es werden also zwei verschiedene Bedeutungen des Begriffs *Hypothese* genutzt, von denen die erste im Sinne problemorientierten,

schüleraktivierenden Unterrichts eher zum Erwerb neuen Fachwissens führt, während nur die zweite dem wissenschaftlich adäquaten Begriffsverständnis entspricht und damit den Kompetenzbereich Erkenntnisgewinnung adressiert:

1. Alltagssprachlich versteht man unter einer Hypothese eine vage, spontan geäußerte Idee, die einen Sachverhalt möglicherweise korrekt, möglicherweise aber auch inkorrekt erklären kann (Kattmann, 2015). Eine solche Aussage ist das Ergebnis einer logischen Schlussweise – der Abduktion –, bei der ausgehend von einem erklärungsbedürftigen Phänomen und unter Nutzung bereits vorhandenen Wissens die bestmögliche Erklärung des Phänomens generiert (*inference to the best explanation*; Harman, 1965) bzw. eine plausible Erklärung erdacht wird (*educated guess*; Clement, 2008). Die dabei entweder theoriegeleitet entwickelten oder kreativ erdachten Erklärungsansätze werden hinsichtlich ihrer Plausibilität und ihrem Potential, die *bestmögliche* Erklärung zu liefern, geprüft und angesichts neu erhobener Daten verändert oder ergänzt (Johnson & Kreams, 2001).
2. Wissenschaftstheoretisch sind Hypothesen „*empirisch überprüfbar, grundsätzlich widerlegbar, eindeutig und widerspruchsfrei [und] werden stets durch Vorwissen bzw. theoretische Konstrukte (z. B. Theorien, Regeln, Modelle) begründet*“ (Krell & Krüger, 2022, S. 381). Es handelt sich um Voraussagen, die mithilfe von Beobachtungen, Vergleichen, Experimenten oder Modellierungen überprüft werden können und die angesichts der dabei erhobenen Daten entweder bestätigt oder widerlegt werden. Eine solche Aussage ist das Ergebnis einer logischen Schlussweise – der Deduktion –, bei der ausgehend von einer Theorie oder einem Modell auf einen konkreten Einzelfall geschlossen und somit ein Ereignis vorhergesagt wird. Auf dem Wege der Falsifikation werden Hypothesen, die durch die erhobenen Daten nicht bestätigt werden, verworfen, so dass auf diese Weise Phänomene untersucht werden können (Popper, 1934/2005).

Keineswegs wird hier der Standpunkt vertreten, dass das Deduzieren einer Voraussage aus einem Modell, einer Theorie o.ä. eine höherwertige kognitive Tätigkeit sei, der daher der Begriff *Hypothese* zustehe, wohingegen das abduktive Entwickeln spontaner Erklärungsversuche minderwertig sei und daher mit dem Begriff *Erklärung* auskommen müsse. Krüger und Upmeier zu Belzen (2021) haben die wissenschaftstheoretischen Betrachtungen des abduktiven Schlussfolgerns unlängst in ihr Kompetenzmodell der Modellierkompetenz integriert und damit den ersten Versuch unternommen, die Unterscheidung zwischen abduktiven Erklärungen und deduktiven Hypothesen am Beispiel der Arbeitsweise Modellieren zu illustrieren. Angesichts der anspruchsvollen Denkprozesse, die beim abduktiven Schlussfolgern durchlaufen werden müssen (Johnson & Kreams, 2001), ist es nachvollziehbar, dass im Vergleich zum alten Kompetenzmodell der

Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) im neuen Kompetenzmodell der Modellierkompetenz (Krüger & Upmeier zu Belzen, 2021) das abduktive Erklären als neue Niveaustufe 3a eingeführt wurde, die dem deduktiven Ableiten einer Hypothese auf Niveaustufe 3b hierarchisch nicht untergeordnet ist. Beide Tätigkeiten sind für sich genommen anspruchsvoll und sollten – so die Argumentation in Beitrag 5 – als zu erwerbende Kompetenzen anerkannt werden. Deutlich wird dies aber erst dann, wenn man beide Prozesse nicht gleichermaßen unter dem Begriff *Hypothese*<sup>16</sup> subsumiert, sondern sie begrifflich differenziert.

In Beitrag 5 wird vor diesem Hintergrund der Unterschied zwischen Stunden, in denen tatsächlich Erkenntnisgewinnungskompetenzen mittels einer deduktiv entwickelten Hypothese gefördert werden, und solchen, in denen im Einstieg als Hypothesen bezeichnete abduktive Erklärungen generiert werden, die letztlich in den Erwerb von Fachwissen münden, anhand je eines exemplarischen Staatsexamensentwurfs illustriert (T2).

Im Biologieunterricht gibt es bislang erst einen Versuch (Engelschalt et al., 2022), die Unterscheidung zwischen den drei Schlussweisen Induzieren, Deduzieren und Abduzieren zur Förderung von Erkenntnisgewinnungskompetenzen in den Biologieunterricht zu integrieren; ein weiterer Unterrichtsvorschlag wird in Beitrag 14 unterbreitet.

### 3.3. Zusammenfassende Diskussion

In diesem Abschnitt werden die Beiträge 2, 3, 4 und 5 gemeinsam diskutiert. Dabei werden zunächst wesentliche Limitationen der Untersuchungen erörtert, ehe der Erkenntnisgewinn der vier Beiträge kritisch diskutiert und die Befunde an den Unterrichtsplanungsdiskurs angebunden werden.

#### *Limitationen*

Schriftliche Unterrichtsentwürfe gelten als „*Artefakte*“ (Seel, 2011, S. 40; siehe auch Koberstein-Schwarz & Meisert, 2022), die zwar im Zuge der Lehrkräftebildung nach den jeweiligen inhaltlichen und formalen Vorgaben der Ausbildungsstandorte von angehenden Lehrkräften verfasst werden müssen, die aber keineswegs der Planungsdokumentation entsprechen, wie sie ausgebildete Lehrkräfte im Berufsalltag vornehmen. Die wenigen Studien, die zum zeitlichen Aufwand der Unterrichtsplanung vorliegen, zeigen recht einheitlich, dass pro durchzuführender Stunde jeweils

---

<sup>16</sup> Dabei ist zu bedenken, dass diesem Vorschlag bereits eine Didaktisierung innewohnt, denn in der wissenschaftstheoretischen Literatur werden auch auf dem Wege der Abduktion generierte Erklärungen umstandslos *Hypothesen* genannt (z. B. Johnson & Krems, 2001, S. 905; Oh, 2019, S. 733; Schurz, 2008, S. 205). Zum Zwecke der Lehrkräftebildung sowie eines kompetenzorientierten Biologieunterrichts erscheint die begriffliche Auftrennung dennoch sinnvoll.

ca. 20-30 Minuten Planungszeit erforderlich sind (Engelhardt, 1982; Haas, 1998; Hübner & Werle, 1997; Oehlschläger, 1978; Wengert, 1989). Das ist gewiss ein deutlich geringerer zeitlicher Aufwand, als er für das Verfassen eines schriftlichen Unterrichtsentwurfs zu bewältigen ist. Darüber hinaus handelt es sich bei einem Unterrichtsentwurf neben der geplanten Stunde, die auch Lehrkräfte im Alltag – dann vorwiegend mental – entwickeln (*mentaler Unterrichtsentwurf*, Abb. 1), um die ausführlichen Analysen (z. B. der Sachstruktur des Themas, der Lerngruppenvoraussetzungen) und didaktischen Begründungen der getroffenen Planungsentscheidungen. Die Attribuierung von Unterrichtsentwürfen als „Artefakt“ ist somit gewiss zutreffend. Andererseits spielen Unterrichtsentwürfe zweifelsohne eine wichtige Rolle in der Lehrkräftebildung im Rahmen von Abschlussprüfungen (z. B. in Berlin, VSLVO, 2022/05.08.2022). Sie zum Gegenstand der empirischen Forschung zu erheben resultiert also auch aus Rahmenbedingungen der deutschen Lehrkräftebildung, denen sich empirisch forschende Fachdidaktiken schwerlich entziehen können. Folgerichtig wurden in den vergangenen Jahren verschiedene Forschungsvorhaben umgesetzt, in denen schriftliche Unterrichtsentwürfe – wenngleich mit verschiedenen Zielen und daher aus verschiedenen Perspektiven – analysiert wurden (fachübergreifend: Weingarten, 2019; Deutsch: Fladung, 2022; Geschichte: Wolf, 2022; Physik: Stender, 2014; Sachunterricht: Kantreiter, 2022). Zudem zeigen Roloff et al. (2020), dass die Note der zweiten Staatsexamensprüfung – anders als die Note zum ersten Staatsexamen – gar als Prädiktor für die Unterrichtsqualität von Lehrkräften im späteren Berufsleben (d. h. in der 3. Phase der Lehrkräftebildung) gelten kann. Sie erklären den Befund damit, dass die 1. Phase der Lehrkräftebildung vor allem deklaratives/konzeptuelles Wissen adressiert, während die 2. Phase der Lehrkräftebildung die Fähigkeit fokussiert, das in der Universität erworbene Wissen zu vernetzen und anzuwenden. Somit lasse sich die Note zum zweiten Staatsexamen als Indikator für das prozedurale Wissen angehender Lehrkräfte auffassen (Roloff et al., 2020, S. 12). Vor diesem Hintergrund erscheint die Analyse solcher Artefakte wertvoll für die empirische Professionsforschung zu sein, als sie vertiefte Einblicke in das prozedurale Wissen gewähren, das angehende Lehrkräfte am Ende der sechs- bis siebenjährigen Ausbildung in Deutschland erworben haben und das – wie in Beitrag 1 gezeigt wurde – bislang noch wenig in der empirischen naturwissenschaftsdidaktischen Forschung berücksichtigt wird.

Diesem grundsätzlichen Einwand gegen das hier dargestellte Forschungsvorhaben lässt sich ein spezifischer Einwand hinzufügen, der sich aus der PCK-Forschung ergibt: PCK wird mit vielfältigen Methoden erhoben und entsprechend werden verschiedene Datenquellen genutzt, unter anderem Unterrichtsentwürfe (z. B. Bergqvist & Chang Rundgren, 2017; Friedrichsen et al., 2008; Rollnick, 2017; Roth et al., 2011). Allerdings zeigen Chan und Hume (2019) in ihrem



systematischen Literaturreview zur PCK-Forschung, dass sich keine einzige der untersuchten Studien allein auf Artefakte wie Unterrichtsentwürfe als Datenmaterial stützte, sondern stets zusätzliche Daten erhoben wurden (z. B. Videographie des Unterrichts; Interviews zur geplanten Stunde). Der Vorteil einer solchen Triangulation besteht darin, den Forschungsgegenstand aus verschiedenen Blickwinkeln oder mit verschiedenen Methoden untersuchen und über den Vergleich der jeweiligen Ergebnisse tiefere Einsichten in die Daten erhalten zu können. Dies würde nicht bloß eine weitere Maßnahme zur Qualitätssicherung der qualitativen Inhaltsanalyse darstellen (Göhner & Krell, 2020), sondern wäre im Rahmen des *PCK map approaches* (Park & Oliver, 2008; Park & Suh, 2019) zwingend notwendig: In seinem Literaturreview zu international publizierten Studien, die den *PCK map approach* genutzt haben, zeigt Chan (2022), dass die Methodik oft nicht so rigoros umgesetzt wird wie im Original von Park und Oliver (2008) bzw. Park und Suh (2019) beschrieben. Neben der Nutzung verschiedener Datenquellen, um triangulieren zu können, gehört dazu im Wesentlichen der Fokus auf die Unterrichtssituation, also das ePCK<sub>i</sub>, das durch Unterrichtsvideographie (z. B. Bravo & Cofré, 2016), Unterrichtsaudiographie (z. B. Ekiz-Kiran & Boz, 2020) oder durch Unterrichtsprotokolle (z. B. Norville & Park, 2021) erfasst werden könnte. Während die Datentriangulation jedoch in der Tat insofern als Qualitätskriterium aufgefasst werden kann, als sie die Validität der Dateninterpretation betrifft, lässt sich das Plädoyer für einen Einsatz des *PCK map approaches* anhand von Daten der Unterrichtsdurchführung statt der -planung schwerlich als Indikator für die Studienqualität auffassen. Eher scheint Chan (2022) in seiner Argumentation die vergleichsweise intensive Berücksichtigung der Unterrichtsplanung und die häufige Vernachlässigung der Unterrichtsdurchführung in der PCK-Forschung (Chan & Hume, 2019) im Blick zu haben. Nichtsdestotrotz lässt sich der *PCK map approach* in gleicher Weise auf schriftliche Unterrichtsentwürfe anwenden und das darin enthaltene ePCK<sub>p</sub> erfassen. Da die den Beiträgen 2-5 zugrundeliegenden Forschungsfragen allein auf die Unterrichtsplanung Bezug nehmen und den Zusammenhang zwischen ePCK<sub>p</sub> und ePCK<sub>i</sub> im Sinne prognostischer Validität nicht in den Blick nehmen, ist die fehlende Möglichkeit, die Stundendurchführungen der 46 Referendarinnen und Referendare mit zu analysieren, kein hinreichender Einwand gegen die vorliegenden Arbeiten.

Plausibel erscheint hingegen Chans (2022) Vorschlag, den *PCK map approach* dahingehend zu modifizieren, dass mithilfe einer Ratingskala oder eines Kriterienrasters die Qualität der kodierten PCK-Episoden bewertet wird. Das würde dazu führen, dass neben der zunächst rein quantitativen Darstellung des ePCK<sub>p</sub> in Form der *PCK maps* auch die Qualität der jeweiligen Verbindungen berücksichtigt werden würde. Hierin läge eine potentielle Erklärung für die nicht signifikanten Unterschiede in der Anzahl der berücksichtigten PCK-Facetten und der Anzahl der Verbindungen

zwischen den vom Fachseminarleiter als sehr gut und den als weniger gut bewerteten Staatsexamensentwürfen in Beitrag 3. Kodiert wurde hier jeweils nur das Auftreten von PCK-Elementen. Ganz gleich, ob beispielsweise eine methodische Entscheidung (KISR) geeignet oder nicht ist, um die Kompetenzentwicklung der SuS zu ermöglichen (KSU) – eine solche Episode würde in jedem Fall als eine Verbindung zwischen KISR und KSU in den *PCK maps* visualisiert sein. Eine fruchtbare Möglichkeit, die Qualität von Planungsentscheidungen zu evaluieren, könnte die Unterscheidung zwischen Oberflächen- und Tiefenstrukturen (Decristan et al., 2020; Koberstein-Schwarz & Meisert, 2022; Oser & Baeriswyl, 2001) bieten: Während Beschreibungen und Begründungen auf Oberflächenstrukturebene die methodische Gestaltung des Unterrichts adressieren, würden Begründungen für diese Gestaltung auf Tiefenstrukturebene die kognitiven Prozesse der Schülerinnen und Schüler in den Blick nehmen, was eine anspruchsvolle Analyseleistung darstellt. Die Qualität der Ausführungen in diesem Sinne zu unterscheiden, war jedoch nicht das Ziel dieser Arbeit. Zudem würde eine derartige Qualitätseinschätzung statt der inhaltlich-strukturierenden Inhaltsanalyse eine evaluative Inhaltsanalyse (Schreier, 2014) notwendig machen, die jedoch aufwendiger wäre, weil zunächst zu prüfen wäre, ob die beiden Kodierenden und Kodierer selbst in der Lage sind, in den manchmal nur wenige Zeilen langen PCK-Episoden reliabel Oberflächen- und Tiefenstrukturen zu erkennen und entsprechend zu bewerten. Auf diese Weise ließe sich der Frage begegnen, inwiefern Unterrichtsentwürfe, die im Rahmen der Staatsexamensprüfung verfasst worden sind, überhaupt ein geeignetes Datenmaterial darstellen: Die nicht signifikanten Unterschiede zwischen den beiden in Beitrag 3 untersuchten Gruppen könnte auch damit zusammenhängen, dass Referendarinnen und Referendare aufgrund der großen Bedeutung der Staatsexamensprüfung verleitet sind, besonders extensiv zu planen, ausführlich den Lerngegenstand, den Kompetenzstand der Schülerinnen und Schüler und die getroffenen Planungsentscheidungen zu analysieren und zu begründen. Dies führt dazu, dass eine Vielzahl von PCK-Facetten und -Verbindungen kodiert wurde, die aber oft gar nicht zielführend sein müssen und somit weniger das Ergebnis fachdidaktischer als prüfungsstrategischer Überlegungen sind. Hier würde also eher eine Analyse der Qualität der Verbindungen sinnvoll sein, als bloß die Anzahl der Verbindungen zu zählen.

Hätte man Chan (2022) folgend mit dem *PCK mapping* darauf abgezielt, die Qualität unmittelbar aus den kodierten PCK-Facetten und somit ohne Hinzuziehung eines externen Fachseminarleiters zu bewerten, wäre die von König et al. (2020b) vorgenommene Unterscheidung in generische und fachspezifische Planungsentscheidungen tragfähig: Auch wenn *ad hoc*- und *post hoc*-entwickelte Begründungen aus dem Datenmaterial ohne kommunikative Validierung nicht unterschieden werden können, so wäre es denkbar, als weitere, übergeordnete Ebene im Kategoriensystem

(Beitrag 2) die Unterscheidung in eher generische und eher biologiespezifische Entscheidungen (Neuhaus, 2021) einzuführen. König et al. (2020b) zeigen am Beispiel deutscher Staatsexamensentwürfe, dass es Referendarinnen und Referendaren signifikant leichter fällt, generische Entscheidungen zu treffen und zu begründen (z. B. dass Einstiege motivierend sein sollen). Fachspezifische Entscheidungen (z. B. welche Teilaspekte des Themas didaktisch reduziert werden können) sind offenbar herausfordernder, da hier eine vertiefte Durchdringung des Fachinhalts und der Lehr-Lern-Situation notwendig ist, die verschiedene Wissensressourcen gleichermaßen erfordert (CK, PCK). Im Sinne des Konstrukts PCK, auf das diese Arbeit abzielt, läge hierin neben der Unterscheidung in Oberflächen- und Tiefenstrukturen eine weitere fruchtbare Unterscheidung, die in der Analyse schriftlicher Unterrichtsentwürfe zukünftig beachtet werden könnte.

Ein weiterer Einwand gegen die in Beitrag 3 vorgenommene Analyse (und damit auch gegen das in den Beiträgen 6, 7 und 8 entwickelte und überprüfte Kriterienraster) bezieht sich auf die Fachspezifität: In der aktuellen PCK-Forschung herrscht weitgehend Konsens, dass PCK ein domänen-, disziplin-, themen- und konzeptspezifisches Konstrukt darstellt (z. B. Carlson et al., 2019; Chan & Hume, 2019), wobei diese Kategorien oft nicht hinreichend trennscharf voneinander unterschieden werden. Unklar ist zudem beispielsweise auch, warum in Instrumenten wie dem PCK-IBI (Großschedl et al., 2019) Items zu Themen wie dem Herz-Kreislauf-System, zur Photosynthese und zur DNA gleichermaßen enthalten sind, obwohl die Autorinnen und Autoren PCK explizit als „topic-specific“ (S. 408) definieren. Da es im Rahmen dieser Dissertation nicht möglich war, ausschließlich Unterrichtsentwürfe zu einem bestimmten Kompetenzbereich und Thema zu analysieren, ist die Kodierung und wie im Folgenden zu zeigen sein wird die Entwicklung des Kriterienrasters eher generischer Natur. Die Auffassung von PCK, die dieser Arbeit zugrunde liegt, lässt sich am ehesten mit dem von Veal und MaKinster (1999) eingeführten Begriff „general PCK“ (S. 8) beschreiben, das noch nicht so spezifisch ausgerichtet ist wie das themen- und das konzeptspezifische PCK. Damit schließt diese Arbeit an Şen (2023) an, der für die PCK-Forschung die Unterscheidung zwischen PK-Codes und PCK-Codes wie folgt beschreibt:

*By way of a solution for the overlap of PK and PCK components, researchers can benefit from the topic-specific aspect of PCK to distinguish PK and PCK codes. Researchers can ask the teachers about their plans before the lesson and listen to their planned instructional strategies and assessment methods. If their explanations include topic-specific activities and representations, they can be coded under KOIS [hier: KISR, L.G.]. On the other hand, general explanations without topic-specific examples can be coded as part of PK. The same is also true for coding KOA [hier: KAs, L.G.] and the assessment aspect of PK. [...] Hence, researchers can use the topic-specific aspect of PCK to decide whether the code in question belongs to PCK components or general pedagogical knowledge. It should also be noted that teachers can use general assessment methods such as multiple-choice questions and concept cartoons in their lessons. As the content of such assessments includes the topic being taught, such general assessment methods represent the topic-specific aspect of PCK and are part of ePCK. Therefore, the general assessment methods used in class can also be labeled KOA. (Şen, 2023, S. 7f.)*

Obwohl die in Beitrag 3 kodierten Passagen oft auch generische Planungsentscheidungen betreffen, sind diese stets gebunden an das Fach Biologie, so dass es plausibel ist, diese Passagen nicht als PK, sondern als PCK kodiert zu haben.

Da die in den Beiträgen 2-5 untersuchten Staatsexamensentwürfe nicht unter kontrollierten Bedingungen verfasst worden sind, sondern die Referendarinnen und Referendare Tage oder Wochen vor ihren Prüfungsterminen mit der Planung ihrer Staatsexamensstunde begonnen haben können, ist eine konstruktbezogene Unschärfe nicht auszuschließen: Basierend auf dem RCM (Carlson et al., 2019) wird in Beitrag 2 und 3 argumentiert, dass ePCK<sub>p</sub> erfasst wurde, also der Teil des persönlich verfügbaren PCK (pPCK), der im konkreten Planungsprozess genutzt wird. Zwei Einwände ließen sich gegen diese Behauptung hervorbringen: Erstens ist es schwerlich möglich, ohne Datentriangulation oder kommunikative Validierung (Göhner & Krell, 2020) im geschriebenen Text festzustellen, welche Passagen tatsächlich konkret im Planungsprozess genutztes ePCK<sub>p</sub> und welche allgemeineres pPCK sind. Zu letzterem ließen sich wie in Beitrag 1 geschildert auch Planungsstrategien zählen, die aber gar nicht eigens erfasst wurden und zum Teil auch schwer zu identifizieren wären, da sie in Staatsexamensentwürfen in der Regel nicht explizit formuliert werden. Auch die Unterscheidung zum cPCK ist nicht immer eindeutig, da nicht auszuschließen ist, dass die Referendarinnen und Referendare ihre Entwürfe vorab von anderen Personen (z. B. Mentorinnen und Mentoren) haben Korrektur lesen lassen. Auf diesem Wege könnte cPCK Eingang in die Entwürfe gefunden haben, über das die Referendarinnen und Referendare selbst gar nicht verfügen. Komplizierter wird es, wenn man annimmt, dass cPCK nicht bloß als das in einer Community (z. B. von Fachseminarleitungen, von universitären Fachdidaktikerinnen und Fachdidaktikern) gesicherte, anerkannte Wissen darstellt, sondern cPCK individuell verfügbar ist (Irmer et al., 2023):

*cPCK can be possessed by a single person, meaning that all general knowledge a person possesses that has also been published and discussed can be referred to as his/her cPCK. In our way of understanding, it is, in contrast to a person's pPCK, not necessarily influenced by their personal experiences and could be acquired, e.g., by taking part in a university lecture. (Irmer et al., 2023, S. 3)*

Dies widerspricht allerdings dem RCM (Carlson et al., 2019) und ist derzeit Gegenstand einer sich erst noch formierenden empirischen Forschung, die die Zusammenhänge zwischen ePCK, pPCK und cPCK sowie die jeweils dazwischen wirkenden *amplifiers and filters* untersucht (Doppelpfeile in Abb. 2; Behling et al., 2022). Für die Kodierungen in den Beiträgen 2 und 3 hätte das beispielsweise folgende Bedeutung: Wenn es zutrifft, dass die Referendarinnen und Referendare im Sinne des symbolischen Interaktionismus im Rahmen der Staatsexamensprüfung gar nicht so planen, wie sie normalerweise planen würden und könnten, sondern ihre Entwürfe so schreiben, wie sie

annehmen, dass es ihre Prüfungskommission gutheißen wird, dann ließe sich unter anderem die Länge der Entwürfe plausibel erklären: In zahlreichen Entwürfen wurden Planungsentscheidungen in großer Ausführlichkeit beschrieben und begründet. Der Heuristik von Vogelsang und Riese (2017) folgend lässt sich Unterrichtsplanung als iterativer Prozess aus Kreation und Legitimation auffassen (Abb. 1; Abb. 3). Demnach seien kreierte Ideen einer mentalen Überprüfung unterworfen, die dazu führen kann, dass Ideen wieder verworfen und neue entwickelt werden. Das Kriterium ist im Idealfall die Funktionalität (Beitrag 3), also die Frage, inwiefern eine Idee dazu beiträgt, dass die Schülerinnen und Schüler die angestrebte Kompetenzentwicklung erfahren. In diesem Sinne wäre die geplante Stunde inklusive der gewählten Methoden, Aufgaben usw. *ad hoc* legitimiert. In den Staatsexamensentwürfen hingegen lässt sich jedoch oft nicht feststellen, ob Planungsentscheidungen nicht eher von einer spannenden Methode, die eingesetzt werden soll, oder vom Thema, das als nächstes unterrichtet werden muss, ausgehend getroffen werden. Im Sinne von Weingartens (2019) Einwand bezüglich des symbolischen Interaktionismus wäre es schlüssig, wenn die z. T. sehr elaborierten Begründungen für derartige Entscheidungen nicht aus der Sache heraus, sondern zur Zufriedenstellung der Prüfungskommission *post hoc* entwickelt werden. Solche Begründungen sind dann jedoch allein dem Planungskontext (Karlström & Hamza, 2021) geschuldet und können schwerlich als Indikator für die Unterrichtsplanungskompetenz der Probanden gelten. Der Genese von Planungsentscheidungen und der Qualität ihrer Begründungen ließe sich beispielsweise mithilfe von Interviews beikommen, die mit den Referendarinnen und Referendaren im Anschluss an das Verfassen der Entwürfe hätten geführt werden müssen. Dies war aufgrund der Bereitstellung anonymisierter Entwürfe in diesem Projekt jedoch nicht möglich.

Ein letzter, konstruktbezogener Einwand sei der Vollständigkeit halber erwähnt: Der Versuch, Unterrichtsplanungskompetenz basierend auf der Kompetenzdefinition von Blömeke et al. (2015) als Kontinuum aufzufassen und dies innerhalb der Professionsforschung mit der PCK-Modellierung im RCM (Carlson et al., 2019) zusammenzuführen (Abb. 3), führt dazu, dass im Rahmen der Beiträge 2-5 erfasste Konstrukt  $ePCK_p$  genannt wurde. Demgegenüber würden Alonzo et al. (2019) einwenden, dass  $ePCK$  grundsätzlich gar nicht erfassbar ist:

*The tacit nature of  $ePCK$  presents a clear challenge for researchers seeking to capture this realm of PCK. Even when connected to a particular instance of science instruction, artefacts such as lesson plans or annotated videos capture  $pPCK$  (expressed when teachers' reasoning is made explicit as part of macroprocesses of planning or reflecting), rather than  $ePCK$ . Because  $ePCK$  is transformed into  $pPCK$  as it is made explicit, we argue that it is impossible to capture the true nature of  $ePCK$ . (Alonzo et al., 2019, S. 280)*

Wie Alonzo et al. (2019) selbst ausführen, handelt es sich beim Zusammenhang von  $ePCK$  zu  $pPCK$  zunächst um eine theoretische Verortung, die der empirischen Prüfung erst noch bedarf. Die Annahme, dass  $ePCK$  in dem Moment, in dem es expliziert wird, zu  $pPCK$  transformiert wird, ist nur insofern plausibel, als  $ePCK$  laut RCM (Carlson et al., 2019) Teil des  $pPCK$  ist. Die

Übergänge sind bislang nur vage definiert. Anders als Alonzo et al. (2019) und in Übereinstimmung mit Stender und Brückmann (2020) wird im Rahmen der vorliegenden Arbeit der Standpunkt vertreten, dass ePCK<sub>p</sub> das in einer konkreten Planungssituation angewandte fachdidaktische Wissen zeigt und dass dieses in Planungsprodukten (z. B. durch qualitative Inhaltsanalyse von Unterrichtsentwürfen; Beitrag 3: Großmann & Krüger, 2022c) sowie Planungsprozessen (z. B. durch Lautes Denken beim Planen; Koberstein-Schwarz & Meisert, 2022) erfasst werden kann.

### *Erkenntnisgewinn*

Trotz der zahlreichen Limitationen, die bei der Interpretation der in den Beiträgen 2, 3, 4 und 5 dargestellten Ergebnisse zu berücksichtigen sind, lassen sich einige Befunde festhalten, die die Forschung zur Unterrichtsplanungskompetenz aus biologiedidaktischer Perspektive voranbringen. Die vorliegende Arbeit kommt der von Weingarten (2019) geforderten „*umfangreichen Exploration des Forschungsfeldes im Sinne einer kritischen Bestandsaufnahme der gegenwärtigen Planungspraxen von Lehrkräften*“ (S. 28) nach und leistet somit die erste derartige Untersuchung für die Biologiedidaktik seit Haas (1998). Überließe man das Konstrukt Unterrichtsplanungskompetenz allein der Allgemeindidaktik und Pädagogik (z. B. König & Rothland, 2022; Scholl et al., 2019), würden fachspezifische Aspekte unentdeckt bleiben. Der Unterrichtsplanung kommt jedoch gewissermaßen eine Vermittlerrolle zwischen der fachdidaktischen Forschung einerseits und der schulischen Unterrichtspraxis andererseits zu: Wenn der fachdidaktischen Forschung daran gelegen ist, ihre Forschungsergebnisse wie beispielsweise Kompetenzmodelle (z. B. Modellierkompetenz; Krüger & Upmeyer zu Belzen, 2021; Bewertungskompetenz; Eggert & Bögeholz, 2006) unterrichtsrelevant werden zu lassen, dann wäre ein Blick auf die Art und Weise, wie (angehende) Lehrkräfte Unterricht planen, durchführen und reflektieren notwendig, um Herausforderungen in der Implementierung derartiger Konstrukte zu identifizieren. Die Beiträge 2-5 leisten einen Beitrag, eine Brücke zwischen der Professionsforschung (→ PCK), der Allgemeindidaktik (→ generische Unterrichtsplanung) und der Fachdidaktik (→ biologiespezifische Unterrichtsplanung) zu schlagen. Drei zentrale Erkenntniszuwächse aus den Beiträgen werden im Folgenden diskutiert:

Erstens liegt das Innovationspotential der Beiträge 2-5 zumindest für die deutsche Lehrkräftebildung darin, insbesondere die 2. Phase der Lehrkräftebildung in den Blick zu nehmen, die in der empirischen Forschung insgesamt (Anderson-Park & Abs, 2020; Kunz & Uhl, 2022) und in den Naturwissenschaftsdidaktiken allzu oft nicht berücksichtigt wird. In Bezug auf die Unterrichtsplanungskompetenz folgten bereits Weingarten (2019) sowie die Gruppe um Johannes König (z. B. Fladung, 2022; König et al., 2015; König et al., 2020a) dem Gedanken, anhand von

Staatsexamensentwürfen Rückschlüsse darauf zu ziehen, über welches Wissen und welche Fähigkeiten Lehrkräfte am Ende der sechs- bis siebenjährigen Ausbildung verfügen und welche Herausforderungen dort noch bestehen. Derartige Erkenntnisse könnten beispielsweise im Zuge der Entwicklung eines Kompetenzentwicklungsmodells genutzt werden, das für die Unterrichtsplanung derzeit noch nicht vorliegt. Insbesondere die Ergebnisse der Beiträge 3, 4 und 5 zeigen, dass Unterrichtsplanungskompetenz am Ende der Ausbildung noch keineswegs vollkommen ausgebildet ist. Damit stützen die Befunde die Annahme, dass der Erwerb von Unterrichtsplanungskompetenz einen langfristigen berufsbiografischen Prozess darstellt (Munthe & Conway, 2017). Eine intensivere Untersuchung von erfahrenen Lehrkräften erscheint somit plausibel, um nachzuzeichnen, in welchen Entwicklungsstufen sich Unterrichtsplanungskompetenz entwickelt und welches Wissen tatsächlich praxisrelevant ist und welches nur für Prüfungen wie die Staatsexamensprüfung relevant ist. Für ein solches Vorhaben wären dann jedoch keine Unterrichtsentwürfe mehr zu untersuchen, die erfahrene Lehrkräfte in der 3. Phase nicht mehr schreiben, sondern andere methodische Zugänge wie mikroethnologische Interviews (z. B. Neuweg, 2018) notwendig.

Zweitens begnügen sich die Beiträge 2-5 nicht mit einer „*Exploration*“ (Weingarten, 2019, S. 28). Stattdessen wird neben dem in den Staatsexamensentwürfen kodierten ePCK<sub>p</sub> eine zweite Variable in Form der holistischen Einschätzungen der Planungsqualität durch einen Fachseminarleiter eingeführt. Neben der Stärkung der Zusammenführung von Auszubildenden der 1. und 2. Phase der Lehrkräftebildung liegt der Vorzug dieses Vorgehens darin, einen Zusammenhang zwischen der Vernetztheit des PCK und der holistisch beurteilten Planungsqualität herzustellen. Dass die Qualität von PCK vor allem der Vernetzung einzelner PCK-Elemente liegt (Park & Oliver, 2008; Park & Suh, 2019) konnte in Beitrag 3 nicht bestätigt werden. Dies könnte neben den in den Limitationen angeführten methodischen Aspekten daran liegen, dass die Qualität eines Unterrichtsentwurfs nur in Teilen am darin explizierten PCK liegt: In ihrem innovativen Versuch, für die Physikdidaktik PCK-Niveaustufen zu formulieren und diese für die Auswertung eines Paper-Pencil-Tests zu nutzen, stellen Schiering et al. (2023) fest, dass die im RCM (Carlson et al., 2019) benannten *amplifiers* und *filters* von großer Bedeutung für die Entwicklung von PCK sowie die Interpretation von Studienergebnissen sein können (vgl. Beitrag 1). Insbesondere Unterrichtserfahrung sowie die Einstellung, eher schüler- oder lehrerzentriert (Teo et al., 2008) zu unterrichten, sind wichtige Variablen, die das PCK von angehenden Lehrkräften beeinflussen, wie bereits in anderen Studien (z. B. Brown et al., 2013; Jong et al., 2005) gezeigt wurde. Beides konnte im Rahmen dieser Studie nicht kontrolliert werden. Vor allem die Frage, inwiefern die geplanten Stunden im konstruktivistischen Sinne schülerzentriert geplant worden sind, kann jedoch

maßgeblich in die Beurteilung des Fachseminarleiters mit eingeflossen sein. Wenn das holistische Globalurteil des Fachseminarleiters also neben der fachdidaktischen Qualität noch andere Variablen des RCM (Carlson et al., 2019) miteinbezieht, ist es nicht verwunderlich, wenn die Korrelation zum ePCK<sub>p</sub> nicht signifikant ausfällt. Darüber hinaus ist zu bedenken, dass das Verfassen eines schriftlichen Unterrichtsentwurfs eine sehr herausfordernde Aufgabe darstellt, die neben konzeptuellem und prozeduralem Professionswissen (PK, CK, PCK) auch die Fähigkeit voraussetzt, Planungsgedanken schriftlich angemessen darzustellen: Kaya und Kaya (2023) untersuchten jüngst, welche Faktoren die Güte schriftlicher Unterrichtsentwürfe zur Analyse von PCK beeinflussen. Bei zwei Dritteln ihrer 79 Probandinnen und Probanden identifizierten sie ein „*lack of self-efficacy beliefs in writing*“ (S. 18), das sich unter anderem in der Vielzahl der beim Planen zu berücksichtigenden Aspekte und einer daraus resultierenden Überforderung zeigt, Planungsentscheidungen in ihrer Wechselwirkung schriftlich zu beschreiben. Dies könnte eine weitere Erklärung dafür sein, warum Beitrag 3 keine Bestätigung der Annahme liefert, dass PCK-Qualität und Anzahl der Vernetzungen positiv miteinander korrelieren: Möglicherweise zeigt die geplante Stunde eine angemessene, zielführende Prozessstruktur mit geeigneten Methoden, die der Fachseminarleiter aus dem Stundenverlaufsplan und den Arbeitsmaterialien herauslesen konnte, ohne dass dies im Entwurf aufgrund mangelnder Schreibfähigkeiten expressiv verbis beschrieben und begründet wurde.

Drittens lassen sich in Bezug auf einzelne PCK-Facetten einige wichtige, für die biologiedidaktische Lehrkräftebildung besonders relevante Beobachtungen machen:

- OTS/KSC<sup>17</sup>: In Beitrag 5 wurde deutlich, dass die klare und sachgemäße Zielformulierung als Ausgangspunkt der Unterrichtsplanung (z. B. Wiggins & McTighe, 2005) eine besondere Herausforderung darstellt (*clarity of learning objectives*; König et al., 2021a). Sie ist insofern allen im Folgenden zu diskutierenden Aspekten der Unterrichtsplanung übergeordnet, als sich diese im Sinne des Kriteriums *Funktionalität* daran messen lassen müssen, ob sie einen Beitrag dazu leisten, Schülerinnen und Schülern die angestrebte Kompetenzentwicklung zu ermöglichen. Zielklarheit und Kohärenz der Lernziele haben einen nachgewiesenen Einfluss auf die Unterrichtsqualität und die Kompetenzentwicklung von Schülerinnen und Schülern (Seidel et al., 2005). Am Beispiel des Kompetenzbereichs Erkenntnisgewinnung (KMK, 2005; KMK, 2020) wurde gezeigt, dass die Referendarinnen

---

<sup>17</sup> Zu *Orientation to Teaching Science* (OTS) und *Knowledge of Science Curriculum* (KSC) werden die Ergebnisse an dieser Stelle nicht separat diskutiert, da die Operationalisierung von OTS als Kompetenzen und damit curriculare Ziele nicht an den internationalen Diskurs anknüpfbar ist, in dem diese Aspekte der Facette KSC zugeordnet werden. Dies wird in Beitrag 7 durch eine modifizierte Abbildung des Pentagon-Modells transparent gemacht, in der die in den Beiträgen 2 und 3 noch OTS zugeordneten Kategorien korrekterweise der Facette KSC zugeordnet werden.



und Referendare der hier untersuchten Stichprobe unter Erkenntnisgewinnung überwiegend die Gewinnung neuer fachlicher Erkenntnisse mit praktischen Arbeitsweisen verstehen, was jedoch dem Kompetenzbereich Fachwissen bzw. Sachkompetenz entspräche (Krell & Krüger, 2022). Zwar wurde dem Kompetenzbereich Erkenntnisgewinnung in den vergangenen 20 Jahren in der fachdidaktischen Forschung beispiellose Aufmerksamkeit gewidmet (für einen Überblick, siehe z. B. Rönnebeck et al., 2016) und neben Schülervorstellungen zum naturwissenschaftlichen Erkenntnisprozess und zu Nature of Science die einzelnen Arbeits- und Denkweisen in theoretischen wie empirischen Arbeiten untersucht. Dennoch scheinen diese Erkenntnisse aus der Forschung noch nicht hinreichend in die Praxis überführt worden zu sein. Dies könnte u. a. daran liegen, dass es bislang kaum Versuche dazu gibt, Erkenntnisse der naturwissenschaftsdidaktischen Kompetenzforschung mit professionstheoretischen Modellen in Verbindung zu bringen. Folgt man dem in dieser Arbeit vertretenen Standpunkt, dass der Unterrichtsplanung eine Scharnierfunktion zukommt, da vor dem Hintergrund von fundiertem PK, CK und PCK Unterricht für Schülerinnen und Schüler geplant wird und folglich empirische fachdidaktische Befunde unterrichtsrelevant werden (können), dann könnte es ratsam sein, die Perspektive der Unterrichtsplanung stärker zu berücksichtigen. Für den Kompetenzbereich Erkenntnisgewinnung bildet bislang einzig die Dissertation von Tardent Kuster (2020) den Versuch, PCK zu einem spezifischen Kompetenzbereich und Unterrichtsplanung zusammen zu denken.

- KSU: In Beitrag 4 wurde deutlich, dass die untersuchten Referendarinnen und Referendare unter Schülerkognition vor allem das bereits vorhandene Vorwissen bzw. die bereits entwickelten Kompetenzen verstehen und ausführlich in ihren Entwürfen darlegen, dass sie jedoch kaum auf Schülervorstellungen und Lernschwierigkeiten eingehen. Dies ist erstaunlich, da es inzwischen deutschsprachige, an Lehrkräfte gerichtete Publikationen gibt (z. B. Hammann & Asshoff, 2014; Kattmann, 2015), die beispielsweise als Ausgangspunkt einer fachlichen Klärung (Duit et al., 2012) genutzt werden könnten. Im Sinne der zuvor diskutierten Bedeutung von schüler- bzw. lehrerzentrierten Einstellungen ließe sich hier zumindest indirekt schlussfolgern, dass entweder Schülervorstellungen – zumindest im Rahmen der hier untersuchten Staatsexamensentwürfe – nur eine geringe Bedeutung beigemessen wird, oder dass ihre Erhebung sowie die sich daran anschließende Abstimmung sämtlicher methodisch-didaktischer Entscheidungen auf die Schülervorstellungen der jeweiligen Lerngruppe als zu herausfordernd empfunden wird (z. B. Moodley & Gaigher, 2019; Morrison & Lederman, 2003). Zumindest geben die

Befunde insofern Anlass zur Reflexion, als es sich bei den untersuchten Entwürfen ( $N = 107$ ) immerhin um eine Gesamtstichprobe zweier Prüfungsdurchgänge (2018/19, 2019) im Bundesland Berlin handelt und zumindest in gewissem Maße von einer Verallgemeinerbarkeit der Befunde ausgegangen werden kann, zumindest für das Land Berlin. Wenn eine so erhebliche Diskrepanz zwischen der großen Bedeutung von Schülervorstellungen in der empirischen fachdidaktischen Forschung einerseits und der offenbar verschwindend geringen Berücksichtigung von Schülervorstellungen von Referendarinnen und Referendaren am Ende ihrer Ausbildung besteht, dann wirft das Fragen auf: Wurden Schülervorstellungen in der untersuchten Stichprobe deswegen nicht berücksichtigt, weil es im Rahmen der Staatsexamensprüfung als unwegbar und unsicher empfunden wurde, schließlich setzt ein idealtypisch umgesetzter *Conceptual Change* (vgl. Krüger, 2007) voraus, dass man sich als Lehrkraft bewusst den vielgestaltigen Schülervorstellungen aussetzt und mit diesen zielgerichtet arbeitet – eine gewiss sehr herausfordernde Aufgabe. Umgekehrt gewendet ließe sich jedoch einwenden: Wenn nicht einmal in Staatsexamensprüfungen Schülervorstellungen berücksichtigt werden und man die Gelegenheit nicht nutzt, zu demonstrieren, dass man konstruktivistischen und didaktisch adaptiven Unterricht (König et al., 2015) planen und durchführen und dabei „Aha“-Momente auslösen kann, wann sonst? In Bezug auf die Lehrkräftebildung insgesamt übertragen lässt sich eine Dichotomie zwischen großer Relevanz in der theoretischen und empirischen Forschung (1. Phase der Lehrkräftebildung) und möglicherweise geringer Relevanz unter Fachseminarleitungen (2. Phase der Lehrkräftebildung) annehmen. Wenn in den Biologie-Fachseminaren der hier untersuchten Referendarinnen und Referendare die Berücksichtigung von Schülervorstellungen nicht gefordert und vor allem geübt wird, ist es nur folgerichtig, dass diese dann auch in der Staatsexamensprüfung nicht berücksichtigt werden. Möglicherweise besteht hierin auch eine der empirisch noch unterbelichteten Herausforderungen in der Transformation von cPCK zu pPCK bzw. von pPCK zu ePCK<sub>p</sub> (Alonzo et al., 2019; Behling et al., 2022; Carlson et al., 2019; Irmer et al., 2023), die durch die in Abbildung 2 als Doppelpfeile dargestellten *amplifiers* und *filters* vermittelt wird: Möglicherweise verfügten die Referendarinnen und Referendare der untersuchten Stichprobe über konzeptuelles Wissen über Schülervorstellungen zum jeweils unterrichteten Thema und hätten auch den *Conceptual-Change*-Ansatz zum Unterrichten mit Schülervorstellungen beschreiben und erklären können ( $\rightarrow$  cPCK/pPCK). Dies jedoch konkret in der Planung der Stunde anzuwenden und somit zu konkretisieren ( $\rightarrow$  ePCK<sub>p</sub>), könnte mit Schwierigkeiten behaftet sein, die im Zuge der so bedeutsamen Staatsexamensprüfung vermieden worden sind, möglicherweise aufgrund fehlenden

prozeduralen Wissens (Beitrag 1). Jedoch auch unabhängig von der Staatsexamensprüfung wurde bereits im Allgemeinen (z. B. Halim & Meerah, 2002; Moodley & Gaigher, 2019; Otero & Nathan, 2008) sowie in Bezug auf die Unterrichtsplanung im Besonderen (Gunckel, 2011; Käpylä et al., 2009; Weitzel & Blank, 2020) gezeigt, dass Schülervorstellungen selten berücksichtigt werden.

- KAs: In Beitrag 3 wurde u.a. deutlich, dass die PCK-Facette *Knowledge of Assessment Strategies of Science Learning* (KAs) sowohl in den Entwürfen mit hoher als auch denen mit geringer Planungsqualität kaum berücksichtigt wurde. Obwohl diese Beobachtung in zahlreichen PCK-Studien gemacht wurde (für eine Übersicht, siehe Chan & Hume, 2019), so ist sie konkret in Bezug auf Unterrichtsplanung erstaunlich: Gerade im Zuge eines Unterrichtsbesuchs, wie ihn die Staatsexamensprüfung darstellt, wäre zu erwarten, dass man den Unterrichtsbeobachterinnen und -beobachtern demonstriert, dass es zwischen dem Beginn der Stunde und dem Stundenende einen nachweislichen Kompetenzzuwachs bei den Schülerinnen und Schülern gegeben hat. Dies setzt jedoch voraus, dass am Stundenende ein Element eingeplant wird (z. B. ein Impuls im Unterrichtsgespräch, eine vertiefende Übungsaufgabe o.ä.), anhand dessen man nachweisen kann, dass die Schülerinnen und Schüler das angestrebte Lernziel tatsächlich erreicht haben. Zwar wird dieses Unterrichtsplanungsprinzip vor allem in der angloamerikanischen Didaktik vertreten (z. B. Wiggins & McTighe, 2005), in der deutschsprachigen jedoch kaum. Von prüfungsstrategischen Überlegungen abgesehen ist wertschätzendes Feedback ein wichtiges Element der Unterrichtsplanung, dessen lernförderlicher Effekte empirisch belegt sind (z. B. Hattie, 2023). Auf die sechs kognitiven Herausforderungen in der Unterrichtsplanung (König et al., 2021a) angewendet hieße das, die Verknüpfung zwischen dem angestrebten Lernziel (*clarity of learning objectives*) und der Phasenstruktur (*phasing*), insbesondere der Sicherungsphase, als Ausgangspunkt der Planung anzusehen. Wenn man auf diese Weise vom Lernziel her plant (*Backward Design*; Wiggins & McTighe, 2005), ist eine Grundstruktur der Unterrichtsstunde etabliert, die insofern funktional ist, als letztlich eine Passung zwischen der im Lernziel formulierten angestrebten Kompetenzentwicklung und dem Stundenergebnis in der Sicherungsphase gegeben ist.

Zusammengefasst zeigt sich, dass die Quantität an in der Planung berücksichtigten PCK-Elementen und PCK-Verbindungen keinen hinreichenden Indikator für die Planungsqualität darstellt. Stattdessen ist ein qualitativer Blick nötig: Ausgehend von den in den Beiträgen 3, 4 und 5 dargestellten Ergebnissen sowie von weiteren empirischen und theoretischen Befunden zur Unterrichtsplanung lassen sich Performanzerwartungen formulieren, also Beschreibungen dessen,

wie angehende Biologielehrkräfte ihren Unterricht qualitativ hochwertig planen. Dabei geht es nicht um präskriptive Erwartungen an die inhaltliche und formale Gestaltung von Unterrichtsentwürfen, die es – empirisch nicht fundiert – ja bereits zahlreich gibt, beispielsweise in Form von Planungsratgebern (für eine Übersicht, siehe Vogelsang & Riese, 2017).

Kriterienraster stellen einen geeigneten Typus von Instrument dar, um die bislang dargestellten und diskutierten Ergebnisse qualitativ in Qualitätskriterien zu überführen. Solche Kriterien fehlen derzeit noch (Krepf & König, 2022b). Sie könnten einerseits Auszubildenden der 1. und 2. Phase der Lehrkräftebildung als Orientierungsrahmen oder gar als Beurteilungsraster im Rahmen von Prüfungen dienen, und andererseits angehenden Lehrkräften transparent machen, wie ein gelungener Unterrichtsentwurf gestaltet werden soll. Dies kann aber nur gelingen, wenn ein solches Kriterienraster darauf hin geprüft wurde, ob die mit ihm ermittelten Scores valide Interpretationen erlauben. Dies wurde jedoch für die wenigen bereits verfügbaren Kriterienraster zur Unterrichtsplanung (z. B. Jacobs et al., 2008; Ndiokubwayo et al., 2022) noch nicht geleistet. Im folgenden Kapitel werden daher zunächst die Entwicklung eines Kriterienrasters (Beitrag 6) und anschließend die Überprüfung des Kriterienrasters aus Validitätsperspektive (Beitrag 7) dargestellt, ehe die Kriterien Dozierenden bzw. Fachseminarleiterinnen und Fachseminarleitern (Beitrag 8) sowie Studierenden bzw. Referendarinnen und Referendaren (Beitrag 9) vorgestellt werden.

*„Je planmäßiger die Menschen vorgeben,  
desto wirksamer vermag sie der Zufall zu treffen.“  
(Friedrich Dürrenmatt)*

## 4 Entwicklung und Überprüfung eines Kriterienrasters

Trotz der Bedeutung schriftlicher Unterrichtsentwürfe in der (deutschen) Lehrkräftebildung gibt es bislang keine empirisch geprüften Kriterien, mit denen Unterrichtsentwürfe objektiv, reliabel und valide analysiert und bewertet werden können (König et al., 2021a). Solche Kriterien könnten dazu beitragen, die möglicherweise nicht zu Unrecht von Referendarinnen und Referendaren und von Ausbildenden selbst als inkonsistent und intransparent empfundene Beurteilungspraxis im Referendariat (Döbrich & Abs, 2008; Kärner et al., 2019; Strietholt & Terhart, 2009) stärker zu standardisieren, immer schon implizit vorhandene Erwartungen an gute Unterrichtsentwürfe zu explizieren und damit letztlich eine Lernhilfe beim Erwerb von Unterrichtsplanungskompetenz darzustellen.

Zu diesem Zwecke wurden die in den Beiträgen 2, 3, 4 und 5 gewonnenen Erkenntnisse dazu genutzt, ein Kriterienraster zu entwickeln, mit dem die Qualität von Unterrichtsentwürfen eingeschätzt werden kann. Schröder et al. (2020) zufolge werden in der empirischen Unterrichtsplanungsforschung derzeit drei methodische Zugänge zur Erfassung von Unterrichtsplanungskompetenz genutzt: (1) Vignetten, die aufgrund ihres hohen Grades an Standardisierung und eines eindeutigen Erwartungshorizonts testökonomisch betrachtet den Vorzug großer Objektivität und Inhaltsvalidität besitzen, allerdings zulasten ökologischer Validität aufgrund eines zumeist unauthentischen Planungskontexts; (2) reale Unterrichtsplanungen, die unter authentischen Bedingungen von (angehenden) Lehrkräften verfasst worden sind, die jedoch durch die Komplexität realer Lehr-Lern-Situationen und die heterogenen Anforderungen an den Ausbildungsstandorten kaum objektiv und reliabel ausgewertet werden können; und (3) Performanztests, die einerseits die Standardisierung und andererseits die Authentizität der Planungssituation zusammenführen und die Defizite der beiden erst genannten Zugänge kompensieren sollen.

Das im Rahmen dieser Arbeit entwickelte Kriterienraster ist dem zweiten methodischen Zugang zuzuordnen: Es soll dazu dienen, schriftliche Unterrichtsentwürfe objektiv und reliabel auszuwerten zu können, und hätte gegenüber dem von Schröder et al. (2020) präferierten Performanztest den Vorzug, nicht an einen spezifischen Kontext (dort: Mechanik) gebunden zu sein, sondern auf

sämtliche in Unterrichtsentwürfen behandelte Themen und Kompetenzbereiche des Biologieunterrichts<sup>18</sup> anwendbar zu sein (*general rubric*; Brookhart, 2018).

Auch wenn Kriterienraster als Instrument in der biologiedidaktischen Forschung bislang noch kaum vertreten sind, so hat sich in den vergangenen 20 Jahren doch eine erhebliche Anzahl an vor allem internationalen Forschungsvorhaben mit der Entwicklung und Evaluation von Kriterienrastern befasst, wie beispielsweise an den inzwischen zahlreichen Literaturreviews zu Kriterienrastern ersichtlich ist (Brookhart, 2018; Brookhart & Chen, 2015; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Panadero & Jonsson, 2020; Reddy & Andrade, 2010). Um die folgenden Ausführungen auf die für diese Arbeit relevanten Gestaltungsmerkmale zu beschränken, sei das in den folgenden Beiträgen dargestellte Instrument folgendermaßen charakterisiert:

1. Es handelt sich um ein Kriterienraster, das anders als die verwandten Checklisten und Ratingskalen qualitative Performanzbeschreibungen für die verschiedenen Niveaustufen bietet. Checklisten erlauben hingegen bloß eine dichotome Entscheidung (z. B. vorhanden/nicht vorhanden), während Ratingskalen zwar verschiedene Niveaustufen unterscheiden (z. B. sehr gut, gut, befriedigend, ausreichend, mangelhaft), ohne für die jeweiligen Stufen jedoch zu explizieren, welche Performanz jeweils zugrunde liegt (Brookhart, 2018).<sup>19</sup>
2. Es handelt sich um ein analytisches Kriterienraster, das anders als holistische Kriterienraster nicht bloß eine einzige Entscheidung erfordert, bei der alle Kriterien gleichzeitig zu berücksichtigen sind, sondern bei dem sämtliche Kriterien jeweils einzeln beurteilt werden müssen (Brookhart, 2018). Da für jedes Kriterium und jede Niveaustufe (im besten Falle) präzise und trennscharfe Performanzbeschreibungen (Chan & Ho, 2019) vorliegen, ist die Beurteilung letztlich sehr differenziert und nachvollziehbar. Daher gelten analytische Kriterienraster als „*gold standard*“ unter den Kriterienrastern (Schreiber et al., 2012, S. 212).
3. Da das Kriterienraster grundsätzlich für Unterrichtsentwürfe im Fach Biologie eingesetzt werden soll und kein bestimmtes Thema bzw. Kompetenzbereich und keine Lerngruppe vorgegeben ist, handelt es sich um eine *general rubric*, nicht um eine *task-specific rubric* (Brookhart, 2018; Tierney & Simon, 2004). Die Performanzbeschreibungen sind also so allgemein gehalten, dass sie auf sämtliche Unterrichtsentwürfe anwendbar sind.

---

<sup>18</sup> Aufgrund seiner letztlich generischen Gestalt könnte dieses Kriterienraster aber mit kleineren Anpassungen oder Ergänzungen ebenso in anderen Fächern eingesetzt werden.

<sup>19</sup> Kriterienraster sind jedoch nicht per se als höherwertige Instrumente anzusehen. Je nach Einsatzzweck können auch Checklisten und Ratingskalen Vorzüge gegenüber Kriterienrastern haben (Schreiber et al., 2012). Insbesondere da die Entwicklung von Kriterienrastern sehr viel aufwändiger ist als die Entwicklung einer Checkliste oder einer Ratingskala, wird oft aus zeitökonomischen Gründen auf den Einsatz von Kriterienrastern verzichtet, obwohl sie einen nachgewiesenen positiven Einfluss auf Lernprozesse haben (z. B. Brookhart, 2018).

4. Eingesetzt wird das Kriterienraster zunächst durch Dozierende und Fachseminarleitungen, um die Qualität von Unterrichtsentwürfen einzuschätzen. Es handelt sich derzeit noch um eine *scoring rubric*, die anders als *instructional rubrics* noch nicht von der Zielgruppe (angehenden Biologielehrkräften) als Lernhilfe selbst genutzt wird (Andrade, 2005).
5. Daran anknüpfend wird das Kriterienraster bislang im Sinne von *summative assessment* eingesetzt, d. h. um im Rahmen von Prüfungen die Qualität von Unterrichtsentwürfen bewerten zu können. Langfristig gesehen wird das Kriterienraster auch zum Zwecke von *formative assessment* (Panadero & Jonsson, 2013) eingesetzt werden, so dass angehende Biologielehrkräfte beispielsweise im Verlauf des Referendariats wiederholt Rückmeldung im Lernprozess bekommen und somit Feedback zu ihrer Unterrichtsplanungskompetenz erhalten.

In diesem Kapitel werden zunächst aus der theoretischen und empirischen Kriterienraster-Literatur abgeleitete Qualitätskriterien dargestellt, ehe die Entwicklung des Kriterienrasters (Beitrag 6) sowie die Überprüfung des Kriterienrasters aus Validitätsperspektive (Beitrag 7) beschrieben wird. Anschließend wird dargelegt, wie das Kriterienraster von Auszubildenden in der Lehrkräftebildung genutzt werden kann (Beitrag 8). Schließlich werden die entwickelten Qualitätskriterien anhand von Beispielen aus den in den Beiträgen 2-5 analysierten Staatsexamensentwürfen für angehende Biologielehrkräfte der 1. und 2. Phase illustriert und erläutert (Beitrag 9).

### 4.1. Entwicklung

#### 4.1.1. Qualitätskriterien von Kriterienrastern

Abseits der naturwissenschaftsdidaktischen Forschung gibt es einen inzwischen mehr als zwei Jahrzehnte währenden, intensiven Diskurs über Kriterienraster, der insbesondere in Zeitschriften wie *Assessment & Evaluation in Higher Education* geführt wird. Diesem Diskurs sind eine Reihe von Qualitätskriterien zu entnehmen, die in der Entwicklung von Kriterienrastern zu berücksichtigen sind (Tab. 3). Die folgende Darstellung erhebt keinen Anspruch auf Vollständigkeit. Es werden nur diejenigen Kriterien aufgeführt, die für die Entwicklung eines *analytischen* Kriterienrasters relevant sind.

Diese fünf Kriterien wurden in der Entwicklung des Kriterienrasters zur Einschätzung der Planungsqualität berücksichtigt. Insbesondere E1 ist dabei relevant, da hierin die Verbindung zwischen der in den Beiträgen 2-5 dargelegten Analyse von Staatsexamensentwürfen liegt. Mithilfe

der holistischen Qualitätseinschätzung des Fachseminarleiters und seinen Begründungen (Anhang 2) gelang es, insgesamt 24 Kriterien mit jeweils drei Niveaustufen zu entwickeln. Das Strukturprinzip der Kriterien wurde in Beitrag 6 dargestellt.

**Tabelle 3**

*Kriterien für die Entwicklung von Kriterienrastern*

Kriterium	Beschreibung	Quellen	
E1	Induktives Vorgehen	Kriterien und Niveaustufen induktiv aus dem Datenmaterial herleiten, das mit dem Kriterienraster bewertet werden soll	Sadler (2014)
E2	Anzahl der Kriterien	So viele Kriterien wie nötig, aber so wenige wie möglich; Voraussetzung: alle Kriterien stehen im Einklang mit dem Ziel (z. B. den curricularen Standards)	Moskal (2000)
E3	Anzahl der Niveaustufen	So viele Niveaustufen wie nötig, um Qualitätsniveaus voneinander abzugrenzen; Niveaustufen müssen trennscharf sein	Chan & Ho (2019)
E4	Spezifität	Performanzbeschreibungen so generisch bzw. spezifisch formulieren, wie es für die zu bewältigende Aufgabe notwendig ist	Dawson (2017); Tierney und Simon (2004)
E5	Sprache	Objektiv-deskriptiv Performanz beschreiben statt subjektiv-evaluativ („gute Struktur“ o.ä.), da sonst nicht deutlich wird, welche Performanz erstrebenswert ist	Brookhart (2018); Chan & Ho (2019)

#### 4.1.2. Beitrag 6: *Identifying Performance levels of enacted pedagogical content knowledge in trainee biology teachers' lesson plans*

**Großmann, L. & Krüger, D. (2023).** Identifying Performance Levels of Enacted Pedagogical Content Knowledge in Trainee Biology Teachers' Lesson Plans. In: G. S. Carvalho, A. S. Afonso & Z. Anastácio (Hrsg.), *Fostering Scientific Citizenship in an Uncertain World. Selected Papers from the ESERA 2021 Conference* (S. 95-116). Cham: Springer. [https://doi.org/10.1007/978-3-031-32225-9\\_7](https://doi.org/10.1007/978-3-031-32225-9_7)

#### **Zusammenfassung**

In diesem Beitrag wurden die folgenden Forschungsfragen untersucht:

- F1. What cognitive demands are PBTs (pre-service biology teachers) confronted with when writing a lesson plan?
- F2. Which performance levels describe PBTs' ePCK<sub>p</sub> when writing a lesson plan?

Um im Kriterienraster diejenigen fachdidaktischen Aspekte abzubilden, die in der Unterrichtsplanung besonders relevant sind, wurde als theoretischer Rahmen das heuristische CODE-PLAN-Modell (König et al., 2021a) gewählt, in dem sechs wesentliche kognitive Herausforderungen in der Unterrichtsplanung genannt werden. Diesen wurden fachdidaktische



Wissensaspekte aus dem Pentagon-Modell (Park & Oliver, 2008) zugeordnet, die in der Analyse der Staatsexamensentwürfe (Beiträge 2-5) untersucht wurden. Durch diese Verbindung von PCK und Unterrichtsplanung wird hier also erneut das Konstrukt ePCK<sub>p</sub> in den Blick genommen.

Um zu identifizieren, welche kognitiven Herausforderungen beim Schreiben eines Unterrichtsentwurfs bewältigt werden müssen (F1), wurden vor allem die Kriterien E1 und E2 (Tab. 3) berücksichtigt: Sämtliche Kriterien eines Kriterienrasters sollen im Einklang mit dem Ziel stehen (Moskal & Leydens, 2000). Das Ziel in diesem Fall wären diejenigen Standards aus den Standards für die Lehrerbildung (KMK, 2019), die sich auf die Unterrichtsplanung beziehen (Tab. 1). Diese sollen von Lehrkräften am Ende ihrer Ausbildung erfüllt werden, insofern ist es plausibel, sie zum Ausgangspunkt der Kriterienentwicklung zu machen. Wie jedoch bereits gezeigt wurde (Tab. 1), sind diese Standards für die Unterrichtsplanung noch nicht hinreichend differenziert, um die Vielfalt der in der Unterrichtsplanung zu leistenden Teiloperationen (Abb. 1) abzudecken. Sie stellen also eine notwendige, aber noch keine hinreichende Bedingung dafür dar, dass das Kriterienraster die Qualität von Unterrichtsentwürfen valide erfasst. Daher wurden wie von Sadler (2009) vorgeschlagen induktiv weitere Kriterien entwickelt, die sich aus der Analyse der Staatsexamensentwürfe (Beiträge 2-5) ergeben haben. Dieses Vorgehen ist deswegen relevant, weil hierdurch ökologische Validität (Wegener & Blankenship, 2007) sichergestellt wurde, d. h. die Kriterien bilden diejenigen Merkmale von Unterrichtsentwürfen ab, wie sie tatsächlich in der Ausbildungspraxis (zumindest im Land Berlin; SenBJF, 2017) verfasst werden. Auf diese Weise wurden insgesamt 24 Kriterien entwickelt, von denen anzunehmen ist, dass sie inhaltlich relevante Aspekte schriftlicher Unterrichtsentwürfe abdecken. Diese und weitere Annahmen zur Qualität der Kriterien aus Beitrag 6 werden anschließend in Beitrag 7 überprüft.

Ausgehend von dieser Auswahl an Kriterien mussten Niveaustufen entwickelt und jeweils Performanzbeschreibungen verfasst werden (F2). Früh in der Entwicklung wurde deutlich, dass drei Niveaustufen das optimale Maß an Graduierung darstellen. Beim induktiven Vergleichen von Textstellen aus verschiedenen Entwürfen zeigten sich große Schwierigkeiten, allzu nuancierte und vor allem trennscharfe Niveaus zu unterscheiden, so dass entgegen der weit verbreiteten Nutzung von vier oder gar fünf Niveaustufen (Brookhart, 2018) das vorliegende Kriterienraster nur drei Stufen umfasst (E3, Tab. 3). Außerdem muss das Kriterienraster handhabbar bleiben, was angesichts der großen Anzahl an Kriterien ohnehin schon eine Herausforderung darstellt.

Als Strukturprinzip wurden in der sozialwissenschaftlichen Forschung entwickelte und empirisch geprüfte Graduierungsbegriffe (Rohrman, 1978; Rohrman, 2007) genutzt. Drei Skalen wurden identifiziert, die für die Graduierung der 24 Kriterien notwendig sind: Für manche Kriterien ist es sinnvoll, die *Häufigkeit* zu unterscheiden, mit der ein Aspekt berücksichtigt wurde (*nie – selten – oft*).

Dies wurde jedoch nur für sehr wenige Kriterien angewendet, denn wie in Beitrag 3 gezeigt wurde, steckt die Qualität weniger in der Häufigkeit als in der Tiefe, in der ein Aspekt berücksichtigt wird. Daher wurde die *Qualität* unterschieden, indem im Sinne von Oberflächen- und Tiefenstrukturen (Decristan et al., 2020; Oser & Baeriswyl, 2001) geprüft wird, ob ein Aspekt vorwiegend auf der Ebene der methodischen Gestaltung (Oberflächenstrukturebene) oder auf der Ebene der kognitiven Prozesse mit Blick auf den angestrebten Kompetenzzuwachs (Tiefenstrukturebene) dargestellt und begründet wird. Schließlich wurde für die meisten Kriterien unterschieden, inwiefern eher generisch oder fachspezifisch (König et al., 2020b) beschrieben und begründet wird. Für jede der drei Skalen werden in Beitrag 6 jeweils ein Beispiel aus den Staatsexamensentwürfen auf der mittleren Niveaustufe und eines auf der höchsten Niveaustufe einander gegenübergestellt. Darunter wird dann jeweils abstrahierend beschrieben, welche Performanz dort zu beobachten ist. Hierbei wurde beachtet, dass die Beschreibungen so formuliert sind, dass sie für sämtliche Themen und Kompetenzbereiche anwendbar sind (Kriterium E4; Tab. 3). Gewährleistet wird dies über den Begriff *Funktionalität* (Beitrag 3), der bereits insofern ein relatives Beurteilungsmaß darstellt, als methodisch-didaktische Entscheidungen nicht per se richtig oder falsch sind, sondern angesichts der jeweiligen Lerngruppe, des angestrebten Kompetenzzuwachses und des Themas (Abb. 1; Abb. 3) als eher funktional oder eher dysfunktional zur Kompetenzförderung eingeschätzt werden können. Zudem wurde versucht, nicht in einer wertenden, sondern in einer deskriptiven Sprache die Performanz zu beschreiben (Kriterium E5; Tab. 3), d. h. auf Ausdrücke wie „... ist gut gelungen“ wurde verzichtet, da hierdurch nicht deutlich werden würde, was die jeweilige Performanz im Detail als *gut* auszeichnet. Auf diese Weise liegt ein Kriterienraster vor, in dem 24 Kriterien, die überwiegend die Vernetzung verschiedener ePCK<sub>p</sub>-Aspekte erfordern, dargestellt sind.

## 4.2. Überprüfung

Dem Kriterienraster liegen die folgenden Annahmen zugrunde: Hat eine angehende Biologielehrkraft mithilfe des Kriterienrasters einen hohen Score erhalten, so handelt es sich um einen sehr guten Unterrichtsentwurf. Aus dieser Performanz lässt sich im Sinne von Blömeke et al. (2015) schlussfolgern, dass diese Person über ausgeprägte Unterrichtsplanungskompetenz verfügt. Umgekehrt würde man schlussfolgern, dass eine angehende Biologielehrkraft, die einen geringen Score erreicht, über geringe Unterrichtsplanungskompetenz verfügt.

Ob diese Annahmen aber zutreffen, bedarf einer empirischen Überprüfung:

*Die reliable und valide Erfassung professioneller Kompetenzen ist nicht nur Voraussetzung für empirische Erkenntnisse über die Bedingungen und Konsequenzen professioneller Kompetenz, sondern [...] auch Grundlage für wichtige Entscheidungen und Handlungsempfehlungen, etwa bei der Evaluation von Aus- und Weiterbildungsangeboten oder der Begleitung von Reformprozessen. Wie aber können die Fähigkeiten und Kenntnisse von Lehrkräften valide erfasst werden? (Kunter & Klusmann, 2010, S. 74f.)*

Da das im Rahmen dieser Arbeit entwickelte Kriterienraster dazu dienen soll, empirisch gesicherte Qualitätskriterien zur Beurteilung schriftlicher Unterrichtsentwürfe (Krepf & König, 2022b) zu liefern, und da beansprucht wird, mithilfe des Kriterienrasters die Qualität von Unterrichtsentwürfen zuverlässiger einschätzen zu können als mit Checklisten oder Ratingskalen, ist es von großer Bedeutung, die oben genannten Annahmen zu prüfen. Schließlich wäre es beispielsweise möglich, dass die Performanzbeschreibungen auf der höchsten Niveaustufe vieler Kriterien so einfach zu erreichen wären, dass man selbst mit einem offenkundig eher überarbeitungsbedürftigen Unterrichtsentwurf dennoch einen relativ hohen Score erreichen kann, und vice versa. Das Kriterienraster böte nur dann einen Mehrwert für die empirische Unterrichtsplanungsforschung einerseits und die Lehrkräftebildung andererseits, wenn sichergestellt ist, dass der mithilfe des Kriterienrasters ermittelte Score tatsächlich Planungsqualität ausdrückt. Es müssen daher Argumente dafür gesammelt werden, dass das Kriterienraster *valide* Interpretationen erlaubt. Reddy und Andrade (2010) weisen jedoch darauf hin, dass eine Validitätsprüfung im Rahmen der Entwicklung von Kriterienrastern im Allgemeinen sehr selten vorgenommen wird. Im Folgenden wird zunächst kurz der aktuelle Validitätsdiskurs skizziert, in den sich Beitrag 7, in dem das entwickelte Kriterienraster überprüft wird, einreicht.

#### **4.2.1. Validität**

Lange Zeit fristete die Validität ein Schattendasein in der empirischen Forschung und fand als Diskurs unter Psychometrikerinnen und Psychometrikern sowie Testtheoretikerinnen und Testtheoretikern statt. In den letzten 10 Jahren gibt es vermehrte Bemühungen in den Naturwissenschaftsdidaktiken, für die entwickelten Instrumente Validitätsevidenzen zu sammeln und somit zu gewährleisten, dass die Interpretationen der Testscores tatsächlich etwas über die Kompetenz aussagen, die gemessen werden soll (z. B. Krüger et al., 2020; Schreiber & Gut, 2022). Die vorliegende Arbeit reiht sich in diese noch junge Forschungstradition ein und bereichert die Unterrichtsplanungsforschung insofern, als sie das entwickelte Kriterienraster aus verschiedenen Blickwinkeln prüft. Damit ergänzt es die wenigen bereits vorhandenen Kriterienraster, die sich zumeist – wenn überhaupt – auf eine Experteneinschätzung zur Angemessenheit der Kriterien beschränken (z. B. Enugu & Hokayem, 2017; Forbes & Davis, 2010; Kademian & Davis, 2018).

Zur klaren Verortung der folgenden Ausführungen sei folgender Hinweis vorangestellt: Im deutschsprachigen Diskurs wird weiterhin die tradierte Trias aus Objektivität, Reliabilität und Validität angeführt, wenn man über Gütekriterien empirischer Forschung spricht (z. B. Döring & Bortz, 2016). Als Referenzrahmen dieser Arbeit werden jedoch die inzwischen etablierten *Standards for Educational and Psychological Testing* (AERA et al., 2014) genutzt, denen zufolge Objektivität und Reliabilität Voraussetzungen für Validität sind und sich diese beiden Gütekriterien den fünf *sources of validity evidence* (AERA et al., 2014) zuordnen lassen. Aus diesem Grund gibt es neben diesem Unterkapitel zu Validität keine weiteren Unterkapitel zu Objektivität und Reliabilität.

AERA et al. (2014) definieren Validität wie folgt:

*Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests. The process of validation involves accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores for proposed uses that are evaluated, not the test itself.* (AERA et al., 2014, S. 11)

Diese Definition steht im Widerspruch zur in den Naturwissenschaftsdidaktiken immer noch verbreiteten Annahme, dass Validität die Frage beantwortet, ob ein Testinstrument das misst, was es messen soll (z. B. Schmiemann & Lücken, 2014). Demnach wäre Validität weiterhin ein Merkmal des Testinstruments selbst. Zumeist über statistische Verfahren (z. B. Faktorenanalysen, Cronbach's  $\alpha$  als Maß für interne Konsistenz der Skalen) wird dann einmalig geprüft, ob der Test *valide* ist. Dass dies jedoch keine hinreichende Bedingung dafür ist, von valider Testwertinterpretation auszugehen, lässt sich am entwickelten Kriterienraster nachvollziehen: Theoretisch ließe sich nach Anwendung des Kriterienrasters auf eine hinreichend große Anzahl an Unterrichtsentwürfen mittels Faktorenanalysen bestimmen, wie viele latente Variablen der mit dem Kriterienraster gemessenen Unterrichtsplanungskompetenz zugrunde liegen und inwiefern die Skalen im Falle von Mehrdimensionalität jeweils in sich und unter einander korrelieren. Damit wäre aber noch nicht geprüft, ob nicht beispielsweise Kriterien enthalten sind, die für das Konstrukt Unterrichtsplanungskompetenz eigentlich irrelevant sind („*construct-irrelevant variance*“; Messick, 1995, S. 742), oder ob Kriterien fehlen, die angesichts des empirischen und theoretischen Forschungsstandes zur Unterrichtsplanung eigentlich enthalten sein müssten („*construct underrepresentation*“; Messick, 1995, S. 742).

Es bedarf also einer globaleren Perspektive, die insbesondere von Michael Kane entwickelt wurde und die letztlich in die *Standards* (AERA et al., 2014) mündete. Demzufolge müssen in einem *argument-based approach* empirische Evidenzen dafür gesammelt werden, dass die Testscores so interpretiert werden können, wie man es anstrebt (Kane, 1992; Kane, 2013). Zu diesem Zwecke kann auf fünf *sources of validity evidence* zurückgegriffen werden:

- *Evidence based on test content* stellt sicher, dass die Kriterien des Kriterienrasters (*test content*) tatsächlich das Konstrukt messen, das sie messen sollen (hier: Qualität schriftlicher Unterrichtsentwürfe). Dies kann zum Beispiel über eine Expertenbefragung sichergestellt werden (AERA et al., 2014), indem erfragt wird, inwiefern Kriterien ergänzt werden müssten, oder ob Kriterien gestrichen werden sollten, die im Allgemeinen nicht in schriftlichen Unterrichtsentwürfen erfüllt werden müssen. Damit würde Messicks (1995) o. g. Einwänden einer möglichen *construct-irrelevance* und einer *construct-underrepresentation* vorgebeugt werden. Darüber hinaus ist insbesondere in der Lehrkräftebildung darauf zu achten, dass es ein „*alignment*“ (AERA et al., 2014, S. 15), d. h. eine Passung zwischen den Kriterien und den zugrunde liegenden Standards gibt, in diesem Fall den Standards für die Lehrerbildung (Tabelle 1; KMK, 2022). Dieser Aspekt wird auch in der Kriterienraster-Literatur betont (z. B. Moskal & Leydens, 2000).
- *Evidence based on internal structure* stellt sicher, dass die mit dem Kriterienraster erhobenen Daten die theoretisch erwartbare Struktur des untersuchten Konstrukts abbilden. Unter interner Struktur werden dabei in der Regel „*dimensionality, measurement invariance, and reliability*“ (Rios & Wells, 2014, S. 108) verstanden. Hinsichtlich der Dimensionalität würde beispielsweise über eine konfirmatorische Faktorenanalyse geprüft werden, ob ein zugrundeliegendes eindimensionales Konstrukt in den Daten auch eine eindimensionale Struktur ergibt. Hinsichtlich der Messinvarianz müsste gezeigt werden, dass statistische Kennwerte wie die Trennschärfe oder der Schwierigkeitsindex in verschiedenen Gruppen oder unter verschiedenen Bedingungen zu vergleichbaren Ergebnissen führen. Hinsichtlich der Reliabilität müsste beispielsweise geprüft werden, ob die interne Konsistenz (Cronbach’s  $\alpha$ ) für die jeweiligen Skalen innerhalb des Kriterienrasters angemessen hoch ist<sup>20</sup>.
- *Evidence based on response processes* stellt sicher, dass bei denjenigen Personen, die das Kriterienraster anwenden, die angestrebten kognitiven Prozesse ausgelöst werden. Anders als bei gewöhnlichen Testinstrumenten, auf die sich die AERA et al. (2014) beziehen, ist in Bezug auf das in dieser Arbeit entwickelte Kriterienraster zu definieren, ob es um die *response processes* derjenigen geht, die das Kriterienraster zur Beurteilung nutzen (z. B. Fachseminarleiterinnen und Fachseminarleiter), oder ob es um die *response processes* der angehenden Lehrkräfte geht, die das Kriterienraster in ihrem Planungsprozess als

---

<sup>20</sup> Ausgehend von der Beobachtung, dass in den Naturwissenschaftsdidaktiken ein inkonsistenter Umgang mit den Begriffen Reliabilität, interne Konsistenz und Cronbach’s  $\alpha$  vorherrscht, zeigt Taber (2018), dass die Orientierung an proklamierten Grenzwerten für ein akzeptables  $\alpha$  wenig ergiebig ist. Stattdessen sollte in jedem Einzelfall theoretisch hergeleitet werden, warum für das jeweilige Instrument ein ermitteltes  $\alpha$  akzeptabel ist oder nicht.

Hilfsmittel oder im Rahmen eines *peer-assessments* (Panadero & Jonsson, 2013; Panadero et al., 2018) bzw. eines *self-assessments* (Krebs et al., 2022) nutzen. Beide Perspektiven sind relevant: Hinsichtlich der Fachseminarleiterinnen und Fachseminarleiter wäre beispielsweise zu prüfen, ob sie die Formulierungen im Kriterienraster verstehen und so auf die Unterrichtsentwürfe anwenden können (*summative assessment*), wie es im Rahmen der Entwicklung des Kriterienrasters durch den Autor dieser Arbeit und eine trainierte studentische Hilfskraft geschehen ist (Kodierleitfaden, siehe Anhang 3). Hinsichtlich der angehenden Lehrkräfte wäre zu prüfen, ob sie die Kriterien verstehen und auf Entwürfe anderer Studierender oder Referendarinnen und Referendare (*peer-assessment*) oder ihre eigenen Entwürfe (*self-assessment*) anwenden können. Dies könnte beispielsweise im Rahmen von Interviews und Lautem Denken (Sandmann, 2014) geprüft werden, ist aber nicht Gegenstand der vorliegenden Dissertation und daher nicht Teil des hier entwickelten Validitätsarguments.

- *Evidence based on relations to other variables* stellt sicher, dass die mit dem Kriterienraster erhobenen Scores in theoretisch erwartbarem Zusammenhang zu anderen Variablen stehen. Auf diese Weise würde geprüft, ob das mit dem Kriterienraster gemessene Konstrukt Unterrichtsplanungskompetenz beispielsweise im erwartbaren Zusammenhang mit PCK steht. Würde man also ein etabliertes PCK-Instrument (z. B. den PCK-IBI; Großschedl et al., 2019) nutzen, um das pPCK angehender Lehrkräfte zu erheben, und zudem das Kriterienraster auf von denselben angehenden Lehrkräften verfasste Unterrichtsentwürfe anwenden, wäre eine positive Korrelation zu erwarten, da PCK auch dem Kriterienraster als theoretische Grundlage dient und von einem starken Zusammenhang zwischen dem pPCK und dem mit dem Kriterienraster erhobenen ePCK<sub>p</sub> ausgegangen werden kann (Alonzo et al., 2019; Carlson et al., 2019). In diesem Fall läge konvergente Validitätsevidenz vor (AERA et al., 2014). Auch wenn das Kriterienraster eher generisch angelegt ist, erfordern die meisten Kriterien doch, dass Überlegungen zur Diagnose der Lerngruppenvoraussetzungen mit dem Lernziel in Beziehung gesetzt werden und reflektiert wird, wie dieses am jeweiligen biologischen Fachinhalt erreicht werden kann. Würde man in gleicher Weise mit einem CK- oder einem PK-Test verfahren, müsste nicht unbedingt angenommen werden, dass eine positive Korrelation zur Planungsqualität besteht. Wenn unter PK nämlich vor allem allgemeine Planungsstrategien, Classroom Management, Wissen über Lerntheorien und Diversität von Schülerinnen und Schülern verstanden wird (vgl. Leijen et al., 2022), ist nicht davon auszugehen, dass derartiges Wissen positiv mit der Anwendung in einer konkreten Planungssituation korrelieren muss, zumal derart allgemeine PK-Variablen (z. B. Classroom Management) nicht als Kriterien in das

Kriterienraster aufgenommen wurden. Zudem zeigen Vogelsang et al. (2022), dass zwar für PCK, nicht aber für CK eine positive Korrelation zur Unterrichtsplanungskompetenz gemessen in Form eines Performanztests im Fach Physik besteht. In einem solchen Fall läge diskriminante Validitätsevidenz vor (AERA et al., 2014). Während derartige Zusammenhänge im Rahmen dieser Arbeit nicht geprüft wurden, wurde als Variable der Ausbildungsstand untersucht – in der Annahme, dass Referendarinnen und Referendare im Staatsexamen aufgrund größerer Praxis- und damit auch Planungserfahrung im Durchschnitt einen signifikant höheren Score erreichen als Master-Studierende im Praxissemester. Durch den Vergleich zweier Teilstichproben aus der 1. und der 2. Phase der Lehrkräftebildung kann geprüft werden, ob das Kriterienraster instruktional sensitiv ist, d. h. ob es auf die erwartbaren Kompetenzunterschiede zwischen den Teilstichproben reagiert.

- *Evidence based on consequences of testing* stellt sicher, dass die Anwendung des Kriterienrasters keine unbeabsichtigten Folgen für die angehenden Lehrkräfte mit sich bringt, deren Unterrichtsentwürfe mit dem Kriterienraster bewertet werden. Folglich sollten nur solche Interpretationen aus den Testscores abgeleitet werden, die mit der transparent zu machenden Intention des Kriterienrasters in Einklang stehen. Kane (2013) spricht in diesem Zusammenhang vom „*interpretation/use argument*“, d. h. dem „*network of inferences and assumptions inherent in the proposed interpretation and use*“ (S. 2). Es muss demnach expliziert werden, „*why, how, by whom and in what contexts*“ (Turley & Gallagher, 2008, S. 87) das Kriterienraster eingesetzt werden soll. Im vorliegenden Fall soll es von Auszubildenden der 1. und 2. Phase der Lehrkräftebildung dazu eingesetzt werden, schriftliche Unterrichtsentwürfe von Studierenden bzw. Referendarinnen und Referendaren zu beurteilen. Die Scores werden folglich als Indikatoren für die Qualität der Unterrichtsentwürfe aufgefasst, aus denen auf die Unterrichtsplanungskompetenz der angehenden Lehrkräfte geschlossen wird. Es muss also geprüft werden, ob ein hoher mit dem Kriterienraster erhobener Score tatsächlich einer hohen Planungsqualität entspricht und vice versa. Anderenfalls hätte die Nutzung des Kriterienrasters beispielsweise im Rahmen von Prüfungen zur Beurteilung von Unterrichtsentwürfen seinen Anspruch verfehlt, die von Krepf und König (2022b) geforderten empirisch geprüften Qualitätskriterien zur Beurteilung von Unterrichtsentwürfen zu liefern. Die Konsequenz, dass angehende Lehrkräfte aufgrund eines ungeeigneten Instruments im Rahmen ihrer Ausbildung unfair beurteilt werden, wäre eine unitendierte Konsequenz und müsste auf dem Wege einer empirischen Überprüfung ausgeschlossen werden. Daher wird über einen Vergleich zwischen holistisch ermittelnden Noten zur Beurteilung der Qualität der

Unterrichtsentwürfe einerseits und den analytisch mithilfe des Kriterienrasters ermittelten Scores sichergestellt, dass die Scores tatsächlich Qualität ausdrücken (Tomas et al., 2019).

Für vier der fünf genannten *sources* wurden in Beitrag 7 empirische Evidenzen gesammelt und auf diese Weise ein Validitätsargument entwickelt, das plausibel nahelegt, dass die Testscores des Kriterienrasters tatsächlich eine valide Beurteilung der Qualität von Unterrichtsentwürfen und damit eine Messung der Unterrichtsplanungskompetenz angehender Biologielehrkräfte erlauben.

#### **4.2.2. Beitrag 7: *Assessing the Quality of Science Teachers' Lesson Plans: Evaluation and Application of a Novel Instrument***

Beitrag 7 wurde publiziert: Großmann, L. & Krüger, D. (2023). Assessing the Quality of Science Teachers' Lesson Plans: Evaluation and Application of a Novel Instrument. *Science Education*, 1-37. <https://doi.org/10.1002/sce.21832>

#### **Zusammenfassung**

In diesem Beitrag wurden die folgenden Forschungsfragen untersucht:

- F1.1. To what extent do science teacher educators consider the RALP (*Rubric to Assess Science Lesson Plans*) criteria relevant for writing lesson plans? (*Validity evidence based on test content*)
- F1.2. To what extent does the scoring rubric allow for reliable measures?
  - a. To what extent is the RALP scoring procedure stable over a two-week interval when applying the scoring rubric to PSTs (pre-service science teachers)' and TSTs' (trainee science teachers') lesson plans? (*Intra-rater reliability*)
  - b. To what extent do teacher educator and a trained student assistant reach an intersubjective agreement when applying the RALP to TSTs' lesson plans? (*Inter-rater reliability*)
- F1.3. To what extent do data reflect the intended interconnectedness of PCK components? (*Validity evidence based on internal structure*)
- F1.4. To what extent is the selection of criteria appropriate to assess lesson plan quality? (*Validity evidence based on internal structure*)
- F1.5. To what extent do the RALP scores correlate with science teacher educators' holistic quality assessment? (*Validity evidence based on consequences of testing*)
- F1.6. To what extent do PSTs and TSTs achieve different scores? (*Validity evidence based on relations to other variables*)
- F2. What are the most significant differences between PSTs' and TSTs' lesson plans?



Wie den sieben Fragestellungen zu entnehmen ist, liegt Beitrag 7 eine zweiteilige Struktur zugrunde: Die Fragen F1.1-F1.6 dienen der Entwicklung des Validitätsarguments unter Berücksichtigung von vier der fünf *sources of validity evidence* (AERA et al., 2014). Frage F2 knüpft daran an und dient dazu, qualitative Einblicke in die Unterrichtsentwürfe zu geben, indem vergleichend zu den drei Kriterien mit dem größten Unterschied im Score zwischen Master-Studierenden sowie Referendarinnen und Referendaren exemplarisch analysiert wird, inwiefern die Kriterien berücksichtigt wurden. Auf diese Weise wird illustriert, dass die Graduierung in Niveaustufen im Kriterienraster geeignet ist, um die Qualitätsunterschiede in den Entwürfen zu erfassen.

Als Voraussetzung für Validität wurde zunächst die Stabilität des Auswertungsverfahrens (Intra-Rater-Reliabilität) und die intersubjektive Nachvollziehbarkeit (Inter-Rater-Reliabilität) überprüft. Im Durchschnitt ergab sich jeweils ein nahezu perfektes Cohen's  $\kappa$  ( $\kappa_{intra} = .93$ ;  $\kappa_{inter} = .88$ ; Landis & Koch, 1977). Die Ergebnisse der Fragen 1.1-1.6 werden in Tabelle 4 zusammengefasst.

**Tabelle 4**

*Übersicht über das entwickelte Validitätsargument*

<i>Validity evidence based on...</i>	Fragestellung	Ergebnis
<i>test content</i>	Kriterien relevant? (F1.1)	Ja, nur die Nutzung von Literatur in der didaktischen Analyse sowie die Berücksichtigung von Schülervorstellungen wurde von Fachseminarleiterinnen und Fachseminarleitern als eher irrelevant eingeschätzt.
<i>internal structure</i>	Dimensionalität? (F1.3)	Eindimensionales Modell erklärt die Daten signifikant besser als das zweidimensionale Modell ( $\chi^2 [2] = 7.72, p < .05$ ); kein signifikanter Unterschied zum vier- und fünfdimensionalen Modell; Cronbach's $a = .78$
<i>internal structure</i>	Trennschärfe? (F1.4)	Ja, 20/24 Kriterien sind hinreichend trennscharf ( $r_{i(i-1)} > .30$ ).
<i>internal structure</i>	Itemschwierigkeit? (F1.4)	Angemessene Verteilung der Kriterien in der Wright-Map; Entwicklung eines Erwartungshorizonts ist zu leicht, Berücksichtigung von Schülervorstellungen zu schwer in dieser Stichprobe.
<i>consequences of testing</i>	Hoher Score = hohe Planungsqualität? (F1.5)	Ja, es gibt eine starke Korrelation zwischen den Scores und den holistisch erzeugten Noten zur Einschätzung der Planungsqualität sowohl bei den Master-Studierenden ( $r = -.69, p < .001$ ) als auch bei den Referendarinnen und Referendaren ( $r = -.72, p < .001$ ).
<i>relations to other variables</i>	Instruktionale Sensitivität? (F1.6)	Ja, Referendarinnen und Referendare ( $Mdn = 28$ ) erreichen im Durchschnitt einen signifikant höheren Score als Master-Studierende ( $Mdn = 20$ ; $U = 421, z = -5.26, d = 1.23, p < .001$ ; großer Effekt).

Insgesamt gibt es überzeugende Evidenz dafür, dass entsprechend dem vorab formulierten „*interpretation/use argument*“ (Kane, 2013, S. 2) davon ausgegangen werden kann, dass das Kriterienraster dazu geeignet ist, die Qualität schriftlicher Unterrichtsentwürfe valide

einzuschätzen. Durch die Rückmeldung der Fachseminarleiterinnen und Fachseminarleiter zu F1.1 ist sichergestellt, dass das Kriterienraster nur solche Kriterien abdeckt, die am Ende der deutschen Lehrkräftebildung im Rahmen der Staatsexamensprüfung erfüllt werden sollten. Insbesondere die Ergebnisse zu F1.5 sind relevant, da sie in der Tat dafür sprechen, dass ein hoher Testscore tatsächlich einer hohen Planungsqualität entspricht. In Ergänzung bestätigen die Ergebnisse zu F1.6 die Annahme, dass Referendarinnen und Referendare im Durchschnitt signifikant bessere Unterrichtsentwürfe schreiben als Master-Studierende, so dass insgesamt ein plausibles Validitätsargument entwickelt werden konnte.

### 4.3. Erträge für die Lehrkräftebildung

#### 4.3.1. Beitrag 8: *Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen*

Beitrag 8 wurde publiziert: Großmann, L. & Krüger, D. (2022a). Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen, *SEMINAR – Lehrerbildung und Schule*, 2022 (1), 91-110.

#### **Zusammenfassung**

In diesem Beitrag wird das in Beitrag 6 entwickelte und in Beitrag 7 überprüfte Kriterienraster für Auszubildende der 1. und 2. Phase der Lehrkräftebildung vorgestellt. Er lässt sich insofern als Synopse dieses Dissertationsvorhabens verstehen, als ausgehend von der hier angewandten Operationalisierung von Unterrichtsplanungskompetenz (Abb. 3) die in der Unterrichtsplanung zu bewältigenden Probleme analysiert und darauf aufbauend das Kriterienraster für Dozierende sowie Fachseminarleiterinnen und Fachseminarleiter vorgestellt wird. Adressatengerecht wurde hierbei eine Bezugnahme auf den PCK-Diskurs und somit auf die theoretische Grundlage vermieden. Stattdessen wurde u. a. durch erläuternde Ausführungen zu den fünf Abschnitten innerhalb des Kriterienrasters (Tabelle 1 in Beitrag 8) sowie durch die Angabe überwiegend deutschsprachiger Belege für die Relevanz der 24 Kriterien versucht, an die derzeitige Ausbildungspraxis anzuknüpfen.

#### 4.3.2. Beitrag 9: *Probleme beim Planen?! Acht Ratschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie*

Beitrag 9 wurde publiziert: Großmann, L., Lotz, A., Mulke, S. & Krüger, D. (2022). Probleme beim Planen?! Acht Ratschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie, *MNU-Journal*, 75 (05), S. 368-376.

##### **Zusammenfassung**

Dieser Beitrag richtet sich explizit an angehende Biologielehrkräfte der 1. und 2. Phase der Lehrkräftebildung. Die 24 Kriterien des Kriterienrasters werden nicht tabellarisch, sondern in Form von Leitfragen präsentiert, die dazu anregen sollen, die eigenen Unterrichtsentwürfe kritisch zu reflektieren. Erneut wird das Problemlösen als zentrale Herausforderung in der Unterrichtsplanung fokussiert, indem die 24 Kriterien insgesamt acht wesentlichen Problemen zugeordnet werden, die sich in der Analyse der Staatsexamensentwürfe (Beiträge 2-5) gezeigt haben. Sie entsprechen den kognitiven Herausforderungen in der Unterrichtsplanung laut dem CODE-PLAN-Modell (König et al., 2021a), ohne dass dies jedoch im Beitrag explizit ausgeführt wird. Statt der in Beitrag 8 geleisteten eher theoretischen Herleitung der Kriterien wird in Beitrag 9 anhand realer Beispiele aus den Staatsexamensentwürfen gezeigt, wie die Probleme in der Unterrichtsplanung gelöst werden können. Zu diesem Zwecke werden erstens weniger gelungene Beispiele für die acht Probleme gezeigt, zweitens Optimierungsvorschläge durch Streichungen oder Ergänzungen unterbreitet, und diese drittens zum besseren Verständnis erläutert. Auf diese Weise wird anhand ökologisch valider Beispiele an die Probleme angeknüpft, auf die angehende Lehrkräfte beim Durchlaufen ihres Planungsprozesses (Abb. 1) stoßen. Dies ist auch insofern innovativ, als es erstaunlicherweise für manche planungsrelevante Aspekte wie z. B. Sequenz- bzw. Reihenplanungen von Meisert (2016) abgesehen keine publizierten Beispiele gibt, an denen sich angehende Lehrkräfte orientieren können. Gleiches gilt für die Formulierung von Indikatoren zum Nachweis des Kompetenzzuwachses von Schülerinnen und Schülern, die zwar vor allem in der englischsprachigen Didaktik weit verbreitet ist (z. B. Anderson & Kratwohl, 2014; Wiggins & McTighe, 2005), für die es aber bislang noch keine praxisnahe Erläuterung mit Beispielen gibt. An dieser Stelle knüpft Beitrag 9 an, der angehenden Biologielehrkräften (und solchen anderer Fächer) eine Hilfe beim Planen sein kann. Der Vorzug dieses Beitrags liegt gewiss darin, dass hier neben dem Verfasser dieser Arbeit und seinem Betreuer, also Dozierenden aus der 1. Phase der Lehrkräftebildung, auch zwei Fachseminarleiter aus der 2. Phase der Lehrkräftebildung als Autoren mitgewirkt haben. Auf diese Weise ist sichergestellt, dass der Beitrag auch die Bedürfnisse von Referendarinnen und Referendaren in der 2. Phase aufgreift.

#### 4.4. Zusammenfassende Diskussion

In diesem Abschnitt werden die Beiträge 6, 7, 8 und 9 gemeinsam diskutiert. Dabei werden zunächst wesentliche Limitationen der Untersuchungen erörtert, ehe der Erkenntnisgewinn der vier Beiträge kritisch diskutiert und die Befunde an den Unterrichtsplanungsdiskurs angebunden werden.

##### *Limitationen*

Traditionell wurden Kriterienraster als *scoring rubrics* in Form von summativem Assessment genutzt, um Prüfungsleistungen bewerten zu können. Heutzutage werden Kriterienraster jedoch vor allem als *instructional rubrics* in Form von formativem Assessment genutzt, um Lernprozesse zu begleiten, qualitatives Feedback zu geben und dabei zugleich die Erwartungen an das zu entwickelnde Lernprodukt transparent zu machen (Andrade, 2005). Das im Rahmen dieser Dissertation entwickelte Kriterienraster entspricht der erstgenannten Perspektive und scheint demnach hinter die aktuellere Literatur (Brookhart, 2018) zurückzufallen. Dabei ist jedoch zweierlei zu berücksichtigen: Erstens ist es ungeachtet der empirisch nachgewiesenen, lernförderlichen Wirkung des Einsatzes von Kriterienrastern als *instructional rubrics* nach wie vor plausibel, Kriterienraster in bestimmten Situationen als *scoring rubrics* einzusetzen und auf diese Weise eine Beurteilung vorzunehmen. In Bezug auf die Unterrichtsplanungsforschung im Allgemeinen und schriftliche Unterrichtsentwürfe im Besonderen wurde einleitend dargelegt, dass es derzeit noch keine empirisch geprüften Qualitätskriterien gibt, mit denen schriftliche Unterrichtsentwürfe bewertet werden können (Krepf & König, 2022b). Einerseits könnte das Kriterienraster im Rahmen der Lehrkräftebildung eingesetzt werden und somit der wiederholt bemängelten, fehlenden Konsistenz und Transparenz der Beurteilungspraxis vor allem in der 2. Phase der Lehrkräftebildung entgegenwirken (Döbrich & Abs, 2008; Kärner et al., 2019; Strietholt & Terhart, 2009). Andererseits würde es der empirischen Forschung die Möglichkeit eröffnen, über einen qualitativ-analytischen Zugang zu Unterrichtsentwürfen ein quantitatives Maß (den Score) zu ermitteln und damit die Planungsqualität des Datenmaterials numerisch darstellen zu können. Kriterienraster zum summativen Assessment sind also nicht per se obsolet.

Neben diesem grundsätzlichen Einwand sind sowohl hinsichtlich der Entwicklung bzw. der Gestaltung des Kriterienrasters als auch hinsichtlich seiner Überprüfung einige spezifischere Einwände zu berücksichtigen. Hinsichtlich der Gestaltung des Kriterienrasters lassen sich die folgenden vier Aspekte kritisch betrachten:

Erstens erhebt das Kriterienraster den Anspruch, so generisch formuliert zu sein, dass es auf Unterrichtsentwürfe zu sämtlichen Themen und Kompetenzbereichen anwendbar ist. Ausgehend

von der Kompetenzauffassung von Blömeke et al. (2015) wird davon ausgegangen, dass aus der Performanz – also dem schriftlichen Unterrichtsentwurf – auf die Kompetenz – also die Unterrichtsplanungskompetenz – geschlossen werden kann. Hierbei ist jedoch zu beachten, dass in der generischen Kompetenzdefinition von Klieme et al. (2007) die Kontextgebundenheit von Kompetenz betont wird. Analog wird wie oben dargestellt davon ausgegangen, dass Unterrichtsplanung nicht nur fachspezifisch, sondern auch genuin themenspezifisch ist (König et al., 2020b; Scholl et al., 2019). Schließt man also aus der Analyse eines einzigen Unterrichtsentwurfs einer angehenden Lehrkraft auf ihre Unterrichtsplanungskompetenz, so ist diese Interpretation insofern mit Skepsis zu betrachten, als die Analyse eines weiteren Unterrichtsentwurfs zu einem anderen Thema, in einer anderen Lerngruppe und mit einer anderen angestrebten Kompetenzentwicklung bereits zu ganz anderen Ergebnissen führen kann. Hierin liegt eine grundsätzliche methodische Herausforderung in der Kompetenzforschung. In Bezug auf die Unterrichtsplanungskompetenz im Sinne der oben dargelegten Konzeptualisierung in dieser Arbeit (Abb. 3) sind in diesem Zusammenhang drei Aspekte zu beachten:

Zum einen erfordert Unterrichtsplanung Professionswissen als Disposition (z. B. Riese et al., 2022), also PK, CK sowie PCK. Es ist davon auszugehen, dass angehende Lehrkräfte zu den verschiedenen Themen der Biologie über unterschiedlich stark ausgeprägtes CK verfügen, das wiederum einen Einfluss auf den Planungsprozess und damit das Planungsprodukt hat. In der Multidimensionalität des Konstrukts, das hier erfasst werden soll, ist die Problematik der Themenspezifität also bereits angelegt und lässt sich forschungsmethodisch nicht kontrollieren, wenn im Sinne von Kang (2017) Unterrichtsplanung in authentischen Kontexten – also ökologisch valide (Wegener & Blankenship, 2007) – erforscht werden soll. Im Kriterienraster selbst wird die Themenspezifität nicht explizit formuliert, ist jedoch im Begriff *Funktionalität* implizit vorhanden: Um beispielsweise potentielle Lernschwierigkeiten antizipieren zu können, bedarf es belastbaren fachlichen Wissens – also CK –, ohne dass eine solche Planungsoperation nicht vorgenommen werden könnte und ohne dass fraglich wäre, inwiefern die geplante Stunde den Schülerinnen und Schülern adaptiv einen Kompetenzzuwachs ermöglicht.

Zum anderen lässt sich aus dem RCM (Carlson et al., 2019) ableiten, dass es einen Unterschied zwischen dem grundsätzlich verfügbaren PCK einer Person – dem pPCK – und dem in einer konkreten Situation angewandten Teil des pPCK – dem ePCK – gibt (Alonzo et al., 2019). Das Kriterienraster erfasst ePCK<sub>p</sub>, also das dem Planungsprodukt zu entnehmende angewandte PCK, das zum Kreieren und Begründen (Vogelsang & Riese, 2017) genutzt wird. Inwieweit die dem internationalen Professionsforschungsdiskurs entstammenden Konstrukte ePCK oder pPCK mit dem vor allem im deutschsprachigen Raum etablierten Kompetenzbegriff gleichzusetzen sind, und ob unter Kompetenz das ePCK, das pPCK oder beides verstanden wird, ist in der Literatur

bislang noch ungeklärt. In dieser Arbeit wird argumentiert, dass in Unterrichtsentwürfen das ePCK<sub>p</sub> im Sinne situationsspezifischer Fähigkeiten (Kreation, Legitimation; Vogelsang & Riese, 2017; Abb. 1; Abb. 3) erfasst wird. Das Kriterienraster kann diese beiden *realms* von PCK nicht unterscheiden, was angesichts der Kompetenzauffassung von Blömeke et al. (2015) nicht als problematisch angesehen wird: Aus der Performanz (dem schriftlichen Unterrichtsentwurf) kann nichtsdestotrotz auf Unterrichtsplanungskompetenz geschlossen werden. Zur Individualdiagnostik ist aufgrund der Fachspezifität von zu bedenken, dass eine einmalige Messung einem *bias* unterliegt. Es wäre möglich, beispielsweise im Rahmen einer längsschnittlichen Untersuchung zum Beispiel innerhalb einer thematisch zusammenhängenden Unterrichtssequenz mehrere Unterrichtsentwürfe entwickeln zu lassen und mit dem Kriterienraster zu untersuchen. Dann wäre sowohl die thematische Variation als Störvariable kontrolliert als auch untersuchbar, wie groß die Streuung der mit dem Kriterienraster ermittelten Scores ist.

Darüber hinaus entstammen die Kriterien des Kriterienrasters der Analyse von Staatsexamensentwürfen (Beitrag 3) und bilden somit auch diejenigen Aspekte ab, die in diesem Datenmaterial gefunden wurden. Es ließe sich zurecht einwenden, dass das PCK zwar als fachdidaktisches Professionswissen die theoretische Grundlage der Kriterien bildet, dass die Kriterien selbst aber kaum spezifisch biologiedidaktisch, sondern eher generisch formuliert sind und gleichermaßen in anderen Fächern eingesetzt werden könnten. Dem ist gewiss so, und ein Grund dafür liegt auch darin, dass in den untersuchten Staatsexamensentwürfen kaum biologiespezifische unterrichtsplanungsrelevante Kriterien identifiziert werden konnten. Wenn Neuhaus (2021) als biologiespezifische Unterrichtsqualitätsmerkmale beispielsweise den Einsatz realer Objekte (also Lebewesen) benennt, in den Staatsexamensstunden allerdings ausnahmslos nicht mit Lebewesen gearbeitet wird, dann taucht dieser Aspekt auch nicht als eigenes Kriterium im Kriterienraster auf, sondern ist unter dem generisch formulierten Kriterium „Medieneinsatz“ subsumiert. Hinsichtlich der Biologiespezifität ist zudem zu bedenken, dass PCK in der Forschung zwar oft als domänen-, disziplin-, themen- und konzeptspezifisch definiert wird (Abb 2: Carlson et al., 2019; vgl. Behling et al., 2022; Gess-Newsome, 2015), ohne dass diese Kategorien jedoch immer trennscharf unterschieden werden. In den in Beitrag 3 untersuchten Staatsexamensentwürfen wurden Planungsentscheidungen überwiegend eher generisch beschrieben und begründet (z. B. warum eine bestimmte Aufgabe zum Thema eher in Gruppen- als in Partnerarbeit bearbeitet werden soll), so dass sich dies auch im Kriterienraster widerspiegelt. Spezifischere Aspekte zu einzelnen Themen, Kompetenzbereichen oder Unterrichtsmethoden würden die Einsatzmöglichkeiten des Kriterienrasters deutlich einschränken. Zudem ist zu beachten, dass das Kriterienraster (Anhang 3) für die meisten Kriterien sehr wohl eine Niveauabstufung vornimmt,

bei der allgemein-pädagogische Ausführungen der Niveaustufe 1 und fachdidaktische Ausführungen der Niveaustufe 2 zugeordnet werden. So erfordern K1, K2 und L3 beispielsweise eine sinnvolle Abstimmung inhalts- und prozessbezogener Kompetenzen, so dass ohne eine sachangemessene Durchdringung und Strukturierung des biologischen Fachinhalts die höchste Niveaustufe nicht erreicht werden kann. I1-I4 beziehen sich allesamt auf eine fachdidaktische Perspektive, die zum jeweiligen Thema einzunehmen ist. S3 und S4 beziehen sich auf Alltagsvorstellungen und Lernschwierigkeiten, die den jeweiligen biologischen Themen innewohnen, und die meisten Instruktionsstrategien wie die Auswahl von Aufgaben (L4) oder Medien (L5) sind erst dann auf der höchsten Niveaustufe ausgeprägt, wenn sie explizit fachdidaktisch beschrieben und begründet werden. Eine noch spezifischere Benennung eher biologiespezifischer Qualitätsmerkmale (z. B. zum Umgang mit Lebewesen, zum empirischen Arbeiten) würde das Kriterienraster in seiner Anwendbarkeit einschränken, da diese Aspekte nicht in sämtlichen Biologiestunden zum Tragen kommen. Aus der Perspektive der neueren praxeologischen Unterrichtsforschung betrachtet lässt sich das hier skizzierte und im Rahmen dieser Arbeit nicht zu lösende Problem im Sinne von Breidenstein (2021) als Interferenz verschiedener Praktiken verstehen: Im Biologieunterricht und auch im methodischen Vorgehen in den Beiträgen 3 und 7 überlagern sich einerseits die fachliche Dimension schulischen Lernens (d. h. die genuin biologiespezifischen Unterrichtsgegenstände; Kompetenzbereich Fachwissen bzw. Sachkompetenz) und die anwendungsbezogenen Problemlösefähigkeiten, die gleichermaßen auch in anderen Fächern relevant sind und gefördert werden sollen (d. h. die eher generischen Praktiken wie das Auswerten von Diagrammen im Kompetenzbereich Kommunikation, das Formulieren von Hypothesen im Kompetenzbereich Erkenntnisgewinnung oder die Bildung eines ethischen Urteils im Kompetenzbereich Bewertung). Inwieweit generische und fachspezifische Praktiken interferieren und inwiefern es sich dabei um störende oder um verstärkende Interaktionen handelt, ist aktuell noch ungeklärt (Breidenstein, 2021). In dieser Arbeit wurde implizit angenommen, dass sich beide Praktiken bedingen und notwendigerweise ergänzen, so dass sowohl in Beitrag 3 als auch in Beitrag 7 keine Beschränkung auf ausschließlich biologiespezifische Aspekte vorgenommen wurde.

Zweitens liegt dem Kriterienraster implizit, jedoch nicht explizit, die vor allem in der deutschsprachigen Unterrichtsqualitätsforschung etablierte heuristische Unterscheidung zwischen Oberflächen- und Tiefenstrukturen (Decristan et al., 2020; Oser & Baeriswyl, 2001) zugrunde. So zielen die meisten Kriterien auf die Frage ab, inwieweit die jeweiligen Planungsentscheidungen funktional für das Erreichen des jeweiligen Ziels sind (Intensitätsskala; Rohrmann, 1978), so dass für die Beurteilung der Planungsqualität weniger die methodisch-didaktische Gestaltung auf

Oberflächenstrukturebene, sondern ihr Beitrag zur angestrebten Kompetenzförderung auf Tiefenstrukturebene adressiert wird (Beitrag 6). Die zu diesem Zwecke entwickelten Kriterien sind allesamt aus der empirischen und theoretischen *fachdidaktischen* Literatur hergeleitet. Wie oben bereits ausgeführt, steht die Unterrichtsplanung aber in einem Spannungsfeld aus Pädagogik (PK) und Fachdidaktik (CK, PCK), so dass eine konsequenter Berücksichtigung von in der Allgemeindidaktik beschriebenen und empirisch geprüften Merkmalen qualitätvollen Unterrichts (z. B. Helmke, 2017) möglich gewesen wäre. In der Unterrichtsqualitätsforschung zeigt sich, dass das etablierte Modell der drei Basisdimensionen *Klassenführung*, *Konstruktive Unterstützung* und *Kognitive Aktivierung* (Klieme et al., 2001) möglicherweise trotz seiner empirisch nachgewiesenen Stärken sowohl um generische Aspekte (z. B. Assessment/Feedback: Taut & Rakoczy, 2016; Üben: Schlesinger et al., 2018) als auch um fachspezifische Subdimensionen (z. B. fachliche Korrektheit: Brunner, 2018; Kohärenz der Inhalte: Lipowsky et al., 2018) erweitert werden könnte (Praetorius et al., 2020). Da die vor allem in der Allgemeindidaktik entwickelten Modelle zur Unterrichtsqualität überwiegend generisch gestaltet sind, müssten die zu den jeweiligen Subdimensionen von Unterrichtsqualität vorhandenen Indikatoren fachspezifisch angepasst werden, um den jeweiligen Lernzielen, Lerngegenständen und fachspezifischen Lehr-Lernprozessen Rechnung zu tragen (Praetorius et al., 2020). Dergleichen ist in der empirischen fachdidaktischen Unterrichtsqualitätsforschung jedoch erst noch zu entwickeln (Begrich et al., 2023) und war nicht Gegenstand dieser Dissertation.

Drittens sind weder im Kriterienraster selbst noch in dem an Auszubildende gerichteten Beitrag 8 Hinweise gegeben, wie viele Punkte im Kriterienraster zu welcher Note führen sollten. Dies stellt ein häufiges Manko von Kriterienrastern dar, die im Sinne summativen Assessments eingesetzt werden (Chan & Ho, 2019). Dabei ist jedoch zu bedenken, dass die Festlegung derartiger Ranges (z. B. 42-48 Punkte: Note 1; 36-41 Punkte: Note 2 usw.) nicht willkürlich sein sollte. Die Anwendung des Kriterienrasters (Beitrag 7) hat gezeigt, dass eine starke Korrelation zwischen den Noten des Fachseminarleiters und dem Score besteht, dass zugleich jedoch der höchste erreichte Score 39 von 48 möglichen Punkten war. Das bedeutet, dass selbst am oberen Ende der Qualitätsskala noch fast 20 % der Punkte nicht erreicht wurden. Wollte man also eine kriteriale Bezugsnorm anlegen, sollte den empirischen Befunden entsprechend bereits in diesem Punktebereich die höchste Note gegeben werden. Allerdings ist zu berücksichtigen, dass das Kriterienraster weder den Studierenden noch den Referendarinnen und Referendaren in der untersuchten Stichprobe vorab bekannt war und dass der langfristige Zweck des Kriterienrasters darin liegt, es vorab transparent zu machen und somit Leistungserwartungen im Vorhinein mitzuteilen. Dann würden gewisse Aspekte, die vorab nicht mitgedacht worden wären,



berücksichtigt, so dass sich die Gesamtverteilung der erreichten Scores vermutlich in Richtung des Maximalscores verschieben würde. Aus diesem Grund wird von einer Angabe solcher Notenbereiche, wie sie Chan und Ho (2019) vorschlagen, sowohl in den Beiträgen 7 und 8 als auch in dieser Rahmenschrift abgesehen.

Viertens hängt damit die Frage zusammen, inwiefern alle Kriterien im Kriterienraster gleichermaßen zum Gesamtscore beitragen sollten. Es ist nicht unüblich, innerhalb von Kriterienrastern einzelne Kriterien stärker zu gewichten als andere (vgl. Brookhart, 2018). Theoretisch ließe sich plausibel begründen, dass im Zuge der Unterrichtsplanung beispielsweise die Formulierung eines klaren Ziels bedeutsamer ist als die Entwicklung eines Erwartungshorizonts für die Aufgabenstellungen. Eine Gewichtung der Kriterien sollte jedoch evidenzbasiert vorgenommen werden. So wäre zu entscheiden, wie nuanciert die Gewichtung sein soll (einfach, zweifach, dreifach gewichtete Kriterien?) und welche Kriterien jeweils wie gewichtet werden sollen. Diese Zuordnung sollte nicht subjektiv erfolgen. Die Überprüfung des Kriterienrasters (Beitrag 7) zeigte jedoch beispielsweise, dass die Fachseminarleiterinnen und Fachseminarleiter der Berücksichtigung von Schülervorstellungen geringe Bedeutung beigemessen haben, während sie in der naturwissenschaftsdidaktischen Forschung nach wie vor große Relevanz hat. Abhängig davon, ob man nun aus akademisch-universitärer Perspektive (1. Phase der Lehrkräftebildung) oder unterrichtspraktischer Perspektive (2. Phase der Lehrkräftebildung) argumentiert, würde man das Kriterium zu Schülervorstellungen (S3, Anhang 3) unterschiedlich stark gewichten. Hier wäre es notwendig, einen größeren Kreis an Personen aus beiden Phasen der Lehrkräftebildung zur Relevanz der Kriterien zu befragen, um eine belastbarere Einschätzung zu erhalten. Dies ist im Rahmen dieser Arbeit nicht geschehen, so dass von einer Gewichtung der Kriterien derzeit noch abgesehen wird.

Neben diesen Aspekten zur Entwicklung bzw. Gestaltung des Kriterienraster lassen sich hinsichtlich seiner Überprüfung die folgenden fünf Aspekte kritisch betrachten:

Erstens wurde zwar versucht, die Kriterien so objektiv wie möglich zu formulieren und dem Anspruch gerecht zu werden, allgemeingültige empirisch geprüfte Qualitätskriterien (Krepf & König, 2022b) zu entwickeln. Allerdings basiert die Prüfung auf *validity evidence based on test content* auf der Rückmeldung eines Professors aus Berlin (dem Betreuer dieser Arbeit), einer Professorin für Biologiedidaktik aus Niedersachsen und einem Professor für Biologiedidaktik aus Nordrhein-Westfalen (1. Phase der Lehrkräftebildung) sowie von zehn Fachseminarleiterinnen und Fachseminarleitern aus dem Land Berlin (2. Phase der Lehrkräftebildung). Es ist nicht auszuschließen, dass eine breitere Beteiligung von Personen in der Lehrkräftebildung dazu beigetragen hätte, weitere Kriterien zu ergänzen, andere zu streichen oder umzuformulieren.

Zumindest für das Land Berlin kann behauptet werden, dass die 24 Kriterien die dort in der 2. Phase der Lehrkräftebildung relevanten Aspekte abdecken. Hinsichtlich der Fachseminarleiterinnen und Fachseminarleiter ist zudem zu bedenken, dass Bewertungen im Referendariat allzu oft nicht anhand transparenter und vergleichbarer Kriterien vorgenommen wird (Döbrich & Abs, 2008; Kärner et al., 2019; Strietholt & Terhart, 2009), so dass eine strengere Strichprobenziehung notwendig gewesen wäre. Anders als in anderen Bundesländern stellt die Leitung eines Fachseminars in Berlin kein Beförderungsamt dar, so dass auch Lehrkräfte ohne viel Berufserfahrung und ausgewiesene und objektiv nachweisbare fachdidaktische Kompetenz als Fachseminarleiterin bzw. Fachseminarleiter in Berlin tätig sein können.

Zweitens wohnt dem Abgleich zwischen den mithilfe des Kriterienrasters analytisch gewonnenen Scores einerseits und dem holistischen Urteil eines Fachseminarleiters andererseits eine Schwierigkeit inne: Das Kriterienraster wurde aufgrund der empirisch belegten Beobachtung entwickelt, dass die Bewertung von Prüfungsleistungen durch Fachseminarleiterinnen und Fachseminarleiter oft intransparent und inkonsistent ist oder zumindest so empfunden wird (Döbrich & Abs, 2008; Kärner et al., 2019; Strietholt & Terhart, 2009). Dass die Bewertung der Staatsexamensentwürfe sowohl in Beitrag 3 als auch in Beitrag 7 nun wiederum von einem solchen Fachseminarleiter vorgenommen wird, dessen Expertise-Status im Beitrag allein mit seiner langen Berufserfahrung begründet wird, ist kritisch zu sehen. Da sich Expertise nicht in erster Linie in Berufserfahrung ausdrückt, sondern in der Qualität des verfügbaren Professionswissens und der Nutzung desselben in problemhaltigen Situationen, ist die Güte der Noten, die hier als Validitätsevidenz große Bedeutung haben, mit Vorsicht zu bewerten. Zur Prüfung der Verlässlichkeit des Fachseminarleiters wurde jedoch in Beitrag 3 eine kleine Teilstichprobe von einem zweiten Fachseminarleiter bewertet. Dabei zeigt sich eine hohe Korrelation zu den Noten des ersten Fachseminarleiters. In Beitrag 7 ergab auch der Vergleich zwischen den Noten des Fachseminarleiters und den originalen Staatsexamensnoten von 17 Entwürfen eine hohe Korrelation, so dass intersubjektive Nachvollziehbarkeit zumindest teilweise gewährleistet ist.

Drittens wird in der empirischen Forschung zu Kriterienrastern gelegentlich mithilfe explorativer oder konfirmatorischer Faktorenanalysen geprüft, inwiefern die Daten die theoretisch zu erwartende Struktur des Konstrukts abbilden (Brookhart, 2018). Aufgrund des Pentagon-Modells (Park & Oliver, 2008) wäre es zum Beispiel plausibel, eine fünf-dimensionale Struktur anzunehmen. Eine solche Faktorenanalyse wurde zwar durchgeführt, aufgrund eines zu geringen Kaiser-Meyer-Olkin-Werts (.57, *miserable*; Kaiser & Rice, 1974) und sehr geringer Kommunalitäten vieler Items wurde sie jedoch in Beitrag 7 nicht berichtet. Grundsätzlich gilt: „[F]actor analysis is a large-sample procedure“ (Norman & Streiner, 2014, p. 223), so wird für explorative Faktorenanalysen ein

Verhältnis von Stichprobengröße zur Anzahl der Items von 5:1, 10:1 oder 20:1 vorgeschlagen (z. B. Mundfrom et al., 2005). Zusätzlich wäre jedoch sicherzustellen, dass jeder Faktor überdeterminiert ist, dass also pro Faktor mehrere Items entwickelt werden (Fabrigar et al., 1999). Watkins (2018) fasst zusammen, dass sowohl die Stichprobengröße als auch die Anzahl der Items pro Faktor wichtige Voraussetzungen für eine Faktorenanalyse darstellen. Zwar wäre eine Erhöhung der Stichprobengröße theoretisch möglich gewesen. Bei 24 Kriterien würde die Empfehlung, ein 10:1-Verhältnis von Stichprobengröße zu Kriterien sicherzustellen, jedoch bereits die Analyse von 240 Entwürfen, für das Verhältnis von 20:1 sogar von 480 Entwürfen erfordern – in Beitrag 7 liegt das Verhältnis bei 100:24, d. h. die Anzahl der untersuchten Entwürfe entspricht dem 4,2-fachen der Anzahl an Kriterien. Eine beträchtlich größere Anzahl an Entwürfen war in den beiden Teilpopulationen nicht zu erhalten und wäre im Rahmen dieses Dissertationsvorhabens aus zeitökonomischen Gründen auch nicht zu analysieren gewesen. Hinsichtlich der Anzahl der Items pro Faktor ist Folgendes zu bedenken: Anders als Leistungstests oder Fragebögen zielt das Kriterienraster nicht auf eine quantitative, sondern auf eine qualitative Auswertung ab. Es würde den Zweck des Kriterienrasters konterkarieren, wenn für die verschiedenen Kriterien jeweils mehrere Items entwickelt werden würden, um eine Überdetermination herzustellen. Dann würden Aspekte mehrfach gewertet, was weder für das summative noch für das formative Assessment einen Mehrwert liefert.

Aus diesen Gründen hätte eine Faktorenanalyse das Validitätsargument (*validity evidence based on internal structure*) für das Kriterienraster nicht gestärkt. Da für jedes Instrument sorgsam zu prüfen ist, welche Validitätsevidenzen notwendig sind und welche nicht (AERA et al., 2014), erscheint es angesichts der anderen, überzeugenden Hinweise auf valide Testwertinterpretation plausibel, im Fehlen einer Faktorenanalyse keinen grundsätzlichen Einwand gegen das Kriterienraster zu sehen.

Viertens besagt das *interpretation/use argument* (Kane, 2013), dass das entwickelte Kriterienraster von Auszubildenden der 1. und 2. Phase der Lehrkräftebildung genutzt werden soll, um schriftliche Unterrichtsentwürfe zu bewerten. Das setzt voraus, dass die entsprechende Zielgruppe (z. B. Fachseminarleiterinnen und Fachseminarleiter) das Kriterienraster als nützlich wahrnimmt. Nützlichkeit meint im Falle von Kriterienrastern, dass ein eindeutiger Mehrwert gegenüber Checklisten, Ratingskalen oder anderen Formen der Beurteilung erkannt werden muss (Brookhart & Chen, 2015), denn anderenfalls würde ein Kriterienraster nicht genutzt werden. Es wird zudem empfohlen, die Anwendung eines Kriterienrasters zu videografieren und die Zielgruppe bei der Anwendung des Kriterienrasters laut denken zu lassen, um möglicherweise auftretende Verständnisschwierigkeiten, Unzufriedenheit oder Überforderung zu erfassen und das Kriterienraster entsprechend überarbeiten zu können. In diesem Zusammenhang wäre auch zu

prüfen, ob beispielsweise verschiedene Fachseminarleiterinnen und Fachseminarleiter bei Anwendung des Kriterienrasters auf dieselben Unterrichtsentwürfe zu einer hohen Übereinstimmung gelangen oder ob die Niveaustufen möglicherweise doch nicht hinreichend trennscharf sind. Somit wäre die Gefahr abgeschwächt, objektive Validitätsevidenz mithilfe subjektiver Einschätzungen von Fachseminarleiterinnen und Fachseminarleitern zu sammeln. All diese Aspekte wären *validity evidence based on response processes*, die im Rahmen der vorliegenden Arbeit nicht erhoben worden sind.

Fünftens würde der empirisch bislang kaum untersuchte Zusammenhang zwischen der Unterrichtsplanung und der Unterrichtsdurchführung (Rothland, 2021) Aufschlüsse darüber geben, inwiefern mithilfe des Kriterienrasters die Qualität des durchgeführten Unterrichts vorausgesagt werden kann. Dies entspräche der prognostischen Validität als *validity evidence based on relations to other variables* (AERA et al., 2014). Es ist plausibel anzunehmen, dass ein Unterrichtsentwurf von hoher Planungsqualität eher in eine gut durchgeführte Stunde mündet und damit zu Lernzuwachs der Schülerinnen und Schüler führt, als ein Unterrichtsentwurf von geringer Planungsqualität. Wenn also die Scores des Kriterienrasters positiv mit der in welcher Form auch immer gemessenen Unterrichtsqualität korrelieren, dann ließe sich schlussfolgern, dass das Kriterienraster nicht irgendwelche, sondern besonders unterrichtsrelevante Planungsaspekte adressiert. Der Zusammenhang zwischen Planung und Durchführung ist wohl auch deshalb bislang wenig erforscht, weil die

*Datengewinnung in der Unterrichtsqualitätsforschung [...] sehr aufwändig [ist]. So ist es nicht erstaunlich, dass aussagekräftige Längsschnittstudien mit mehr als zwei Messzeitpunkten oder experimentelle Studien mit Randomisierung auf (der häufig primär interessierenden) Klassenebene in der Unterrichtsqualitätsforschung vergleichsweise selten realisiert werden und sich deshalb in der Literatur für viele der in den Modellen postulierten kausalen Annahmen nur schwache empirische Belege finden lassen [...].* (Begrich et al., 2023, S. 71)

Eine derartige Prüfung des Zusammenhangs würde die Beobachtung des durchgeführten Unterrichts erfordern, die für die Teilstichprobe der Referendarinnen und Referendare im Staatsexamen schon aus prüfungsrechtlichen Gründen nicht möglich gewesen wäre. Aufschlussreich und methodisch umsetzbar könnte die Anwendung des Kriterienrasters im Rahmen des Praxissemesters sein, in dem Dozierende aus der 1. Phase der Lehrkräftebildung ohnehin schriftliche Unterrichtsentwürfe erhalten und die Studierenden an den Schulen besuchen, den Unterricht beobachten und gemeinsam reflektieren. Dieser Schritt wurde im Rahmen der vorliegenden Arbeit jedoch nicht unternommen.

#### *Erkenntnisgewinn*

Insgesamt lässt sich feststellen, dass das entwickelte Kriterienraster geeignet ist, um die Qualität schriftlicher Unterrichtsentwürfe einschätzen zu können. Somit liegen nun empirisch geprüfte

Kriterien vor (Kreppf & König, 2022b), die in der Lehrkräftebildung und in der Forschung zur Unterrichtsplanung eingesetzt werden können. Die Befunde deuten darauf hin, dass das Kriterienraster relevante Kriterien umfasst (*evidence based on test content*), instruktional sensitiv ist (*evidence based on relations to other variables*) und Scores produziert, die Planungsqualität anzeigen (*evidence based on consequences of testing*).

Wie in Beitrag 1 gezeigt wurde, wird schriftlichen Unterrichtsentwürfen im derzeit noch heterogenen Feld der empirischen naturwissenschaftsdidaktischen Unterrichtsplanungsforschung insofern große Bedeutung beigemessen, als sie oft als Datenquelle dienen. Das entwickelte Kriterienraster bietet nun die Möglichkeit, als weitere Variable die Qualität der schriftlichen Unterrichtsentwürfe unterscheiden zu können. Beispielsweise könnte es in denjenigen Studien, die die Unterrichtsentwürfe als Datenmaterial nutzen, neue Fragestellungen ermöglichen, oder zumindest als belastbares Kriterium bei der Stichprobenziehung (*purposeful sampling*; Patton, 1990) dienen.

In Ergänzung zu den vor allem in Beitrag 7 diskutierten Befunden, aus denen das Validitätsargument entwickelt wurde, lassen sich einige Rückschlüsse auf die Entwicklung von Unterrichtsplanungskompetenz ziehen, die empirisch bislang kaum untersucht ist (König & Rothland, 2022). Zwar wird stets behauptet, dass die Entwicklung von Unterrichtsplanungskompetenz ein langfristiger, berufsbiografischer Prozess sei (Munthe & Conway, 2017), empirische Befunde liegen bislang allerdings noch nicht vor. Diese könnten jedoch nützlich sein, um zielgerichtet Interventionsmaßnahmen zu entwickeln, die Studierende (1. Phase der Lehrkräftebildung) sowie Referendarinnen und Referendare (2. Phase der Lehrkräftebildung) dabei unterstützen, kompetenzorientierten Biologieunterricht zu planen. Zur Stützung des Validitätsarguments wurde in Beitrag 7 ein Vergleich zwischen Master-Studierenden sowie Referendarinnen und Referendaren im Staatsexamen durchgeführt. Erwartungskonform erreichten die Referendarinnen und Referendare im Durchschnitt einen signifikant höheren Score als die Studierenden, was als *validity evidence based on relations to other variables* (AERA et al., 2014) die Annahme stützt, dass das Kriterienraster valide Interpretationen zulässt. Auch wenn hier keine Längsschnittstudie durchgeführt wurde, stützen die Befunde die Annahme, dass Unterrichtsplanungskompetenz am Ende der Lehrkräftebildung stärker ausgeprägt ist als nach Ende des Praxissemester im Master-Studium und damit kurz nach der ersten längerfristigen Praxiserfahrung. Eine mögliche, an diese Arbeit anknüpfende Forschungsarbeit könnte in der Entwicklung eines Kompetenzstrukturmodells (Mayer & Wellnitz, 2014; Schecker & Parchmann, 2006) bzw. einer *learning progression* (Upmeier zu Belzen et al., 2019) bestehen. Dies wird im Ausblick genauer ausgeführt.

*„Pläne machen und Vorsätze fassen bringt viel gute Empfindungen mit sich, und wer die Kraft hätte, sein ganzes Leben lang nichts als ein Pläne-Schmiedender zu sein, wäre ein sehr glücklicher Mensch: Aber er wird sich gelegentlich von dieser Tätigkeit ansruben müssen, dadurch dass er einen Plan ausführt – und da kommt der Ärger und die Ernüchterung.“*

(Friedrich Nietzsche)

## 5 Vorschläge für die Unterrichtspraxis

Lehrkräfte haben die Aufgabe, die Kompetenzen von Schülerinnen und Schülern zu fördern. Dies tun sie, indem sie Lehr-Lern-Szenarien planen (und durchführen), die entsprechende in den Bildungsstandards (KMK, 2005; KMK, 2020) bzw. den darauf aufbauenden Rahmenlehrplänen (für Berlin: SenBJF, 2015; SenBJF & MBSJ, 2021) formulierte Kompetenzen adressieren. Wie in Beitrag 1 gezeigt wurde, spielt die Nutzung und Anpassung von *curriculum materials* im Planungsprozess eine große Rolle (z. B. Beyer & Davis, 2012; Forbes, 2011). Lehrkräfte nutzen also bereits vorhandene und beispielsweise von Schulbuchverlagen reichhaltig angebotene Unterrichtsmaterialien, die sie entweder direkt einsetzen oder im Sinne didaktischer Adaptivität (König et al., 2015) so anpassen, dass die Aufgaben und Materialien dem Kompetenzstand der eigenen Lerngruppe angemessen sind (vgl. Abb. 1, oben rechts).

Bezüglich der Nutzung von Materialien lässt sich in Ergänzung zu den in den Beiträgen 1-9 berichteten Ergebnissen feststellen, dass in den Staatsexamensentwürfen vergleichsweise wenig auf Unterrichtsmaterialien Bezug genommen wird. Zwei Befunde sind auffällig: Erstens nutzen Referendarinnen und Referendare in der analysierten Stichprobe von  $N = 107$  Staatsexamensentwürfen nur wenige Quellen ( $Mdn = 7$ ) zur Stützung ihrer Argumentation. Während fachwissenschaftliche Quellen (CK) am häufigsten zitiert werden, werden fachdidaktische (PCK) im Median kein einziges Mal und pädagogische Quellen (PK) einmal zitiert. Auf Unterrichtsmaterial (z. B. Schulbücher oder Unterrichtsvorschläge aus Zeitschriften wie *Unterricht Biologie*) wird im Median zwei Mal Bezug genommen.

Zweitens zeigt eine Auflistung der Quellen, die hinsichtlich der im Rahmen dieser Arbeit besonders relevanten Kategorie PCK (Tab. 5) genutzt wurden, dass in vergleichsweise wenigen Staatsexamensentwürfen auf einschlägige biologiedidaktische Literatur Bezug genommen wird. Auffällig ist, dass zum Teil mehr als zehn Jahre alte Quellen genutzt werden, und dass es sich bei den genutzten Quellen mit Ausnahme des Beitrags von Terzer und Upmeier zu Belzen (2008) nicht um Publikationen in fachdidaktischen Zeitschriften handelt, sondern überwiegend um Beiträge in Sammelbänden zur Einführung in die Didaktik bzw. Methodik. Zudem wird zwar Literatur zu Schülervorstellungen (Hammann & Asshoff, 2014; Kattmann, 2015; Kattmann, 2017) genutzt, allerdings sehr selten. Dies kann als ein weiterer Hinweis auf die geringe Relevanz von

Schülervorstellungen im Rahmen der Staatsexamensprüfung gedeutet werden (vgl. Beitrag 4, Beitrag 7). Möglicherweise greift hier die in Beitrag 1 skizzierte Unterscheidung zwischen *conceptual knowledge* und *procedural knowledge*: Wissen darüber zu haben, dass es zu vielen Themen empirisch bereits beschriebene Schülervorstellungen gibt (*conceptual knowledge*), wäre eine notwendige Grundlage, aber noch keine hinreichende. Es bedarf darüber hinaus Strategien, dieses Wissen in der Unterrichtsplanung anzuwenden, Diagnoseaufgaben zur Erhebung von Schülervorstellungen einzusetzen, die Ergebnisse auszuwerten und darauf ausgerichteten Unterricht zu planen (*procedural knowledge*) – eine zweifelsohne herausfordernde Aufgabe. Erst wenn dieses Wissen als prozedural genutzt werden kann, schlägt es sich in der Planung nieder, so dass möglicherweise erst dann auch entsprechende Quellen zitiert werden.

### Tabelle 5

In der Gesamtstichprobe (N = 107) mindestens zwei Mal zitierte PCK-Quellen

PCK-Quellen	n
Gropengießer, H., Harms, U. & Kattmann, U. (2013). <i>Fachdidaktik Biologie</i> . Hallbergmoos: Aulis.	15
Spörhase, U. & Ruppert, W. (2016): <i>Biologie Methodik. Handbuch für die Sekundarstufe I und II</i> . 3. Auflage, Cornelsen Verlag, Berlin	9
Spörhase-Eichmann, U. & Ruppert W. (2004). <i>Biologie Didaktik. Praxisbandbuch für die Sekundarstufe I und II</i> . Berlin: Cornelsen.	7
Gropengießer, H., Kattmann, U. & Krüger, D. (2010): <i>Biologiedidaktik in Übersichten</i> , Hallbergmoos: Aulis Verlag.	6
Köhler, K. (2004): Nach welchen Prinzipien kann Biologieunterricht gestaltet werden? In U. Spörhase-Eichmann & W. Ruppert (Hrsg.), <i>Biologie Didaktik. Praxisbandbuch für die Sekundarstufe I und II</i> (S. 124-145). Berlin: Cornelsen.	4
Kattmann, U. (2015). <i>Schüler besser verstehen</i> , Hallbergmoos: Aulis Verlag.	3
Hammann, M., & Asshoff, R. (2014). <i>Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten</i> . Seelze: Klett Kallmeyer.	2
Berck, K.-H. & Graf, D. (2018). <i>Biologiedidaktik – Grundlagen und Methoden</i> , 5. Auflage, Wiebelsheim: Quelle & Meyer.	2
Terzer, E., & Upmeyer zu Belzen, A. (2008). Naturwissenschaftliche Erkenntnisgewinnung durch Modelle – Modellverständnis als Grundlage für Modellkompetenz. <i>Zeitschrift für Didaktik Der Biologie (ZDB) - Biologie Lehren Und Lernen</i> , 16(1), 33–56. <a href="https://doi.org/10.4119/zdb-1653">https://doi.org/10.4119/zdb-1653</a>	2
Killermann, W., Hiering, P. & Starosta, B. (2009). <i>Biologieunterricht heute. Eine moderne Fachdidaktik</i> . Donauwörth: Auer.	2
Weitzel, H. & Schaal, S. (Hrsg.) (2016). <i>Biologie unterrichten: planen, durchführen, reflektieren</i> . Berlin: Cornelsen.	2
Kattmann, U. (2017). <i>Biologie unterrichten mit Alltagsvorstellungen: Didaktische Rekonstruktion in Unterrichtseinheiten</i> . Seelze: Klett Kallmeyer.	2

Auch wenn aufgrund der oben genannten Limitationen der vorliegenden Arbeit, insbesondere der fehlenden Möglichkeit zur kommunikativen Validierung, keine Erkenntnisse darüber abzuleiten sind, aus welchen Gründen bestimmte Quellen (nicht) genutzt wurden, zeigt sich eine insgesamt geringe Nutzung fachdidaktischer Literatur und vor allem in Fachzeitschriften veröffentlichter empirischer oder theoretischer Studien. Dieser Befund ist zumindest für das Land Berlin insofern aussagekräftig, als es sich bei den durch die Senatsverwaltung für Bildung, Jugend und Familie des

Landes Berlin zur Verfügung gestellten Staatsexamensentwürfen um die Gesamtstichprobe der beiden Prüfungsdurchgänge 2018/19 und 2019 handelt.

Möglicherweise benötigen angehende Lehrkräfte statt zumeist noch stärker der Forschung verhafteter fachdidaktischer Literatur mehr oder weniger unmittelbar einsetzbare Unterrichtsmaterialien, mit denen sie die Kompetenzen ihrer Schülerinnen und Schüler fördern können. Im Sinne der Theorie geplanten Verhaltens (Ajzen, 1991, vgl. Graf, 2007) ließe sich jedoch auch annehmen, dass (angehenden) Lehrkräften die Relevanz und der Nutzen eines fachdidaktisch gut begründeten Planungsprozesses oft nicht klar ist und sie daher vorrangig auf das vorhandene Material zurückgreifen, ohne empirische oder theoretische Literatur zu rezipieren. Technisch gesprochen ist ihre Verhaltensintention möglicherweise darauf ausgerichtet, fachdidaktisch belastbaren Unterricht zu planen. Die subjektive soziale Norm gegenüber diesem Verhalten (hier: die Ausbildungspraxis im Referendariat, die Erwartungshaltung im Fachseminar) sowie die subjektiv wahrgenommene Verhaltenskontrolle (hier: die Selbsteinschätzung, fachdidaktische Literatur recherchieren, rezipieren und mit Mehrwert nutzen zu können) könnten dem jedoch entgegenstehen, so dass im tatsächlich gezeigten Verhalten (hier: dem geschriebenen Unterrichtsentwurf) kaum auf fachdidaktische Quellen rekurriert wird.

Aus den Befunden der Beiträge 4 und 5 wurde abgeleitet, dass sowohl der Umgang mit Schülervorstellungen als auch die Förderung von Kompetenzen im Bereich Erkenntnisgewinnung eine große Herausforderung für angehende Lehrkräfte darstellt. Zur Erhebung von Schülervorstellungen sowie zum produktiven Umgang mit ihnen gibt es bereits eine Reihe von deutschsprachigen Publikationen (Hammann & Asshoff, 2014; Kattmann, 2015; Kattmann, 2017), die allesamt auch in dieser Stichprobe genutzt wurden (Tab. 5). Die Schwierigkeit besteht hierbei eher in der zielgenauen Erhebung der Alltagsvorstellungen von Schülerinnen und Schülern (z. B. zu Beginn der Unterrichtssequenz) sowie der daraufhin angepassten didaktischen Strukturierung des Lehr-Lern-Prozesses in den jeweiligen Einzelstunden.

Ein Bedarf an Unterrichtsmaterialien scheint dagegen eher hinsichtlich des Kompetenzbereichs Erkenntnisgewinnung zu bestehen. Es wurde exemplarisch an der Planung hypothesengeleiteten Unterrichts gezeigt, dass die inhaltliche Strukturierung einer Unterrichtsstunde eine herausfordernde Aufgabe darstellt (Beitrag 5). Zwar gibt es auch hier inzwischen an Lehrkräfte gerichtete Unterrichtsideen (*allgemein*: z. B. Lübeck, 2020; *Beobachten*: z. B. Zabel, 2007; *Ordnen und Vergleichen*: z. B. Wellnitz, 2013; *Experimentieren*: z. B. Peter, 2007; *Modellieren*: z. B. Fleige et al., 2012). Dennoch scheint es nach wie vor herausfordernd zu sein, den Kompetenzbereich Erkenntnisgewinnung vom Kompetenzbereich Fachwissen bzw. Sachkompetenz abzugrenzen und unter Erkenntnisgewinnung nicht das Erarbeiten neuer biologischer Fachinhalte durch den



Einsatz von Arbeitstechniken wie das Mikroskopieren oder Arbeitsweisen wie das Experimentieren, sondern ein Wissen über die Methoden und die Charakteristika der Naturwissenschaften zu verstehen (z. B. Capps & Crawford, 2013; Gyllenpalm & Wickman, 2011; Krell & Krüger, 2022). Aus diesem Grunde wurden auf der Basis der empirischen Befunde dieser Arbeit mehrere Unterrichtsvorschläge entwickelt, mit denen explizit Erkenntnisgewinnungskompetenzen gefördert werden können. Die Beiträge adressieren Aspekte, die in unterrichtspraktischen Publikationen bislang eher selten berücksichtigt werden, wie das Beobachten (Beitrag 12), das Modellieren mit Simulationen (Beitrag 13), den Unterschied zwischen Regeln und Gesetzen im alltagssprachlichen und im naturwissenschaftlichen Sinne als kognitiv-epistemischen *Nature-of-Science*-Aspekt (Beitrag 10), die Bedeutung von Kooperation und Konkurrenz als sozial-institutionellen *Nature-of-Science*-Aspekt (Beitrag 11) sowie übergeordnet die drei logischen Schlussweisen Induzieren, Deduzieren und Abduzieren (Beitrag 14).

## 5.1. Mysterys

Mysterys stellen eine geeignete Methode dar, um Kompetenzen von Schülerinnen und Schülern in offenen Lehr-Lern-Arrangements zu fördern und dabei binnendifferenziert die Bedürfnisse der Schülerinnen und Schüler zu adressieren (Grospietsch & Lins, 2023). Ausgehend von einer mysteriösen oder paradox erscheinenden Aussage sollen Schülerinnen und Schüler auf einzelnen Kärtchen dargebotene Informationen sinnvoll miteinander verknüpfen und das Problem lösen. Die ursprünglich aus der Geografiedidaktik stammende Methode (Leat, 1998) wurde inzwischen auch in der Biologiedidaktik aufgegriffen, so dass eine Reihe von Unterrichtsvorschlägen für den Biologieunterricht vorliegt (z. B. Mülhausen & Pütz, 2013; Ziepprecht & Meier, 2019). Während jedoch Mülhausen und Pütz (2013) beispielsweise das didaktische Potential der Mystery-Methode eher mit allgemeinen Begriffen wie „*multiperspektivisches und systemisches Denken*“ (S. 9), „*Handlungsfähigkeit*“ (S. 10) oder „*problemlösendes Denken*“ (S. 10) kennzeichnen, zielen die im Folgenden dargestellten Beiträge 10, 11 und 12 unmittelbar und spezifisch auf die Förderung von Erkenntnisgewinnungskompetenzen ab. Gewiss sind die von Mülhausen und Pütz (2013) benannten Aspekte ebenfalls wichtig bei der Bearbeitung von Mysterys. Aber angesichts der Notwendigkeit, ausgehend von curricularen Standards Kompetenzen zu fördern und darauf aufbauend eine Lehr-Lern-Struktur zu entwickeln, die Schülerinnen und Schülern diese Kompetenzentwicklung ermöglicht (Abb. 3; Blömeke et al., 2015), wurde im Rahmen dieses Dissertationsprojekts in Zusammenarbeit mit dem Cornelsen-Verlag ein Heft herausgegeben, in dem insgesamt 22 Beiträge versammelt sind, die verschiedene Aspekte von

Erkenntnisgewinnungskompetenz (Krell & Krüger, 2022) adressieren. Im Sinne der in Beitrag 5 gewonnenen Erkenntnis, dass es angehenden Lehrkräften schwerfällt, Stunden zur Förderung von Erkenntnisgewinnungskompetenzen inhaltlich zu strukturieren und dabei nicht versehentlich Sachkompetenz zu fördern, wird in diesem Heft die Mystery-Methode als Instruktionsstrategie genutzt, mit deren Hilfe Erkenntnisgewinnungskompetenzen gefördert werden können.

Die folgenden drei Beiträge entstammen dieser Publikation und setzen sich exemplarisch mit drei Aspekten von Erkenntnisgewinnung auseinander: erstens mit dem Unterschied zwischen Regeln und Gesetzen als einem kognitiv-epistemischen Aspekt von Nature of Science (Erduran & Dagher, 2014; Reinisch & Fricke, 2022; Beitrag 10), zweitens mit dem Verhältnis von Kooperation und Konkurrenz in den Naturwissenschaften als einem sozial-institutionellen Aspekt von Nature of Science (Erduran & Dagher, 2014; Reinisch & Fricke, 2022; Beitrag 11) und drittens mit dem Beobachten als naturwissenschaftliche Arbeitsweise (z. B. Wellnitz & Mayer, 2013; Beitrag 12).

### 5.1.1. Beitrag 10: Bergmann'sche Regel

Beitrag 10 wurde publiziert: Großmann, L. (2022). Coole Kriminelle? Von ungestraften Regelbrüchen in der Natur. In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 102-108). Berlin: Cornelsen Verlag.

#### Zusammenfassung

In diesem Beitrag wird der folgende Bildungsstandard für die Sekundarstufe II adressiert: Die Schülerinnen und Schüler „reflektieren Möglichkeiten und Grenzen des konkreten Erkenntnisgewinnungsprozesses sowie der gewonnenen Erkenntnisse (z. B. Reproduzierbarkeit, Falsifizierbarkeit, [...] Vorläufigkeit)“ (E15; KMK, 2020). Konkret unterscheiden sie am Beispiel der Bergmann'schen Regel zwischen dem alltagssprachlichen und dem naturwissenschaftlichen Begriffsverständnis des Begriffs *Regel*. Damit wird auf Post (2018) rekurriert, der die Verwendung des Begriffs „Bergmann'sche Regel“ in Schulbüchern und Abituraufgaben erstmalig in der Biologiedidaktik kritisch reflektiert hat. Anders als der Begriff *Regel* suggeriert, handelt es sich dabei in den Naturwissenschaften um keine feststehenden, unumstößlichen Kausalzusammenhänge: Regeln beschreiben Korrelationen.

Mithilfe dieses Mysterys lassen sich also zwei wichtige Merkmale des naturwissenschaftlichen Erkenntnisprozesses reflektieren: Erstens stellen Regeln Korrelationshypthesen dar und sind somit keine spontan geäußerten Vermutungen, so wie es die alltagssprachliche Verwendung des Begriffs suggeriert, sondern empirisch fundierte und begründete Voraussagen eines Zusammenhangs zwischen zwei Variablen, der mithilfe der Arbeitsweise *Beobachten* untersucht

werden kann (Krell & Krüger, 2022). Zweitens wird am Beispiel von Carl Bergmann und ausgewählten Zitaten aus seiner Originalpublikation dazu angeregt, eine solche Hypothese als Teil eines naturwissenschaftlichen Forschungsprozesses zu betrachten, in dem es nicht unbedingt um Bestätigung, sondern im Sinne Popper (1934/2005) um Widerlegung der Hypothese geht, da nur dies Erkenntniszuwachs ermöglicht. Dass die Bergmann'sche Regel inzwischen für zahlreiche Arten bestätigt wurde, ist bereits Gegenstand vieler Schulbücher. Dass sie inzwischen jedoch auch für viele Arten widerlegt wurde (Gohli & Voje, 2016), ist für Schülerinnen und Schüler aber ebenso relevant. Gerade die Widerlegung der Regel lässt sich als didaktischer Anknüpfungspunkt nutzen, um im Sinne des Kompetenzbereichs Erkenntnisgewinnung über die Methoden der Naturwissenschaften zu reflektieren.

### 5.1.2. Beitrag 11: Darwin und Wallace

Beitrag 11 wurde publiziert: Großmann, L. (2022). Charles Darwin und Alfred Russel Wallace. In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 116-122). Berlin: Cornelsen Verlag.

#### Zusammenfassung

In diesem Beitrag wird der folgende Bildungsstandard für die Sekundarstufe II adressiert: Die Schülerinnen und Schüler „reflektieren Bedingungen und Eigenschaften biologischer Erkenntnisgewinnung“ (E17; KMK, 2020). Konkret untersuchen sie das Verhältnis zwischen Charles Darwin und Alfred Russel Wallace. Letzterer wird zwar in manchen Schulbüchern am Rande erwähnt, seine Bedeutung für die Evolutionsforschung und sein Verhältnis zu Darwin werden dabei jedoch nicht eingehend berücksichtigt. Ausgehend von einer aktuellen Wallace-Biografie (Glaubrecht, 2014) wird in diesem Mystery das Verhältnis von Kooperation und Konkurrenz in den Naturwissenschaften untersucht.

Ohne jeden Zweifel sind die Verdienste Charles Darwins nicht hoch genug einzuschätzen, der über viele Jahre sowohl in England durch Züchtungen als auch auf seiner Weltreise Daten gesammelt und ausgewertet hat, ehe er seine Evolutionstheorie niederschrieb und sie im Jahre 1859 unter dem Titel „*On the Origin of Species*“ (Darwin, 1859) publizierte. Als im Jahre 1855 Alfred Russel Wallace, der zu dieser Zeit im Malayischen Archipel unterwegs war, seinen sogenannten Sarawak-Aufsatz (Wallace, 1855) publizierte und Darwin gar um kritische Rückmeldung zu seinen Gedanken bat, in dem er zentrale biogeografische Zusammenhänge der Evolutionstheorie Darwins vorwegnahm, wurde Darwin klar, dass er seine Jahrzehnte lange Arbeit rasch publizieren muss, wenn er die Anerkennung für die Evolutionstheorie nicht an Wallace verlieren will.

Mithilfe dieses Mysterys lässt sich anhand der originalen Briefe das Verhältnis von Kooperation, die vor allem Wallace gesucht hat, und Konkurrenz, die vor allem Darwins Freunde wahrgenommen haben, reflektieren. Anders als in Beitrag 10 geht es hierbei also weniger um einen kognitiv-epistemischen Aspekt von Nature of Science, sondern um einen sozial-institutionellen (Erduran & Dagher, 2014): Dass Wissenschaft ohne Kooperation von Forschenden kaum denkbar ist, ist sicher ein wertvoller Aspekt, der im Biologieunterricht bislang noch zu kurz kommt. Dass dabei jedoch natürlich auch Konkurrenz eine wichtige und möglicherweise ja auch produktive Rolle spielt (Fang & Casadevall, 2015), sollte ebenso nicht verschwiegen werden. Am Beispiel der Entwicklung der Evolutionstheorie von Darwin und Wallace lässt sich das Wechselspiel von Kooperation und Konkurrenz exemplarisch nachvollziehen und reflektieren.

### 5.1.3. Beitrag 12: Beobachten

Beitrag 12 wurde publiziert: Großmann, L. & Krüger, D. (2022). Guckst du noch oder beobachtest du schon? In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 39-45). Berlin: Cornelsen Verlag.

#### Zusammenfassung

In diesem Beitrag wird der folgende Bildungsstandard für die Sekundarstufe I adressiert: Die Schülerinnen und Schüler „erörtern Tragweite und Grenzen von Untersuchungsanlage, -schritten und -ergebnissen“ (E8; KMK, 2005). Konkret setzen sie sich kritisch mit der naturwissenschaftlichen Arbeitsweise *Beobachten* auseinander und unterscheiden es vom gewöhnlichen Hingucken. Am Beispiel der Beobachtung von Wölfen in einem Zoogehege wird das Vorgehen zweier Schüler und einer Schülerin kontrastiert: Während die eine streng hypothesen- und kriteriengeleitet vorgeht, ihre Beobachtungen sorgfältig dokumentiert und diese strikt von der Interpretation der Beobachtungen trennt, geht einer der anderen Schüler unsystematisch vor und guckt eher, als dass er beobachtet. Ein weiterer Schüler stellt eine Zwischenform dar. Auf diese Weise wird dazu angeregt, das Vorgehen beim wissenschaftlichen Beobachten zu reflektieren und dabei über die Bedeutung einer initialen Hypothese sowie von Kriterien beim Beobachten nachzudenken.

Wie in den beiden vorhergehenden Mysterys wird hierbei deutlich, dass es nicht um die Wölfe und somit nicht um den Erwerb biologischen Fachwissens, sondern explizit um ein wissenschaftsmethodisch adäquates Verständnis der Arbeitsweise *Beobachten* und somit um einen wichtigen Teil des Kompetenzbereichs Erkenntnisgewinnung geht.

## 5.2. Beitrag 13: Simulation von Fischschwärmen

Beitrag 13 wurde publiziert: Großmann, L., Vogt, N. & Tietjen, B. (2023). Bietet Schwarmverhalten einen Überlebensvorteil für Fische? Testen von Hypothesen mithilfe eines Modells. *Unterricht Biologie*, 47(1), S. 5-9.

### Zusammenfassung

In diesem Beitrag wird der folgende Bildungsstandard für die Sekundarstufe II adressiert: Die Schülerinnen und Schüler „planen und führen hypothesengeleitete Modellierungen durch und protokollieren sie“ (E4; KMK, 2020). Konkret untersuchen sie mithilfe einer NetLogo®-Simulation, inwiefern Schwarmverhalten Fischen einen Überlebensvorteil bietet. Zu diesem Zweck wurde eine Simulation programmiert, in der verschiedene Variablen (Anzahl der Fische, Futtermenge, Anwesenheit eines Raubfisches, Möglichkeit zur Schwarmbildung) variiert werden können. Nach Start der Simulation kann man die Bewegung der Fische in einem Fenster beobachten. Zudem werden zeitgleich Plots generiert, in denen die Anzahl der überlebenden, der verhungerten und der gefressenen Fische als Kurven aufgetragen sind. Durch gezielte Variation der Variablen werden Daten erhoben und zur Überprüfung von vorab aufgestellten Hypothesen genutzt.

Im Sinne der in Beitrag 5 gewonnenen Erkenntnis, dass es angehenden Lehrkräften schwerfällt, zwischen den Kompetenzbereichen Fachwissen bzw. Sachkompetenz auf der einen Seite und Erkenntnisgewinnung auf der anderen Seite zu unterscheiden, wird in Beitrag 13 eine Thematisierung von Fischschwärmen aus evolutionsbiologischen, ökologischen oder verhaltensbiologischen vermieden. Stattdessen werden die Schülerinnen und Schüler mithilfe gestufter Aufgabenstellungen dazu angeregt, über den Zweck und das Testen von Modellen und somit zwei Teilkompetenzen von Modellierkompetenz (Krüger & Upmeyer zu Belzen, 2021) nachzudenken. Auf diese Weise soll deutlich werden, dass es im Sinne des Kompetenzbereichs Erkenntnisgewinnung nicht darum geht, biologische Erkenntnisse über Fischschwärme, sondern wissenschaftsmethodische Erkenntnisse über die Arbeitsweise Modellieren zu gewinnen.

## 5.3. Beitrag 14: Historisch-evolutionäres Schlussfolgern

Beitrag 14 wurde noch nicht publiziert, sondern befindet sich derzeit im Druck: Vogt, N., Großmann, L. & Krüger, D. (im Druck). Logisches Schließen im Biologieunterricht – Wenn Aufgaben evolvieren. *MNU-Journal*.

### **Zusammenfassung**

In diesem Beitrag wird der folgende Bildungsstandard für die Sekundarstufe II adressiert: Die Schülerinnen und Schüler „*reflektieren Bedingungen und Eigenschaften biologischer Erkenntnisgewinnung*“ (E17; KMK, 2020).

Erneut ausgehend von Beitrag 5 und dem Unterschied zwischen Fachwissen bzw. Sachkompetenz auf der einen Seite und Erkenntnisgewinnung auf der anderen Seite wird auch in Beitrag 14 ein Aspekt des Kompetenzbereichs Erkenntnisgewinnung adressiert, der vor allem im Themenfeld Evolution bisher implizit eine Rolle spielte und der hier explizit gemacht wird. Es geht um die logischen Schlussweisen *Deduktion*, *Induktion* und *Abduktion*. Anders als in den anderen Teilbereichen der Biologie können Evolutionsbiologinnen und Evolutionsbiologen keine Hypothesen im Sinne von Voraussagen postulieren und diese dann in der Zukunft testen, wenn sie ausgestorbene Arten und Ökosysteme aus der Vergangenheit untersuchen. Stattdessen versuchen sie, Phänomene aus der Vergangenheit zu erklären, indem sie die Ursachen für diese Phänomene identifizieren. Ausgehend von einer an die Lehrkräfte gerichteten theoretischen Hinführung wird aus einer von der Erstautorin entwickelten Leistungsaufgabe eine Übungsaufgabe, in der Schülerinnen und Schüler identifizieren sollen, ob in den aufgeführten Beispielen deduktiv, induktiv oder abduktiv geschlossen wird.

Dieser Beitrag ist insofern innovativ, als die explizite Reflexion über das rückwärtsgerichtete Abduzieren im Gegensatz zum vorwärtsgerichteten Deduzieren (Krüger & Upmeyer zu Belzen, 2021) bislang erst in wenigen Unterrichtsmaterialien aufgegriffen wird (z. B. Engelschalt et al., 2022). Beitrag 14 leistet einen weiteren Beitrag dazu, dass Schülerinnen und Schüler über diesen zwar sehr abstrakten, aber doch sehr grundlegenden Zugang zu den naturwissenschaftlichen Denkweisen nachdenken.

### **5.4. Zusammenfassende Diskussion**

In diesem Abschnitt werden die Beiträge 10, 11, 12, 13 und 14 gemeinsam diskutiert und ihr innovatives Potential für die schulische Unterrichtspraxis kritisch reflektiert. Allen fünf Beiträgen liegt die Idee zugrunde, am Beispiel des Kompetenzbereichs Erkenntnisgewinnung die Komplexität der Unterrichtsplanung (Abb. 1) für (angehende) Lehrkräfte dergestalt zu reduzieren, dass zentrale Probleme 1. und 2. Art (Abb. 3) bereits gelöst sind und Gestaltungsmerkmale wie die Aufgaben, das Arbeitsmaterial und die ausgewählten Themen in Passung zu den curricularen

Standards stehen. Nur darüber hinausgehende Anpassungen an den Kompetenzstand der jeweiligen Schülerinnen und Schüler müssten Lehrkräfte noch selbst leisten.

Statt einer Diskussion im Lichte der in den jeweiligen Beiträgen didaktisierten Erkenntnisgewinnungsaspekte, die eine Einbeziehung der jeweiligen fachdidaktischen Literatur (z. B. zu Modellierkompetenz) erforderlich machen würde, wird im Sinne des Themas der vorliegenden Dissertation – der Unterrichtsplanung – im Folgenden herausgearbeitet, inwiefern die fünf Beiträge (angehende) Lehrkräfte dabei unterstützen könnten, kompetenzfördernden Biologieunterricht zu planen.

Zwei der sechs kognitiven Herausforderungen, die in der Unterrichtsplanung bewältigt werden müssen und für die in den fünf dargestellten Beiträgen Hilfe angeboten wird, sind die Festlegung eines klaren Lernziels (*clarity of learning objectives*), das den Schülerinnen und Schülern transparent gemacht wird, sowie die daraufhin ausgerichtete Entwicklung von kompetenzfördernden Aufgaben (*task creation*; König et al., 2021a). Hinsichtlich kompetenzfördernder Aufgaben sind Gestaltungsmerkmale zu beachten, die einen Einfluss auf das kognitive Aktivierungspotential der Aufgaben und somit auf ihre im besten Falle lernförderliche Wirkung haben (Maier et al., 2014). Zu diesen Merkmalen gehören unter anderen die adressierte Wissensart (*Fakten-Prozeduren-Konzepte-Metakognition*), der ausgelöste kognitive Prozess (*Reproduktion-Transfer-Problemlösen*), der Lebensweltbezug (*kein-konstruiert-authentisch-real*), die Offenheit (*definiert/ konvergent-definiert/ divergent-ungenau/ divergent*), die sprachlogische Komplexität (*niedrig-mittel-hoch*) und die Repräsentationsformen (*eine-Integration-Transformation*). Je nach Ziel sind die Aufgaben so zu gestalten, dass sie einerseits hinreichend kognitiv aktivieren und andererseits nicht überfordernd sind. An zwei Beispielen aus den vorgestellten Beiträgen lässt sich dies skizzieren:

Beitrag 14 adressiert ein sehr abstraktes Konzept, die logischen Schlussweisen (*Wissensart*), das schwer zu verstehen ist, da es nur einen mittelbaren Bezug zum eigenen Erleben der Schülerinnen und Schüler gibt (*Lebensweltbezug*). Mit der Abstraktheit hängt zusammen, dass eher ungewöhnliche, fremdsprachliche Begriffe wie „*Abduktion*“ genutzt werden, die auf Beispielaufgaben angewendet werden müssen (*sprachlogische Komplexität*). Vor diesem Hintergrund wurde Wert daraufgelegt, das Material so zu reduzieren, dass die Schülerinnen und Schüler die Informationen zu den drei Schlussweisen reproduzieren und anwenden können (*kognitiver Prozess*) und zu diesem Zweck neben Texten nicht auch noch mit Grafiken oder Bildern arbeiten müssen (*Repräsentationsformen*). Zudem wird die von den Schülerinnen und Schülern erwartete Performanz klar definiert und umfasst eindeutig richtige Lösungen, die keinen Raum für eine multiperspektivische Diskussion bieten (*Offenheit*).

Die drei Mysterys (Beiträge 10, 11 und 12) hingegen sind gänzlich anderer Natur: Zwar geht es auch hier jeweils darum, ein Konzept zu erfassen (*Wissensart*), doch adressiert die Mystery-Methode ihrem Wesen nach das Problemlösen (*kognitiver Prozess*). Aus 24 Kärtchen, die neben Texten auch Diagramme, Symbole, Fotos oder Tabellen enthalten (*Repräsentationsformen*), müssen die Schülerinnen und Schüler eine mysteriöse Frage beantworten oder eine widersprüchliche Aussage auflösen. Diese Fragen bzw. Aussagen haben sind im Falle von Beitrag 10 (Bergmann'sche Regel) authentisch, im Falle von Beitrag 11 (Darwin und Wallace) real, und im Falle von Beitrag 12 (Beobachten von Wölfen) konstruiert (*Lebensweltbezug*). Diese Gestaltungsentscheidungen hängen jeweils mit dem Thema zusammen (König et al., 2021a), denn während das Mystery zur Darwin und Wallace in reduzierter Form einen historischen Fall nacherzählt (daher *real*) und im Falle der Bergmann'schen Regel Carl Bergmanns tatsächliche Ausführungen mit didaktisierten Beispielen gestützt werden (daher immer noch *authentisch*), handelt es sich bei den Wölfen bloß um einen einigermaßen plausiblen Kontext, an dem die Arbeitsweise Beobachten analysiert werden kann (daher *konstruiert*). Davon unabhängig wurde aufgrund der Menge an Einzelinformationen, die hier gesichtet und geordnet werden müssen, darauf geachtet, dass die Kärtchen in möglichst einfacher Sprache verfasst und somit leicht verständlich sind (*sprachlogische Komplexität*). Dies ist insbesondere insofern relevant, als das Lernprodukt in Form einer Concept-Map bei den Mysterys zwar klar definiert ist, aber dennoch ganz verschiedene – also divergente – Lösungen möglich sind (*Offenheit*).

Zusammenfassend wird hieran deutlich, dass in den Beiträgen 10-14 Gestaltungsmerkmale berücksichtigt worden sind, die jeweils abhängig vom Ziel der jeweiligen Aufgabe so umgesetzt wurden, dass die Aufgaben herausfordernd, aber nicht überfordernd, Erkenntnisgewinnungskompetenzen fördern. Abschließend ist zu bedenken, dass die fünf Beiträge 10-14 an dieser Stelle nur im Lichte der Unterrichtsplanung betrachtet wurden. Damit sie ihr lernwirksames Potential entfalten können, müssen sie von (angehenden) Lehrkräften jedoch zunächst gefunden, dann eventuell an die jeweiligen Lerngruppen und Zielsetzungen angepasst (z. B. Beyer & Davis, 2012; Kang et al., 2016; König et al., 2021a) und anschließend in der Unterrichtsdurchführung eingesetzt werden. Dies ist nicht trivial, denn der Lernerfolg der Schülerinnen und Schüler hängt maßgeblich davon ab, wie der Prozess der Umsetzung derartiger Aufgaben erfolgt (Stein et al., 1996; Stein & Lane, 1996): Zunächst werden die Schülerinnen und Schüler mit den Aufgaben am Ende des Einstiegs oder zu Beginn der Erarbeitungsphase konfrontiert (*tasks as set up*; Stein et al., 1996, S. 459), ehe die Aufgaben von den Schülerinnen und Schülern mithilfe der zur Verfügung gestellten Materialien bearbeitet werden (*tasks as implemented*; Stein et al., 1996, S. 459) und der Lernprozess sowie die anschließende Sicherungsphase von der Lehrkraft begleitet wird. Am Beispiel des Fachs Mathematik wurde gezeigt, dass das kognitive Aktivierungspotential von Aufgaben durch die Art



und Weise, wie sie in den Unterricht eingeführt werden, oft sinkt (Stein et al., 1996; Stein & Lane, 1996). Auch Jackson et al. (2013) stellten fest, dass in einer Mehrheit der untersuchten Mathematikstunden das kognitive Aktivierungspotential der Aufgaben bereits durch die Art und Weise, wie die Aufgaben in die Unterrichtsstunde eingeführt werden (*tasks as set up*) sinkt und dass zudem die Art und Weise, wie das Unterrichtsgespräch in der Sicherungsphase abläuft, einen großen Einfluss auf die Lernwirksamkeit von Aufgaben haben kann. Kang et al. (2016) bestätigen beide Befunde und stellen fest, dass sowohl zwischen der Planung und der Einführung der Aufgaben in der Unterrichtsstunde sowie zwischen der Einführung der Aufgaben und der Arbeit der Schülerinnen und Schüler mit den Aufgaben Bruchstellen liegen, in denen das eigentlich in den Aufgaben zugrunde liegende kognitive Aktivierungspotential reduziert werden kann, was sich letztlich negativ auf den Lernzuwachs der Schülerinnen und Schüler auswirken kann. Demnach bedarf es Strategien (d. h. prozeduralen Wissens), mit denen das lernförderliche Potential der Aufgaben ausgeschöpft werden kann, beispielsweise indem man im Einstieg die Aufgaben über ein alltagsrelevantes Beispiel einführt, das an die Lebenswelt der Schülerinnen und Schüler anknüpft und dem die Schülerinnen und Schüler folglich Relevanz beimessen (Whittington et al., 2021). Kompetenzfördernde Aufgaben, wie sie hier vorgestellt wurden, können für angehende Lehrkräfte sicher eine Bereicherung in ihrer Unterrichtsplanung darstellen, sie allein sind aber noch keine hinreichende Grundlage dafür, dass der dann durchgeführte Unterricht auch wirklich gut ist. Derartige Aufgaben sollten in das komplexe Wechselgefüge der anderen Variablen eingebettet sein, die in der Unterrichtsplanung gleichermaßen zu berücksichtigen sind (d. h. des Lernziels, des Kompetenzstandes der Schülerinnen und Schüler, der Phasenstruktur; Abb. 1; Abb. 3).

*„[...] I have always found that plans are useless, but planning is indispensable“ (Dwight D. Eisenhower)*

## 6 Schluss

### 6.1. Fazit

Ausgangspunkt dieser Arbeit war die Forderung nach empirisch geprüften Qualitätskriterien zur Beurteilung schriftlicher Unterrichtsentwürfe (Krepf & König, 2022b), mit denen der Kritik an der Beurteilungspraxis in der deutschen Lehrkräftebildung (Döbrich & Abs, 2008; Howe, 2006; Kärner et al., 2019; Strietholt & Terhart, 2009) in Bezug auf die Unterrichtsplanung begegnet werden sollte. Angesichts der Relevanz schriftlicher Unterrichtsentwürfe in der 1. und 2. Phase der Lehrkräftebildung (z. B. König et al., 2015; Roloff et al., 2020), schien es notwendig, hier auf dem Wege der Empirie Qualitätskriterien zu entwickeln. Mithilfe einer qualitativen Inhaltsanalyse wurde untersucht, inwiefern ein Zusammenhang zwischen der Vernetzung fachdidaktischen Wissens (ePCK<sub>p</sub>) und der Planungsqualität schriftlicher Unterrichtsentwürfe besteht.

Die Annahme, dass die Quantität der Verbindungen zwischen den PCK-Facetten als Qualitätsindikator gelten kann (Abell, 2008; Chan et al., 2019; Hashweh, 2005; Park & Chen, 2012; Park & Oliver, 2008; Reynolds & Park, 2021; Schiering et al., 2023; Schneider & Plasman, 2011), konnte in der Analyse von Berliner Staatsexamensentwürfen (Beitrag 3: Großmann & Krüger, 2022c) nicht bestätigt werden. Die Qualität eines schriftlichen Unterrichtsentwurfs liegt demnach eher in der Art und Weise, wie die verschiedenen in der Planung zu berücksichtigenden Aspekte aufeinander abgestimmt werden (Problem 1. Art; Abb. 3) und wie diese Entscheidungen begründet werden (Problem 2. Art, Abb. 3).

Auf der Grundlage dieses Befundes wurde ein Kriterienraster entwickelt, das 24 Kriterien auf jeweils drei Niveaustufen umfasst (Beitrag 6: Großmann & Krüger, 2023b) und dass die Verbindungen zwischen den PCK-Facetten fokussiert. Eine umfassende Prüfung auf Validität der Testwertinterpretationen (AERA et al., 2014) lässt den Schluss zu, dass das Kriterienraster geeignet ist, um die Qualität schriftlicher Unterrichtsentwürfe valide einschätzen zu können (Beitrag 7: Großmann & Krüger, 2023a). Somit liegen nun erstmalig empirisch geprüfte Kriterien zur Beurteilung schriftlicher Unterrichtsentwürfe vor.

## 6.2. Implikationen und Desiderate

### 6.2.1. Implikationen für die Forschung zur Unterrichtsplanungskompetenz

Schriftliche Unterrichtsentwürfe stellen ein häufig genutztes Datenmaterial in der naturwissenschaftsdidaktischen Forschung dar (Beitrag 1: Großmann et al., eingereicht). Mithilfe des in dieser Arbeit entwickelten Kriterienrasters lassen sich derartige Forschungsdesigns dergestalt erweitern, dass die Qualität der von den angehenden Lehrkräften entwickelten Entwürfe valide eingeschätzt werden kann. Zudem lässt sich daraus wiederum auf die Unterrichtsplanungskompetenz angehender Lehrkräfte schließen, was zahlreiche Implikationen für die empirische Forschung hat. Beispielsweise lassen sich nun mithilfe der ermittelten Scores Korrelationen zu anderen Konstrukten wie dem mit einem Fragebogen gemessenen pPCK (vgl. Stender & Brückmann, 2020) oder der Unterrichtsqualität berechnen.

Schließt man mithilfe des im Rahmen dieser Arbeit entwickelten Kriterienrasters im Sinne von Blömeke et al. (2015) vom ermittelten Score einer Person (also der Performanz) auf deren Unterrichtsplanungskompetenz, ist Folgendes zu berücksichtigen: Die Annahme, dass der Erwerb von Unterrichtsplanungskompetenz eine langfristige, berufsbiografische Aufgabe ist (z. B. Munthe & Conway, 2017), ist plausibel. Es ist nicht davon auszugehen, dass angehende Lehrkräfte mit dem Ende der formalen Ausbildung im Rahmen der Prüfung zum zweiten Staatsexamen über voll ausgebildete Kompetenzen verfügen (Rothland, 2021). Folglich wäre es inkonsistent, zu behaupten, dass diejenigen angehenden Lehrkräfte, die sehr hohe Scores mithilfe des Kriterienrasters erreichen, über *Expertise* in Bezug auf Unterrichtsplanung verfügen. Zudem werden schriftliche Unterrichtsentwürfe nur im Rahmen der Lehrkräftebildung verfasst, so dass erfahrene Lehrkräfte, denen man eher den *Expertise*-Status zuspricht (Berliner, 2001; Neuweg, 2022) mit langer Berufserfahrung schwerlich als Vergleichsgruppe herangezogen werden können. Empirische Befunde dazu, wie sich Unterrichtsplanungskompetenz entwickelt und worin sich Expertinnen und Experten einerseits sowie Novizinnen und Novizen andererseits unterscheiden, liegen derzeit nur in Ansätzen vor (z. B. Westerman, 1991). Am Beispiel der Unterrichtsplanungskompetenz lässt sich dies mithilfe des Stufenmodells des *Expertise*erwerbs (Dreyfus & Dreyfus, 1987) veranschaulichen (Tab. 6).

Im Novizen-Stadium (*novice*) ist es noch nicht möglich, die Gesamtsituation in ihrer Komplexität zu erfassen und in einer konkreten Situation die wichtigen von den weniger wichtigen Elementen zu unterscheiden. Stattdessen wird versucht, Regeln zu erinnern und diese in einer solchen Situation anzuwenden. Als Beispiel nennt Neuweg (2022) Regeln wie „*Informieren Sie am Beginn des*

Unterrichts über die Lernziele“ (S. 194) – dieser Regel würde eine Novizin oder ein Novize auch dann beharrlich folgen, wenn es in einer Stunde auch andere, funktionalere Einstiegsszenarien gäbe. Novizinnen und Novizen verfügen für diese Abwägung noch nicht über das notwendige Professionswissen und die notwendige Erfahrung.

Im Stadium des fortgeschrittenen Anfängers (*advanced beginner*) gelingt es bereits, neben kontextfreien auch situationale Aspekte zu berücksichtigen. Das bedeutet, dass sich die Unterrichtswahrnehmung ausschärft und nicht mehr bloß kontextfreie Aspekte in konkreten Situationen identifiziert werden, sondern dass kontextabhängige Aspekte wahrgenommen und in die Entscheidungsfindung miteinbezogen werden. Zwar erfolgt die Auffassung der Gesamtsituation weiterhin analytisch ohne Sinn für das Wesentliche. Jedoch gelingt es bereits, neben starren Regeln auch flexiblere Richtlinien zu berücksichtigen und an den jeweiligen Kontext anzupassen. Als Beispiel ließe sich das *Backward Design* (Wiggins & McTighe, 2005) benennen, also eine Planungsstrategie, bei der vom Ende der Stunde und dem zu erreichenden Ziel her geplant wird, ehe methodisch-didaktische Entscheidungen getroffen werden. Diese Planungsstrategie kann als Richtlinie genutzt werden, muss aber auf die jeweils zu planende Stunde eigenständig angepasst werden. Auf dieser Stufe lassen sich viele der in Beitrag 7 (Großmann & Krüger, 2023) untersuchten Master-Studierenden im Praxissemester verorten.

Im Kompetenz-Stadium (*competent*) entwickelt sich die Fähigkeit, die wichtigen von den weniger wichtigen Aspekten einer Situation zu unterscheiden. Auf analytischem Wege gelangt man durch extensive Planung zum Ziel. Hier wird nun nicht mehr bloß Regeln oder Richtlinien gefolgt, sondern es werden bewusste, kontextabhängige Planungsentscheidungen zu den Zielen und der Gestaltung der Lehr-Lern-Struktur getroffen. Dieser Prozess ist insofern extensiv, als er bewusst und zielgerichtet abläuft und zahlreiche Merkmale einer Situation (z. B. Kompetenzstand der Lerngruppe, Rahmenlehrplan, Sachstruktur des Themas, Sequenzplanung) berücksichtigt werden. Auf dieser Stufe lassen sich viele der in den Beiträgen 2-7 untersuchten Referendarinnen und Referendare am Ende der 2. Phase der Lehrkräftebildung verorten. Die Staatsexamensentwürfe stellen die Materialisierung dieser extensiven Planungsprozesse dar, die ausgebildete Lehrkräfte in dieser Form nicht mehr zu leisten brauchen.

Im Stadium des gewandten Könners (*proficient*) verfügt man bereits über viel Erfahrung, die einem dabei hilft, neue Situationen unmittelbar und holistisch wahrzunehmen. Man verfügt inzwischen über eine Sammlung an Erinnerungen an prototypische Situationen – im Sinne von Stender et al. (2017) sogenannte Skripte – und kann diese in Verbindung zu einer neuen Situation setzen. Dieser

Prozess läuft unmittelbar und unbewusst ab – für die konkrete Entscheidungsfindung ist jedoch noch ein Moment des bewussten Planens notwendig.

Im Experten-Stadium (*expert*) schließlich entzieht sich die Verhaltensentscheidung einer bewussten, zielgerichteten Reflexion. Auf der Grundlage umfangreichen Erfahrungswissens werden Situationen unmittelbar und ganzheitlich erfasst und intuitiv in eine Entscheidung überführt. Anders als auf den anderen vier Stufen greifen Expertinnen und Experten allenfalls implizit auf ihr zu Beginn der Ausbildung erworbenes Regelwissen zurück. Sie sind zwar in der Lage, tragfähige Entscheidungen zu treffen, können diese aber laut Neuweg (2018) nicht immer reflektiert begründen, da sie in Entscheidungssituationen intuitiv und ohne bewussten Bezug zu erlerntem Ausbildungswissen getroffen werden.

**Tabelle 6**

*Stufenmodell des Expertiseerwerbs*

	<i>novice</i>	<i>advanced beginner</i>	<i>competent</i>	<i>proficient</i>	<i>expert</i>
<b>Berücksichtigte Faktoren</b>	kontextfrei	kontextfrei, situational	kontextfrei, situational	kontextfrei, situational	kontextfrei, situational
<b>Sinn für das Wesentliche</b>	nein	nein	ja, durch Planung	ja, unmittelbar	ja, unmittelbar
<b>Auffassung der Gesamtsituation</b>	analytisch	analytisch	analytisch	holistisch	holistisch
<b>Orientierung bei der Verhaltensentscheidung</b>	Regeln	Regeln und Richtlinien	extensive Planung	begrenzte Planung	Intuition

*Anmerkung.* Das Stufenmodell des Expertiseerwerbs basiert auf Dreyfus und Dreyfus (1987), die hier vorgenommene Anwendung auf die Unterrichtsplanung entstammt einem Vortrag von Neuweg (2018). Die Benennung der dritten Stufe (*competent*) kollidiert mit dem Kompetenz-Begriff, der in dieser Arbeit in Bezug auf die Unterrichtsplanung gemeint ist; als vollständig unterrichtsplanungskompetent würde man erst auf der fünften Stufe (*expert*) gelten.

Wie oben bereits angedeutet wird in der Lehrkräftebildung einseitig davon ausgegangen, dass erfolgreiches Können den vorangehenden Erwerb von Wissen voraussetzt (Neuweg, 2022). Die bisherigen Forschungsaktivitäten in der empirischen Unterrichtsplanungsforschung nehmen entsprechend Stichproben in den Blick, die man meistens den Stufen *novice* bzw. *advanced beginner* (Studierende) und seltener der Stufe *competent* (Referendarinnen und Referendare) zuordnen könnte. Die Stufen *proficient* (Lehrkräfte in den ersten Berufsjahren; für einen Überblick, siehe Navy et al., 2022) und *expert* (Lehrkräfte mit vielen Berufsjahren Erfahrung) werden kaum berücksichtigt. Dabei ist jedoch zu bedenken, dass die Erfahrung als solche keine hinreichende Bedingung für Expertise darstellt. Berliner (2004) geht davon aus, dass die meisten Lehrkräfte nicht einmal das

Stadium des gewandten Könners (*proficient*; Tab. 6; Dreyfus & Dreyfus, 1987) erreichen, möglicherweise, weil sie trotz jahrelanger Erfahrung über wenig reflektiertes cPCK/pPCK verfügen und ihr Professionswissen nicht hinreichend gut miteinander vernetzen können, um funktionale und damit letztlich lernförderliche Planungsentscheidungen zu treffen. Überlegungen dazu, wie man die Vernetzung von Professionswissen im Allgemeinen (z. B. durch explizit daraufhin ausgerichtete universitäre Seminare, Reynolds et al., 2020) und in Bezug auf die Unterrichtsplanung im Besonderen (z. B. durch Modelle als Planungshilfe, Karlström & Hamza, 2021; z. B. durch *scaffolds*, Zaragoza et al., 2023) fördern kann, sind noch vergleichsweise jung und in ihrer Wirkung bislang empirisch noch nicht hinreichend überprüft. Das im Rahmen dieser Arbeit entwickelte Kriterienraster (Beitrag 6: Großmann & Krüger, 2023b; Beitrag 7: Großmann & Krüger, 2023a; Beitrag 8: Großmann & Krüger, 2022a) stellt eine weitere Möglichkeit dar, Lehrkräfte im Erwerb von Unterrichtsplanungskompetenz zu unterstützen. Dabei ist zu beachten, dass die schriftlichen Unterrichtsentwürfe ihrem Wesen nach auf eine extensive Planung abzielen, so dass in Bezug auf das Schreiben schriftlicher Unterrichtsentwürfe die Stufe *competent* als Zielgröße angemessen ist. Die Stufen *proficient* und *expert* werden vom Kriterienraster nicht adressiert, da diese Lehrkräfte das Instrument selbst gar nicht mehr benötigen würden.

Die weitgehend fehlende Berücksichtigung solcher Lehrkräfte, die man als Expertinnen und Experten charakterisieren könnte, in der empirischen Forschung (Beitrag 1) ist bemerkenswert, denn auf diese Weise besteht die Gefahr, dass sämtliche Bemühungen des emergierenden Forschungsfeldes zur Unterrichtsplanungskompetenz ins Leere laufen: Ohne diejenigen Lehrkräfte in den Blick zu nehmen, denen man Expertise im Bereich der Unterrichtsplanung zusprechen würde, kann man die optimale Performanz in der Unterrichtsplanung nicht beschreiben und die Lehrkräftebildung entsprechend schwerlich so gestalten, dass angehende Lehrkräfte zielgerichtet auf diese höchste Ausprägung der Unterrichtsplanungskompetenz vorbereitet werden:

*Wenn unterrichtliche Planungskompetenz modelliert, operationalisiert und empirisch erfasst wird, dann wäre [...] zuerst bei denen anzusetzen, die über die hier im Fokus stehende Kompetenz bereits in unterschiedlicher Ausprägung verfügen (können): den im Beruf stehenden, routinierten Lehrerinnen und Lehrern. Mit der Modellierung, Operationalisierung und empirischen Erfassung von Lesekompetenz würde auch nicht im Kindergarten oder in der ersten Klasse angesetzt. Anders: Unterrichtsplanungskompetenz als „neues“ Konstrukt lässt sich womöglich besser in potenziell voller Ausprägung identifizieren und empirisch erfassen als im Entstehen, in der Entwicklung. (Rothland, 2021, S. 368)*

Im Sinne von Neuweg (2018, 2022) würde das erfordern, Unterrichtsplanungskompetenz entwicklungsstufensensitiv zu diagnostizieren: Der im Rahmen dieser Arbeit vorgenommene methodische Ansatz, schriftliche Unterrichtsentwürfe zu analysieren, ist gewiss geeignet, um die Unterrichtsplanungskompetenz von Master-Studierenden sowie Referendarinnen und

Referendaren zu untersuchen. Würde man jedoch erfahrene und vor allem erfolgreiche Lehrkräfte (*experts*) dazu auffordern, ebenfalls einen schriftlichen Unterrichtsentwurf für eine Stunde zu verfassen, dann würde man sie zwingen, auf eine Entwicklungsstufe zu regredieren, die sie eigentlich bereits überwunden hatten. Entscheidungen, die sie intuitiv unter holistischer Betrachtung der Gesamtsituation treffen, müssten sie versuchen zu explizieren und auf deklaratives Wissen, Regeln und Richtlinien zurückführen, auf die sie sich im Alltag allenfalls implizit stützen. Neuweg (2018) folgend bestünde ein methodischer Ansatz darin, beispielsweise mithilfe von mikrophänomenologischen *explicitation*-Interviews (Petitmengin, 2006) unmittelbar an das innere Erleben solcher *experts* zu gelangen und zu rekonstruieren, welche Gedanken sich erfahrene Lehrkräfte in der Unterrichtsplanung machen und mithilfe welcher Strategien sie in welchen Situationen wie zu welchen Planungsentscheidungen gelangen. Ein derartiger methodischer Zugang kann insofern fruchtbar sein, als er dazu beitragen könnte, auf der Basis eines nach wie vor fehlenden, konsensuellen Unterrichtsplanungskompetenzstrukturmodells ein Unterrichtsplanungskompetenzentwicklungsmodell bzw. eine *learning progression*<sup>21</sup> zu entwickeln. Demnach wären perspektivisch drei Ansätze in der Unterrichtsplanungsforschung zu verfolgen.

Erstens müsste auf der Basis theoretischer und empirischer Befunde ein Strukturmodell entwickelt werden, das Unterrichtsplanungskompetenz in hypothetisch angenommene Teilkompetenzen und Niveaustufen unterteilt. Diese Struktur müsste dann qualitativ oder quantitativ überprüft werden. Konzeptionell könnte dabei der in dieser Dissertation verfolgte Ansatz fruchtbar sein, heuristische Modelle wie das *CODE-PLAN model* (König et al., 2021a) aus der Unterrichtsplanungsforschung mit professionstheoretischen Konstrukten wie dem PCK (Alonzo et al., 2019; Carlson et al., 2019) zu verknüpfen.

Zweitens müsste vor allem auf Basis empirischer Befunde zur Planungsperformanz von (angehenden) Lehrkräften ein hypothetischer Lernweg entwickelt werden. Längsschnittliche Analysen, die die Wirkung von Interventionen innerhalb der Lehrkräftebildung untersuchen, wären zu diesem Zwecke notwendig. Wie in Beitrag 1 (Großmann et al., eingereicht) gezeigt wurde, sind derartige Längsschnittstudien in der empirischen Unterrichtsplanungsforschung bislang noch eine Ausnahme, und aufgrund fehlender Experimental-Kontrollgruppen-Designs sind Aussagen zu kausalen Zusammenhängen zwischen bestimmten Variablen aus dem Modell der professionellen Handlungskompetenz (Baumert & Kunter, 2006) und der Unterrichtsplanungskompetenz bisher

---

<sup>21</sup> Zum Unterschied zwischen Kompetenzmodellen und *learning progressions*, siehe Upmeyer zu Belzen et al. (2019). An dieser Stelle wird der dort vertretenen Auffassung gefolgt, dass Kompetenzmodelle in der Regel als Strukturmodelle aufzufassen sind, während *learning progressions* in der Regel Entwicklungsmodelle darstellen (Ebd., S.260). Die Unterscheidung zwischen Struktur- und Entwicklungsmodellen basiert auf Schecker und Parchmann (2006).

kaum belastbar. Eine solche *learning progression* könnte beispielsweise wie von Neuweg (2018) vorgeschlagen auf Basis des Modells des Expertiseerwerbs (Dreyfus & Dreyfus, 1987) entwickelt werden. Wenngleich es sich nur um querschnittliche Analysen handelt, können die Befunde aus Beitrag 7 (Großmann & Krüger, 2023a) einen Teil dazu beitragen, die Progression in der Entwicklung von Unterrichtsplanungskompetenz zu beschreiben.

Auf diese Weise erscheint es möglich, die von Neuweg (2022) beschriebene Dichotomie aufzulösen: Er problematisiert, dass die empirische Forschung bislang einseitig vom Wissen her denkt und Können als bloße Anwendung in der 1. Phase der Lehrkräftebildung erworbenen Wissens betrachtet. Er schlägt vor, stärker vom Können derjenigen her zu denken, denen man den Expertise-Status zusprechen würde, und davon ausgehend zu rekonstruieren, wie Professionalisierung hin zu diesem Können gestaltet sein könnte. Der oben genannte Vorschlag würde diese Dichotomie überwinden, denn einerseits wäre ein vor allem theoretisch zu entwickelndes Unterrichtsplanungskompetenzstrukturmodell zu entwickeln und empirisch zu überprüfen, und andererseits ein Unterrichtsplanungskompetenzentwicklungsmodell vorzuschlagen, das vom Können erfahrener, erfolgreicher Lehrkräfte her den Kompetenzerwerb in Teilschritten zerlegt. Eine derartige Prüfung des Zusammenhangs würde die Beobachtung des durchgeführten Unterrichts erfordern, die für die Teilstichprobe der Referendarinnen und Referendare im Staatsexamen schon aus prüfungsrechtlichen Gründen nicht möglich gewesen wäre. Aufschlussreich und methodisch umsetzbar könnte die Anwendung des Kriterienrasters im Rahmen des Praxissemesters sein, in dem Dozierende aus der 1. Phase der Lehrkräftebildung ohnehin schriftliche Unterrichtsentwürfe erhalten und die Studierenden an den Schulen besuchen, den Unterricht beobachten und gemeinsam reflektieren. Dieser Schritt wurde im Rahmen der vorliegenden Arbeit jedoch noch nicht unternommen.

Drittens wäre über diese beiden generischen Aspekte hinaus eine noch stärker explizit biologiedidaktische Perspektive auf die Unterrichtsplanung einzunehmen. Scholl et al. (2019) stellen fest, dass das Verhältnis von Generik und Fachspezifik in Bezug auf Unterrichtsplanungskompetenz weiterhin unbestimmt ist. Dies stellte auch Breidenstein (2021) unlängst fest. Eine stärker biologiespezifische Perspektive wirft jedoch unmittelbar die Frage auf, mit welchem Körnungsgrad man sich den Unterrichtsgegenständen annähert, zu denen geplant werden soll. So wie in der PCK-Forschung weitgehend fraglich bleibt, was unter Disziplin-, Domänen-, Themen- und Konzeptspezifität zu verstehen ist, stellt sich diese Frage für die Unterrichtsplanung gleichermaßen: Inwiefern ist überhaupt davon auszugehen, dass eine (angehende) Lehrkraft unterschiedlich erfolgreich planen kann, wenn sie zu verschiedenen biologischen Themen (z. B. Photosynthese versus Blutkreislauf) oder Basiskonzepten (z. B.



Struktur und Funktion versus evolutive Entwicklung) plant? Angesichts der eher generischen Überlegungen, die in Planungsprozessen angestellt werden (Abb. 1), lässt sich annehmen, dass vor allem das von Thema zu Thema unterschiedlich ausgeprägte CK von Lehrkräften einen maßgeblichen Einfluss auf die Unterrichtsplanungskompetenz haben könnte. Zumindest wurde wiederholt gezeigt, dass CK einen Prädiktor für die Ausprägung von PCK darstellt (z. B. Großschedl et al., 2015; Kind & Chan, 2019; Şen et al., 2022). Bedeutsam für die Unterrichtsplanungsforschung könnte in diesem Zusammenhang der Vorschlag von Mavhunga und Ndlovu (2023) sein, das Konstrukt CK in akademisches Fachwissen einerseits und unterrichtsbezogenes Fachwissen (*teacher-related science content knowledge*, TerSCK) andererseits zu unterteilen. TerSCK adressiert analog zur fachlichen Klärung im Modell der didaktischen Rekonstruktion (Duit et al., 2012) eine didaktische Perspektive, so dass die Fachinhalte als Lerngegenstände in einem Schulfach (hier: Biologie) betrachtet werden und nicht in der Komplexität, mit der sich die jeweilige Fachwissenschaft (hier: z. B. Molekulargenetik, Neurophysiologie, aquatische Ökologie) den Forschungsgegenständen annähert. Insofern wäre zu fragen, inwiefern Lehrkräfte weniger über fachwissenschaftliches CK, sondern eher über schulrelevantes TerSCK verfügen sollten und inwiefern dieses mit der Unterrichtsplanungskompetenz korreliert. Möglicherweise gelingt es Lehrkräfte mit ausgeprägtem TerSCK zu einem bestimmten Thema guten Unterricht zu planen (z. B. gemessen mithilfe eines Kriterienrasters; Großmann & Krüger, 2023a), zu einem anderen Thema, zu dem sie über weniger elaboriertes TerSCK verfügen, hingegen nicht. Mehr noch: Möglicherweise korreliert die Unterrichtsplanungskompetenz eher mit TerSCK gemessen in Form noch zu entwickelnder Instrumente und weniger mit CK, das in verfügbaren Instrumenten möglicherweise eher akademisches Fachwissen als für den Unterricht notwendiges Fachwissen adressiert (z. B. enthält der CK-Test von Großschedl et al. (2014) Items darüber, wie Muschelschalen aufgebaut sind und worin sich natürliche von künstlichen Miesmuschelbänken unterscheiden – Aspekte, die in der marinen Ökologie sicher wichtiges CK darstellen, aber für die Schule eher nicht dem TerSCK zuzuordnen sind).

Noch globaler betrachtet liegt derzeit auch keine Evidenz dazu vor, ob Lehrkräfte in ihren beiden Fächern gleichermaßen erfolgreich oder weniger erfolgreich planen können. Gerade das deutsche Bildungssystem, in dem Lehrkräfte in der Regel in zwei Fächern ausgebildet werden, bietet die Möglichkeit, diesen möglichen Zusammenhang vergleichend zu untersuchen. So ist beispielsweise fraglich, ob Unterrichtsplanungskompetenz als so generisches Konstrukt aufzufassen ist, dass eine Lehrkraft, die in Biologie guten Unterricht planen kann, dies in gleicher Weise auch im Fach Deutsch, Geschichte oder Englisch kann, oder ob nicht jedem Fach eine innere Logik innewohnt,

die sich dann im CK bzw. TerSCK, somit im PCK und letztlich auch im ePCK<sub>p</sub> niederschlägt. Jüngste theoretische (Müller & Mahler, in Druck) und empirische biologiedidaktische Forschungsvorhaben (Reinisch & Fricke, 2022) zeigen über die von Neuhaus (2021) benannten eher biologiespezifischen Unterrichtsqualitätsmerkmale hinaus, dass die Biologie Eigenheiten aufweist, die sie maßgeblich von anderen Fächern im Allgemeinen (z. B. durch empirisches Arbeiten: Kompetenzbereich Erkenntnisgewinnung; KMK, 2005, KMK, 2020) und von den anderen naturwissenschaftlichen Fächern im Besonderen (z. B. durch die Variabilität und Komplexität biologischer Systeme und damit verbundene ethische Aspekte im Umgang mit Lebewesen) unterscheidet. Inwiefern solche Besonderheiten der jeweiligen Fächer in der Unterrichtsplanung berücksichtigt werden und welche Herausforderungen dies mit sich bringt, sind derzeit noch offene Fragen der fachdidaktischen Unterrichtsplanungsforschung.

### 6.2.2. Implikationen für die Lehrkräftebildung

Nicht bloß das Kriterienraster als Ergebnis dieses Forschungsvorhabens, sondern auch der Entwicklungsprozess kann als Versuch verstanden werden, die Potentiale der theoretischen 1. sowie der praktischen 2. Phase zu nutzen und phasenübergreifend zu arbeiten. Bislang sind solche phasenübergreifend angelegten empirischen Forschungsprojekte rar (Anderson-Park & Abs, 2020; Kunz & Uhl, 2022), obwohl sie gerade in Bezug auf professionstheoretische Fragestellungen zwingend nötig wären, da Professionalisierungsprozesse keineswegs allein in der 1. Phase stattfinden und dann abgeschlossen sind. Weber und Czerwenka (2021) stellen diesbezüglich fest, dass

*die Kooperation und Anschlussfähigkeit zwischen den beiden Phasen bislang einen deutlichen personellen, zeitlichen und finanziellen Mehraufwand für alle an der Ausbildung beteiligten Institutionen [...] bedeuten, so dass sich die Frage stellt, inwieweit dieser Mehraufwand gerechtfertigt scheint und wie dieser ggf. reduziert werden kann. (Weber & Czerwenka, 2021, S. 261)*

Möglicherweise kann das hier entwickelte Kriterienraster insofern einen Beitrag zur Reduzierung des Mehraufwandes der Beteiligten und zur Stärkung der Anschlussfähigkeit zwischen 1. und 2. Phase leisten, als die Kriterien sowohl in universitären Seminaren im Praxissemester als auch an Studienseminaren im Referendariat genutzt werden können und die Entwicklung und Anpassung jeweils eigener Instrumente obsolet wird – die Bereitschaft vorausgesetzt, statt eines in Teilen auf subjektiven Präferenzen basierenden und oft nicht aussagekräftigen eigenen Instruments (Strietholt & Terhart, 2009) ein externes zu nutzen, das bei Bedarf angepasst werden könnte.

In Bezug auf die Unterrichtsplanungskompetenz kann ein solcher phasenübergreifender Zugang insbesondere für die angehenden Lehrkräfte von Nutzen sein. Die adressatengerecht in den Beiträgen 8 (Großmann & Krüger, 2022a) und 9 (Großmann et al., 2022) vorgestellten und an Beispielen illustrierten Kriterien können als Orientierungsrahmen bereits in der 1. Phase genutzt werden, beispielsweise in der Vorbereitung auf das Praxissemester im Master-Studium. Zwei Ideen zum weiteren Umgang mit dem Kriterienraster in der Lehrkräftebildung sollen abschließend kurz skizziert werden:

Erstens wurde bereits angedeutet, dass das Kriterienraster nicht bloß im Sinne summativen Assessments durch Dozierende sowie Fachseminarleiterinnen und Fachseminarleiter genutzt werden kann, sondern dass es auch formativ für Self- und Peer-Assessment genutzt werden könnte. Zwar zeigen Ozogul et al. (2008), dass der Lernzuwachs bei der Bewertung von Unterrichtsentwürfen am größten ist, wenn Dozierende mit ihrer größeren Erfahrung und ihrer solideren Wissensbasis Feedback mit einem Kriterienraster geben. Doch auch Peer- und Self-Assessment führten zu signifikanten Lernzuwächsen und scheinen somit auch hilfreich sein zu können. In Bezug auf das hier entwickelte Kriterienraster würde das erfordern, dass es sach- und adressatengerecht eingeführt, Verständnisschwierigkeiten in den Formulierungen der Niveaustufen besprochen und der Umgang mit dem Kriterienraster schrittweise eingeübt wird. Ohne eine solche behutsame Heranführung können analytische Kriterienraster überfordernd sein (Brookhart & Chen, 2015; Chan & Ho, 2019; Dawson, 2017). In Bezug auf die hier untersuchten Stichproben wäre eine Heranführung im Praxissemester in der 1. Phase sicher möglich, ehe das Kriterienraster dann in der 2. Phase bei den fünf Unterrichtsbesuchen pro Fach im Land Berlin von den Referendarinnen und Referendaren zum Self-Assessment oder durch Hospitierende zum Peer-Assessment genutzt werden kann. Da die Performanz, die spätestens zur zweiten Staatsexamensprüfung erwartet wird, in der höchsten Niveaustufe transparent gemacht wird, können die angehenden Lehrkräfte auf diese Weise üben und ihren eigenen Lernfortschritt überwachen. Sowohl für das Self-Assessment als auch für das Peer-Assessment sind zahlreiche lernförderliche Effekte beschrieben worden, z. B. eine Reduzierung des *cognitive loads* bei der Beurteilung von Lernprodukten und eine präzisere Wahrnehmung gelungener und weniger gelungener Aspekte (Krebs et al., 2022), eine gesteigerte Lernmotivation (Brookhart & Chen, 2015) und gesteigerte Selbstregulationsfähigkeiten (Panadero & Jonsson, 2013).

Zweitens ließe sich – im besten Falle basierend auf einem wie oben geschildert noch zu entwickelnden Unterrichtsplanungskompetenzentwicklungsmodell bzw. einer *learning progression* – ein Konzept entwickeln, wie das Schreiben von Unterrichtsentwürfen möglichst ertragreich gefördert werden kann. Bislang gibt es nur wenige Studien, die sich explizit mit der Frage befassen,

wie Unterrichtsplanungskompetenz in der Lehrkräftebildung erfolgreich gefördert werden kann (Drost & Levine, 2015; Karlström & Hamza, 2021; Zaragoza et al., 2023). Innovativ ist insbesondere der Ansatz von Zaragoza et al. (2023), die mithilfe eines *scaffolds* versuchen, die Brücke zwischen Theorie und Praxis zu schlagen und angehende Lehrkräfte dabei zu unterstützen, ihr Professionswissen im Planungsprozess erfolgreich anzuwenden. Sie sind somit innerhalb des Modells professioneller Kompetenz von Lehrkräften (Baumert et al., 2010) zu verorten und knüpfen unwissentlich an die dieser Arbeit zugrunde liegende Definition von Unterrichtsplanungskompetenz an (Abb. 3). Schließlich ist die Idee, dass angehende Lehrkräfte nicht bloß viel Professionswissen erwerben müssen, sondern dass sie dieses Wissen in problemhaltigen Situationen auch anwenden können müssen, Kern der Kompetenzauffassung von Blömeke et al. (2015). Dem *scaffold* von Zaragoza et al. (2023) folgend wären die Analyse von Unterrichtsentwürfen einerseits und die eigene Unterrichtsplanung andererseits zwei Schritte eines rekursiven Lernprozesses. In Bezug auf das hier entwickelte Kriterienraster würde das bedeuten, dass man zunächst lernt, die Kriterien auf Unterrichtsentwürfe anzuwenden, die von anderen Personen geschrieben worden sind. Es wäre auch denkbar, dass Unterrichtsentwürfe durch die Auszubildenden fingiert werden, so dass bestimmte Kriterien besonders pointiert und deutlich erkennbar gelungen bzw. nicht gelungen sind. Auf diese Weise würde es zu Beginn leichter fallen, die verschiedenen Niveaustufen in der Anwendung des Kriterienrasters auseinanderhalten zu können. Erst im Anschluss planen sie selbst eine Stunde, schreiben einen Unterrichtsentwurf und führen diesen bestenfalls (im Praxissemester oder im Referendariat) auch durch. Anschließend reflektieren sie allein oder gemeinsam den Unterrichtsentwurf und prüfen, inwiefern das angestrebte Ziel im Unterricht erreicht wurde. Das *scaffolding* bestünde darin, nicht alle 24 Kriterien auf einmal einzuführen, sondern beispielsweise mit besonders wichtigen wie der Formulierung eines Lernziels (z. B. Anderson & Kratwohl, 2014; Wittwer et al., 2020) oder der Strukturierung (z. B. Krepf & König, 2022a, 2022b) zu beginnen und erst später anspruchsvollere Kriterien wie die Berücksichtigung von Schülervorstellungen (z. B. Hammann & Asshoff, 2014; Kattmann, 2015) oder eine Reduzierung der Fachbegriffe (Dorfner et al., 2020) zu verlangen.

Inwieweit das Kriterienraster bereits in einer Form vorliegt, die für angehende Lehrkräfte nachvollziehbar und verständlich ist, wurde im Rahmen dieser Arbeit und der Entwicklung des Validitätsarguments noch nicht geprüft. Da Unterrichtsplanungskompetenz wie gezeigt wurde weit mehr als eine „*Jedermannskompetenz*“ darstellt, ist es ratsam, angehende Lehrkräfte schrittweise an das Kriterienraster heranzuführen, damit sie guten Biologieunterricht planen können.

### 6.2.3. Desiderate

Zusammenfassend ergeben sich die folgenden spezifischen Fragestellungen in Bezug auf das entwickelte Kriterienraster (1-5), und die folgenden allgemeineren Fragestellungen (6-9) für die zukünftige Forschung zur Unterrichtsplanungskompetenz:

- 1) Inwiefern erfüllt das Kriterienraster die Performanzerwartungen von in der Lehrkräftebildung der 1. und 2. Phase tätigen Personen jenseits Berlins und stellt somit ein bundesweit einsetzbares Instrument dar?
- 2) Inwiefern lassen sich die Kriterien des Kriterienrasters nach Relevanz gewichten?
- 3) Inwiefern gelingt es angehenden (Biologie-) Lehrkräften, mithilfe des Kriterienrasters als Lernhilfe besseren Unterricht zu planen als solchen ohne das Kriterienraster?
- 4) Inwiefern unterscheidet sich die Qualität eines zweiten Unterrichtsentwurfs angehender Biologielehrkräfte in einem Prä-Post-Design, nachdem sie zu einem ersten Unterrichtsentwurf a) Feedback von Dozierenden, b) Feedback von einer Kommilitonin oder einem Kommilitonen (*peer assessment*) bekommen oder c) sich selbst Feedback mithilfe des Kriterienrasters gegeben haben (*self-assessment*)?
- 5) Inwiefern ist das Kriterienraster im Sinne prognostischer Validität geeignet, die Qualität der Unterrichtsdurchführung vorherzusagen?
- 6) Auf welche Weise lässt sich die Integration von konzeptuellem Wissen zu prozeduralem Wissen in den Planungsprozessen von (Biologie-) Lehrkräften mit großer Expertise in der Unterrichtsplanung methodisch valide erfassen?
- 7) a) Welche und wie viele Teilkompetenzen von Unterrichtsplanungskompetenz lassen sich in einem Kompetenzstrukturmodell ausdifferenzieren und empirisch belegen?
- 7) b) Welche qualitativen Kompetenzabstufungen lassen sich für (Biologie-) Lehrkräfte mit hoher und niedriger Unterrichtsplanungskompetenz empirisch unterscheiden?
- 8) Wie entwickelt sich die Unterrichtsplanungskompetenz angehender (Biologie-) Lehrkräfte vom Bachelor- und Masterstudium (1. Phase der Lehrkräftebildung) über das Referendariat (2. Phase der Lehrkräftebildung) und darüber hinaus (3. Phase der Lehrkräftebildung)?
- 9) Auf welche Weise werden biologiespezifische Unterrichtssqualitätsmerkmale (z. B. Emergenz komplexer Systeme, Ethik, Variabilität von Lebewesen) von (angehenden) Lehrkräften in der Planung berücksichtigt und welche Herausforderungen begegnen ihnen dabei?

## Literaturverzeichnis

- Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, 30(10), 1405–1416. <https://doi.org/10.1080/09500690802187041>
- AERA, APA & NCME. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human*, 50, 179–211.
- Alonzo, A. C., Berry, A. & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: enacted and personal PCK. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 271–286). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_12](https://doi.org/10.1007/978-981-13-5898-2_12)
- Anderson, L. W. & Kratwohl, D. R. (2014). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Always Learning. Pearson.
- Anderson-Park, E. & Abs, H. J. (2020). Lehrerinnen- und Lehrerbildung im Vorbereitungsdienst. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 332–338). Julius Klinkhardt.
- Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College teaching*, 53(1), 27–31.
- Australian Institute for Teaching and School Leadership. (2018). *Australian Professional Standards for Teachers*. <https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standards-for-teachers.pdf>
- Aydin, S. & Boz, Y. (2013). The nature of integration among PCK components: a case study of two experienced chemistry teachers. *Chemistry Education Research and Practice*, 14(4), 615–624. <https://doi.org/10.1039/C3RP00095H>
- Backfisch, I., Lachner, A., Hische, C., Loose, F. & Scheiter, K. (2020). Professional knowledge or motivation? Investigating the role of teachers' expertise on the quality of technology-enhanced lesson plans. *Learning and Instruction*, 66, Artikel 101300. <https://doi.org/10.1016/j.learninstruc.2019.101300>
- Baer, M., Kocher, M., Wyss, C., Guldemann, T., Larcher, S. & Dörr, G. (2011). Lehrerbildung und Praxiserfahrung im ersten Berufsjahr und ihre Wirkung auf die Unterrichtskompetenzen von Studierenden und jungen Lehrpersonen im Berufseinstieg. *Zeitschrift für Erziehungswissenschaft*, 14(1), 85–117. <https://doi.org/10.1007/s11618-011-0168-5>
- Bahçivan, E. (2016). Investigating the relationships among PSTs' teaching beliefs: are epistemological beliefs central? *Educational Studies*, 42(2), 221–238. <https://doi.org/10.1080/03055698.2016.1160823>
- Barke, H.-D., Harsch, G., Kröger, S. & Marohn, A. (2018). Unterrichtsziele. In H.-D. Barke, G. Harsch, S. Kröger & A. Marohn (Hrsg.), *Chemiedidaktik kompakt* (S. 89–153). Springer Spektrum. [https://doi.org/10.1007/978-3-662-56492-9\\_4](https://doi.org/10.1007/978-3-662-56492-9_4)
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>

- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Beckmann, T. & Ehmke, T. (2023). Informal and formal lesson planning in school internships: Practices among pre-service teachers. *Teaching and Teacher Education*, 132, Artikel 104249. <https://doi.org/10.1016/j.tate.2023.104249>
- Begrich, L., Praetorius, A.-K., Decristan, J., Fauth, B., Göllner, R., Herrmann, C., Kleinknecht, M. P., Taut, S. & Kunter, M. (2023). Was tun? Perspektiven für eine Unterrichtsqualitätsforschung der Zukunft. *Unterrichtswissenschaft*(51), 63–97. <https://doi.org/10.1007/s42010-023-00163-4>
- Behling, F., Förtsch, C. & Neuhaus, B. J. (2022). The Refined Consensus Model of Pedagogical Content Knowledge (PCK): Detecting Filters between the Realms of PCK. *Education Sciences*, 12(9), Artikel 592, 1–21. <https://doi.org/10.3390/educsci12090592>
- Bergqvist, A. & Chang Rundgren, S.-N. (2017). The influence of textbooks on teachers' knowledge of chemical bonding representations relative to students' difficulties understanding. *Research in Science & Technological Education*, 35(2), 215–237. <https://doi.org/10.1080/02635143.2017.1295934>
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463–482. [https://doi.org/10.1016/S0883-0355\(02\)00004-6](https://doi.org/10.1016/S0883-0355(02)00004-6)
- Berliner, D. C. (2004). Describing the Behavior and Documenting the Accomplishments of Expert Teachers. *Bulletin of Science, Technology & Society*, 24(3), 200–212. <https://doi.org/10.1177/0270467604265535>
- Beyer, C. J. & Davis, E. A. (2012). Learning to critique and adapt science curriculum materials: Examining the development of preservice elementary teachers' pedagogical content knowledge. *Science Education*, 96(1), 130–157. <https://doi.org/10.1002/sce.20466>
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Bravo, P. & Cofré, H. (2016). Developing biology teachers' pedagogical content knowledge through learning study: the case of teaching human evolution. *International Journal of Science Education*, 38(16), 2500–2527. <https://doi.org/10.1080/09500693.2016.1249983>
- Breidenstein, G. (2021). Interferierende Praktiken. *Zeitschrift für Erziehungswissenschaft*, 24(4), 933–953. <https://doi.org/10.1007/s11618-021-01037-0>
- Brodhacker, S. (2014). *Unterrichtsplanungskompetenz im Praktikum: Einflussfaktoren auf die Veränderung der wahrgenommenen Kompetenz von Studierenden. Empirische Erziehungswissenschaft: Bd. 50.* Waxmann. <https://elibrary.utb.de/doi/book/10.31244/9783830980971>
- Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens* (1. Auflage). *Huber-Psychologie-Forschung*. Huber.
- Bromme, R. & Seeger, F. (1979). *Unterrichtsplanung als Handlungsplanung: Eine psychologische Einführung in die Unterrichtsvorbereitung*. Scriptor.
- Brookhart, S. M. (2018). Appropriate criteria: key to effective rubrics. *Frontiers in Education*, 3(22), 1–12. <https://doi.org/10.3389/feduc.2018.00022>
- Brookhart, S. M. & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343–368. <https://doi.org/10.1080/00131911.2014.929565>

- Brophy, J. E. & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching: A project of the American Educational Research Association* (3. ed., S. 328–375). Macmillan.
- Brown, P. L., Friedrichsen, P. M. & Abell, S. K. (2013). The development of prospective secondary biology teachers PCK. *Journal of Science Teacher Education*, 24(1), 133–155. <https://doi.org/10.1007/s10972-012-9312-1>
- Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik*, 39(2), 257–284. <https://doi.org/10.1007/S13138-017-0122-Z>
- Bullough, R. (1987). Planning and the first year of teaching. *Journal of Education for Teaching*, 13(3), 231–250. <https://doi.org/10.1080/0260747870130303>
- Calderhead, J. (1996). Teachers' beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Hrsg.), *Handbook of educational psychology* (S. 709–725). Macmillan.
- Campbell, T., Gray, R., Fazio, X. & van Driel, J. H. (2022). Research on Secondary Science Teacher Preparation. In J. A. Luft & M. G. Jones (Hrsg.), *Handbook of Research on Science Teacher Education* (S. 97–118). Routledge. <https://doi.org/10.4324/9781003098478-10>
- Capps, D. K. & Crawford, B. A. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, 24(3), 497–526. <https://doi.org/10.1007/s10972-012-9314-z>
- Carlson, J., Daehler, K. R., Alonzo, A., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Chan, K. K. H., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze\_Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 77–92). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_2](https://doi.org/10.1007/978-981-13-5898-2_2)
- Causton-Theoharis, J. N., Theoharis, G. T. & Trezek, B. J. (2008). Teaching pre-service teachers to design inclusive instruction: a lesson planning template. *International Journal of Inclusive Education*, 12(4), 381–399 (19 Seiten). <http://www.informaworld.com/openurl?genre=article&id=doi:10.1080/13603110601156509>
- Chan, K. K. H. (2022). A critical review of studies using the pedagogical content knowledge map approach. *International Journal of Science Education*, 1–27. <https://doi.org/10.1080/09500693.2022.2035011>
- Chan, K. K. H. & Hume, A. (2019). Towards a consensus model: literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 3–76). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_1](https://doi.org/10.1007/978-981-13-5898-2_1)
- Chan, K. K. H., Rollnick, M. & Gess-Newsome, J. (2019). A grand rubric for measuring science teachers' Pedagogical Content Knowledge. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 253–271). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_11](https://doi.org/10.1007/978-981-13-5898-2_11)
- Chan, Z. & Ho, S. (2019). Good and bad practices in rubrics: the perspectives of students and educators. *Assessment & Evaluation in Higher Education*, 44(4), 533–545. <https://doi.org/10.1080/02602938.2018.1522528>
- Clement, J. (2008). *Creative model construction in scientists and students*. Springer.



- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. Murray. <http://digital.wlb-stuttgart.de/purl/bsz350941645>
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>
- Decristan, J., Hess, M., Holzberger, D. & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale: Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. *Zeitschrift für Pädagogik*, 66(1), 102–116.
- Demirdöğen, B., Hanuscin, D. L., Uzuntiryaki-Kondakçı, E. & Köseoğlu, F. (2016). Development and nature of preservice chemistry teachers' pedagogical content knowledge for nature of science. *Research in Science Education*, 46(4), 575–612. <https://doi.org/10.1007/s11165-015-9472-z>
- Döbrich, P. & Abs, H. J. (2008). Evaluation der zweiten Phase der Lehrerbildung. Pädagogische Entwicklungsbilanzen mit Studienseminaren in Hessen. *Schulverwaltung, Hessen, Rheinland-Pfalz*, 13(3), 70–73.
- Dorfner, T., Förtsch, C. & Neuhaus, B. J. (2020). Use of technical terms in German biology lessons and its effects on students' conceptual learning. *Research in Science & Technological Education*, 38(2), 227–251. <https://doi.org/10.1080/02635143.2019.1609436>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humantwissenschaften* (5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). *Springer-Lehrbuch*. Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Dörner, D. (2021). *Die Logik des Misslingens: Strategisches Denken in komplexen Situationen* (17. Aufl.). *Rowohlt: Bd. 61578*. Rowohlt.
- Dreyfus, H. L. & Dreyfus, S. E. (1987). *Künstliche Intelligenz: Von den Grenzen der Denkmaschine und dem Wert der Intuition*. Rowohlt.
- Drost, B. R. & Levine, A. C. (2015). An analysis of strategies for teaching standards-based lesson plan alignment to preservice teachers. *Journal of Education*, 195(2), 37–47.
- Duit, R., Gropengießer, H., Kattmann, U., Komorek, M. & Parchmann, I. [Ilka]. (2012). The Model of Educational Reconstruction – a Framework for Improving Teaching and Learning Science<sup>1</sup>. In D. Jorde & J. Dillon (Hrsg.), *Science Education Research and Practice in Europe* (S. 13–37). SensePublishers. [https://doi.org/10.1007/978-94-6091-900-8\\_2](https://doi.org/10.1007/978-94-6091-900-8_2)
- Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz - Teilkompetenz "Bewerten, Entscheiden und Reflektieren" für Gestaltungsaufgaben Nachhaltiger Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 177–197.
- Ekiz-Kiran, B. & Boz, Y. (2020). Interactions between the science teaching orientations and components of pedagogical content knowledge of in-service chemistry teachers. *Chemistry Education Research and Practice*, 21(1), 95–112. <https://doi.org/10.1039/c9rp00092e>
- Engelhardt, M. von. (1982). *Die pädagogische Arbeit des Lehrers: Eine empirische Einführung*. Schöningh.
- Engelschalt, P., Penzlin, J. & Upmeyer zu Belzen, A. (2022). Erklärungen finden - die Lärchennadel im Streuhaufen. In L. Großmann, S. H. Nessler & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 123–129). Cornelsen.
- Enugu, R. & Hokayem, H. (2017). Challenges pre-service teachers face when implementing a 5E inquiry model of instruction. *European Journal of Science and Mathematics Education*, 5(2), 178–209.

- Erduran, S. & Dagher, Z. R. (2014). *Reconceptualizing the nature of science for science education: Scientific knowledge, practices and other family categories*. Springer.
- Fabrigar L. R., Wegener, D. T., MacCallum R. C. & Strahan E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*(4), 272–299.
- Fang, F. C. & Casadevall, A. (2015). Competitive science: is competition ruining science? *Infection and immunity*, 83(4), 1229–1233. <https://doi.org/10.1128/IAI.02939-14>
- Fladung, I. (2022). *Deutschunterricht im Vorbereitungsdienst adaptiv planen: Eine empirische Studie zum Stellenwert von Diagnostik und Differenzierung in schriftlichen Unterrichtsplanungen* (1. Auflage). Waxmann.
- Fleige, J., Seegers, A., Upmeyer zu Belzen, A. & Krüger, D. (Hrsg.). (2012). *Modellkompetenz im Biologieunterricht 7-10: Phänomene begreifbar machen - in 11 komplett ausgearbeiteten Unterrichtseinheiten*. Auer.
- Forbes, C. T. (2011). Preservice elementary teachers' adaptation of science curriculum materials for inquiry-based elementary science. *Science Education*, 95(5), 927–955. <https://doi.org/10.1002/sce.20444>
- Forbes, C. T. & Davis, E. A. (2010). Curriculum design for inquiry: Preservice elementary teachers' mobilization and adaptation of science curriculum materials. *Journal of Research in Science Teaching*, 47(7), 820–839. <https://doi.org/10.1002/tea.20379>
- Förtsch, C., Werner, S., Kotzebue, L. von & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. <https://doi.org/10.1080/09500693.2016.1257170>
- Friedrichsen, P. M., Abell, S. K., Pareja, E. M., Brown, P. L., Lankford, D. M. & Volkmann, M. J. (2008). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, 46(4), 357–383. <https://doi.org/10.1002/tea.20283>
- Funke, J. (2003). *Problemlösendes Denken* (1. Aufl.). *Einführungen und Allgemeine Psychologie*. Kohlhammer. <http://gbv.ebib.com/patron/FullRecord.aspx?p=1613642>
- Funke, J. & Glodowski, A.-S. (1990). Planen und Problemlösen: Überlegungen zur neuropsychologischen Diagnostik von Basiskompetenzen beim Planen. *Zeitschrift für Neuropsychologie*, 1(2), 139–148.
- Gassmann, C. (2013). *Erlebte Aufgabenschwierigkeit bei der Unterrichtsplanung: Eine qualitativ-inhaltsanalytische Studie zu den Praktikumsphasen der universitären Lehrerbildung*. Springer. <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10634879> <https://doi.org/10.1007/978-3-658-00243-5>
- Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Hrsg.), *Examining pedagogical content knowledge* (S. 3–17). Kluwer Academic Publishers.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, P. M. Friedrichsen & J. Loughran (Hrsg.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (S. 28–42). Routledge.
- Glaser, R. & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, 40, 631–666.

- Glaubrecht, M. (2014). *Am Ende des Archipels - Alfred Russel Wallace* (2. Aufl.). Galiani. <http://www.lehmanns.de/midvox/bib/9783869710709>
- Gobbo, C. & Chi, M. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development*, 1(3), 221–237. [https://doi.org/10.1016/S0885-2014\(86\)80002-8](https://doi.org/10.1016/S0885-2014(86)80002-8)
- Gohli, J. & Voje, K. L. (2016). An interspecific assessment of Bergmann's rule in 22 mammalian families. *BMC evolutionary biology*, 16(1), 222. <https://doi.org/10.1186/s12862-016-0778-x>
- Göhner, M. & Krell, M. (2020). Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review. *Zeitschrift für Didaktik der Naturwissenschaften*, 26, 207–225. <https://doi.org/10.1007/s40573-020-00111-0>
- Graf, D. (2007). Die Theorie des geplanten Verhaltens. In D. Krüger & H. Vogt (Hrsg.), *Springer-Lehrbuch. Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 33–43). Springer.
- Greiten, S. (2014). Welche Kompetenzen für die Unterrichtsplanung benötigen LehrerInnen an Regelschulen für einen inklusiven, auf individuelle Förderung ausgerichteten Unterricht? Erste Ergebnisse aus einer qualitativ-empirischen Studie. In S. Trumpp, S. Seifried, E. Franz & T. Klauß (Hrsg.), *Inklusive Bildung: Erkenntnisse und Konzepte aus Fachdidaktik und Sonderpädagogik* (107-121). Beltz Juventa.
- Gropengießer, H. (2013). Unterricht planen. In H. Gropengießer, U. Harms & U. Kattmann (Hrsg.), *Fachdidaktik Biologie* (9., völlig überarbeitete Auflage, S. 227–242). Aulis Verlag.
- Gropengießer, H., Harms, U. & Kattmann, U. (Hrsg.). (2013). *Fachdidaktik Biologie* (9., völlig überarbeitete Auflage). Aulis Verlag.
- Grospietsch, F. & Lins, I. (2023). Problembasiertes Lernen mittels Mystery-Methode im Biologieunterricht. *MNU Journal*, 76(3), 194–199.
- Großmann, L. & Krüger, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen. *Erkenntnisweg Biologiedidaktik*, 19, 21–39.
- Großmann, L. & Krüger, D. (2022a). Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen. *SEMINAR – Lehrerbildung und Schule*(1), 91–110.
- Großmann, L. & Krüger, D. (2022b). Students' conceptions as a neglected perspective in trainee teachers' biology lesson plans. In K. Korfiatis & M. Grace (Hrsg.), *Contributions from Biology Education Research. Current research in biology education* (S. 181–193). Springer International Publishing. [https://doi.org/10.1007/978-3-030-89480-1\\_14](https://doi.org/10.1007/978-3-030-89480-1_14)
- Großmann, L. & Krüger, D. (2022c). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1). <https://doi.org/10.1007/s40573-022-00141-w>
- Großmann, L. & Krüger, D. (2023a). Assessing the quality of science teachers' lesson plans: Validation and application of a novel instrument. *Science Education*, 108(1), 153–189. <https://doi.org/10.1002/sc.21832>
- Großmann, L. & Krüger, D. (2023b). Identifying Performance Levels of Enacted Pedagogical Content Knowledge in Trainee Biology Teachers' Lesson Plans. In G. S. Carvalho, A. S. Afonso & Z. Anastácio (Hrsg.), *Contributions from Science Education Research. Fostering Scientific Citizenship in an Uncertain World: Selected Papers from the ESERA 2021 Conference* (S. 95–116). Springer. [https://doi.org/10.1007/978-3-031-32225-9\\_7](https://doi.org/10.1007/978-3-031-32225-9_7)

- Großmann, L., Lotz, A., Mulke, S. & Krüger, D. (2022). Probleme beim Planen?! Acht Ratschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie. *MNU Journal*, 75(5), 368–376.
- Grossmann, P. (1990). *The making of a teacher: Teacher knowledge & teacher education*. Teachers College Press.
- Großschedl, J., Harms, U., Kleickmann, T. & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: structure and learning opportunities. *Journal of Science Teacher Education*, 26(3), 291–318. <https://doi.org/10.1007/s10972-015-9423-6>
- Großschedl, J., Mahler, D., Kleickmann, T. & Harms, U. (2014). Content-Related Knowledge of Biology Teachers from Secondary Schools: Structure and learning opportunities. *International Journal of Science Education*, 36(14), 2335–2366. <https://doi.org/10.1080/09500693.2014.923949>
- Großschedl, J., Welter, V. & Harms, U. (2019). A New Instrument for Measuring Pre-Service Biology Teachers' Pedagogical Content Knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439 (38 Seiten). <https://doi.org/10.1002/tea.21482>
- Gunckel, K. L. (2011). Mediators of a Preservice Teacher's Use of the Inquiry-Application Instructional Model. *Journal of Science Teacher Education*, 22(1), 79–100. <https://doi.org/10.1007/s10972-010-9223-y>
- Gyllenpalm, J. & Wickman, P.-O. (2011). The uses of the term hypothesis and the inquiry emphasis conflation in science teacher education. *International Journal of Science Education*, 33(14), 1993–2015. <https://doi.org/10.1080/09500693.2010.538938>
- Haas, A. (1998). *Unterrichtsplanung im Alltag: Eine empirische Untersuchung zum Planungs Handeln von Hauptschul-, Realschul- und Gymnasiallehrern*. Zugl.: Weingarten, Pädag. Hochsch., Diss., 1998. *Theorie und Forschung: Bd. 540*. Roderer.
- Halim, L. & Meerah, S. M. M. (2002). Science Trainee Teachers' Pedagogical Content Knowledge and its Influence on Physics Teaching. *Research in Science & Technological Education*, 20(2), 215–225. <https://doi.org/10.1080/0263514022000030462>
- Hammann, M. & Asshoff, R. (2014). *Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten* (1. Aufl.). Klett Kallmeyer. <https://www.friedrich-verlag.de/shop/schuelervorstellungen-im-biologieunterricht-14908>
- Harman, G. H. (1965). The inference to the best explanation. *The Philosophical Review*, 74(1), 88. <https://doi.org/10.2307/2183532>
- Hashweh, M. Z. (2005). Teacher pedagogical constructions: a reconfiguration of pedagogical content knowledge. *Teachers and Teaching*, 11(3), 273–292. <https://doi.org/10.1080/13450600500105502>
- Hattie, J. (2023). *Visible learning, the sequel: A synthesis of over 2,100 meta-analyses relating to achievement* (1. Aufl.). Routledge. <https://www.taylorfrancis.com/books/9781003380542> <https://doi.org/10.4324/9781003380542>
- Heimann, P. (1962). Didaktik als Theorie und Lehre. *Die deutsche Schule*, 54(9), 407–427.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (7., korrigierte Auflage). Klett Kallmeyer. <https://www.friedrich-verlag.de/shop/unterrichtsqualitaet-und-professionalisierung-31684>
- Hodson, D. (2014). Learning science, learning about science, doing science: different goals demand different learning methods. *International Journal of Science Education*, 36(15), 2534–2553. <https://doi.org/10.1080/09500693.2014.899722>

- Howe, E. R. (2006). Exemplary teacher induction: an international review. *Educational Philosophy and Theory*, 38(3), 287–297. <https://doi.org/10.1111/j.1469-5812.2006.00195.x>
- Hübner, P. & Werle, M. (1997). Arbeitszeit und Arbeitsbelastung Berliner Lehrerinnen und Lehrer. In S. Buchen (Hrsg.), *Jahrbuch der Lehrerforschung* (S. 203–226). Juventa.
- Irmer, M., Traub, D., Böhm, M., Förtsch, C. & Neuhaus, B. J. (2023). Using Video-Based Simulations to Foster pPCK/ePCK—New Thoughts on the Refined Consensus Model of PCK. *Education Sciences*, 13(3), 261. <https://doi.org/10.3390/educsci13030261>
- Jackson, K., Garrison, A., Wilson, J., Gibbons, L. & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education*, 44(4), 646–682. <https://doi.org/10.5951/jresmetheduc.44.4.0646>
- Jacobs, C. L., Martin, S. N. & Otieno, T. C. (2008). A science lesson plan analysis instrument for formative and summative program evaluation of a teacher education program. *Science Education*, 92(6), 1096–1126. <https://doi.org/10.1002/sce.20277>
- Jin, H., Shin, H., Johnson, M. E., Kim, J. & Anderson, C. W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching*, 52(9), 1269–1295. <https://doi.org/10.1002/tea.21243>
- John, P. D. (2006). Lesson planning and the student teacher: re-thinking the dominant model. *Journal of Curriculum Studies*, 38(4), 483–498. <https://doi.org/10.1080/00220270500363620>
- Johnson, T. R. & Kreams, J. F. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive Science*, 25(6), 903–939. [https://doi.org/10.1016/S0364-0213\(01\)00059-3](https://doi.org/10.1016/S0364-0213(01)00059-3)
- Jong, O. de, van Driel, J. H. & Verloop, N. (2005). Preservice teachers' pedagogical content knowledge of using particle models in teaching chemistry. *Journal of Research in Science Teaching*, 42(8), 947–964. <https://doi.org/10.1002/tea.20078>
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Jüttner, M. & Neuhaus, B. J. (2013). Das Professionswissen von Biologielehrkräften: Ein Vergleich zwischen Biologielehrkräften, Biologen und Pädagogen. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 31–49. [http://archiv.ipn.uni-kiel.de/zfdn/pdf/19\\_Juettner.pdf](http://archiv.ipn.uni-kiel.de/zfdn/pdf/19_Juettner.pdf)
- Kademian, S. M. & Davis, E. A. (2018). Supporting beginning teacher planning of investigation-based science discussions. *Journal of Science Teacher Education*, 29(8), 712–740. <https://doi.org/10.1080/1046560X.2018.1504266>
- Kaiser, H. F. & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34(1), 111–117. <https://doi.org/10.1177/001316447403400115>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kang, H. (2017). Preservice teachers' learning to plan intellectually challenging tasks. *Journal of Teacher Education*, 68(1), 55–68. <https://doi.org/10.1177/0022487116676313>
- Kang, H., Windschitl, M., Stroupe, D. & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific

- thinking. *Journal of Research in Science Teaching*, 53(9), 1316–1340.  
<https://doi.org/10.1002/tea.21329>
- Kantreiter, J. (2022). *Unterrichtsplanung unter die Lupe genommen: Eine empirische Studie zum Zusammenhang der intrinsischen motivationalen Orientierungen von Grundschullehrkräften und dem beigemessenen Wert der Planungsqualität im Sachunterricht. Empirische Forschung im Elementar- und Primarbereich: Bd. 10.* Julius Klinkhardt.
- Käpylä, M., Heikkinen, J.-P. & Asunta, T. (2009). Influence of Content Knowledge on Pedagogical Content Knowledge: The case of teaching photosynthesis and plant growth. *International Journal of Science Education*, 31(10), 1395–1415.  
<https://doi.org/10.1080/09500690802082168>
- Karlström, M. & Hamza, K. (2021). How do we teach planning to pre-service teachers – A tentative model. *Journal of Science Teacher Education*, 1–22.  
<https://doi.org/10.1080/1046560X.2021.1875163>
- Kärner, T., Bonnes, C. & Schölzel, C. (2019). Bewertungstransparenz im Referendariat. *Zeitschrift für Pädagogik*, 65(3), 378–400.
- Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht.* Aulis Verlag.
- Kattmann, U. (Hrsg.). (2017). *Biologie unterrichten mit Alltagsvorstellungen: Didaktische Rekonstruktionen in Unterrichtseinheiten* (1. Aufl.). Klett Kallmeyer. <https://www.friedrich-verlag.de/shop/biologie-unterrichten-mit-alltagsvorstellungen-31068>
- Kaya, Z. & Kaya, O. N. (2023). Gathering Rich Data on Preservice Science Teachers' Pedagogical Content Knowledge Through Their Lesson Plans. *Journal of Teacher Education*, 74(1), 10–22. <https://doi.org/10.1177/00224871221105801>
- Keller, M. M., Neumann, K. & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*(5), 586–614.
- Kind, V. & Chan, K. K. H. (2019). Resolving the amalgam: connecting pedagogical content knowledge, content knowledge and pedagogical knowledge. *International Journal of Science Education*, 41(7), 964-978. <https://doi.org/10.1080/09500693.2019.1584931>
- Kircher, E. (2015). Planung und Analyse von Physikunterricht. In E. Kircher, R. Girwitz & P. Häußler (Hrsg.), *Springer-Lehrbuch. Physikdidaktik* (S. 295–319). Springer.  
[https://doi.org/10.1007/978-3-642-41745-0\\_8](https://doi.org/10.1007/978-3-642-41745-0_8)
- Klafki, W. (1958). Didaktische Analyse als Kern der Unterrichtsvorbereitung. *Die deutsche Schule*, 50(10), 450–471.
- Klafki, W. (1963). Das Problem der Didaktik. In Präsidium des Pädagogischen Hochschultages (Hrsg.), *Das Problem der Didaktik: Bericht über den fünften Deutschen Pädagogischen Hochschultag vom 1. bis 5. Oktober in Trier* (S. 19–62). Julius Beltz.
- Klafki, W. (1996). Zur Unterrichtsplanung im Sinne kritisch-konstruktiver Didaktik. In W. Klafki (Hrsg.), *Neue Neue Studien zur Bildungstheorie und Didaktik: Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik.* (5. Aufl., S. 251-284). Beltz.
- Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik: Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* (6. Aufl.). Beltz.
- Klieme, E., Maag-Merki, K. & Hartig, J. (2007). Kompetenzbegriff und Bedeutung von Kompetenzen im Bildungswesen. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 5–15). Bundesministerium für Bildung und Forschung.

- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS - Impulse für Schule und Unterricht: Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–57). Bundesministerium für Bildung und Forschung.
- KMK. (2005). *Bildungsstandards im Fach Biologie für den mittleren Schulabschluss. Beschluss vom 16.12.2004. Beschlüsse der Kultusministerkonferenz*. Luchterhand.
- KMK. (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife*. Carl Link.
- Koberstein-Schwarz, M. & Meisert, A. (2022). Pedagogical content knowledge in material-based lesson planning of preservice biology teachers. *Teaching and Teacher Education, 116*, 103745. <https://doi.org/10.1016/j.tate.2022.103745>
- König, J., Blömeke, S., Jentsch, A., Schlesinger, L., née Nehls, C. F., Musekamp, F. & Kaiser, G. (2021). The links between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom. *Educational Studies in Mathematics, 107*(1), 189–212. <https://doi.org/10.1007/s10649-020-10021-0>
- König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I. & Glutsch, N. (2020a). Planning Competence of Pre-Service German Language Teachers (PlanvoLL-D): Conceptualization, Measurement, and Validation. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepfer & C. Lautenbach (Hrsg.), *Student Learning in German Higher Education: Innovative Measurement Approaches and Research Results* (S. 53–74). Springer. [https://doi.org/10.1007/978-3-658-27886-1\\_4](https://doi.org/10.1007/978-3-658-27886-1_4)
- König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I. & Glutsch, N. (2020b). Pre-service teachers' generic and subject-specific lesson-planning skills: On learning adaptive teaching during initial teacher education. *European Journal of Teacher Education, 43*(2), 131–150. <https://doi.org/10.1080/02619768.2019.1679115>
- König, J., Bremerich-Vos, A., Buchholtz, C. & Glutsch, N. (2020). General pedagogical knowledge, pedagogical adaptivity in written lesson plans, and instructional practice among preservice teachers. *Journal of Curriculum Studies, 52*(6), 800–822. <https://doi.org/10.1080/00220272.2020.1752804>
- König, J., Bremerich-Vos, A., Buchholtz, C., Lammerding, S., Strauß, S., Fladung, I. & Schleiffer, C. (2017). Die Bedeutung des Professionswissens von Referendarinnen und Referendaren mit Fach Deutsch für ihre Planungskompetenz (PlanvoLL-D). In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 121–133). Julius Klinkhardt.
- König, J., Buchholtz, C. & Dohmen, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft, 18*(2), 375–404. <https://doi.org/10.1007/s11618-015-0625-7>
- König, J., Cammann, F., Bremerich-Vos, A. & Buchholtz, C. (2022). Unterrichtsplanungskompetenz von (angehenden) Deutschlehrkräften der Sekundarstufe: Testkonstruktion und Validierung. *Zeitschrift für Erziehungswissenschaft, 25*(4), 869–894. <https://doi.org/10.1007/s11618-022-01113-z>
- König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M., Strauß, S., Bremerich-Vos, A., Fladung, I. & Kaiser, G. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? *Zeitschrift für Erziehungswissenschaft, 21*(3), 1–38. <https://doi.org/10.1007/s11618-017-0765-z>

- König, J., Krepf, M., Bremerich-Vos, A. & Buchholtz, C. (2021). Meeting cognitive demands of lesson planning: introducing the CODE-PLAN Model to describe and analyze teachers' planning competence. *The Teacher Educator*, 56(4), 466–487. <https://doi.org/10.1080/08878730.2021.1938324>
- König, J. & Rothland, M. (2022). Stichwort: Unterrichtsplanungskompetenz. *Zeitschrift für Erziehungswissenschaft*, 25, 771–813. <https://doi.org/10.1007/s11618-022-01107-x>
- Körkkö, M., Kyrö-Ämmälä, O. & Turunen, T. (2016). Professional development through reflection in teacher education. *Teaching and Teacher Education*, 55, 198–206. <https://doi.org/10.1016/j.tate.2016.01.014>
- Kotzebue, L. von (2022). Beliefs, self-reported or performance-assessed TPACK: What can predict the quality of technology-enhanced biology lesson plans? *Journal of Science Education and Technology*, 31(5), 570–582. <https://doi.org/10.1007/s10956-022-09974-z>
- Krebs, R., Rothstein, B. & Roelle, J. (2022). Rubrics enhance accuracy and reduce cognitive load in self-assessment. *Metacognition and Learning*(17), 627–650. <https://doi.org/10.1007/s11409-022-09302-1>
- Krell, M. (2020). Vorstellung und Kompetenz: Zusammenführung zweier zentraler Konzepte in der naturwissenschaftsdidaktischen Forschung. In B. Reinisch, K. Helbig & D. Krüger (Hrsg.), *Biologiedidaktische Vorstellungsforschung: Zukunftsweisende Praxis* (S. 69–82). Springer.
- Krell, M. & Krüger, D. (2022). Erkenntnisgewinnung im Biologieunterricht: Welches Methodenwissen soll vermittelt werden? *MNU Journal*(5), 376–382.
- Krepf, M. & König, J. (2022a). Structuring lessons as an aspect of pre-service teachers' planning competence: A scaling-up analysis. *Zeitschrift für Erziehungswissenschaft*, 25(4), 917–946. <https://doi.org/10.1007/s11618-022-01125-9>
- Krepf, M. & König, J. (2022b). Structuring the lesson: an empirical investigation of pre-service teacher decision-making during the planning of a demonstration lesson. *Journal of Education for Teaching*, 1–16. <https://doi.org/10.1080/02607476.2022.2151877>
- Krepf, M., Plöger, W., Scholl, D. & Seifert, A. (2018). Pedagogical content knowledge of experts and novices-what knowledge do they activate when analyzing science lessons? *Journal of Research in Science Teaching*, 55(1), 44–67. <https://doi.org/10.1002/tea.21410>
- Krüger, D. (2007). Die Conceptual Change-Theorie. In D. Krüger & H. Vogt (Hrsg.), *Springer-Lehrbuch. Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 81–92). Springer.
- Krüger, D., Hartmann, S., Nordmeier, V. & Upmeyer zu Belzen, A. (2020). Measuring scientific reasoning competencies. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper & C. Lautenbach (Hrsg.), *Student learning in German higher education* (S. 261–280). Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-27886-1\\_13](https://doi.org/10.1007/978-3-658-27886-1_13)
- Krüger, D. & Riemeier, T. (2014). Die qualitative Inhaltsanalyse – eine Methode zur Auswertung von Interviews. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Lehrbuch. Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 133–145). Springer. [https://doi.org/10.1007/978-3-642-37827-0\\_11](https://doi.org/10.1007/978-3-642-37827-0_11)
- Krüger, D. & Upmeyer zu Belzen, A. (2021). Kompetenzmodell der Modellierkompetenz – Die Rolle abduktiven Schließens beim Modellieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 27(1), 127–137. <https://doi.org/10.1007/s40573-021-00129-y>
- Kruse, C. (2022). Die Beurteilung angehender Lehrer\*innen im Referendariat. Relevanz, Kontext und erste Explorationen. *journal für lehrerInnenbildung*, 22(1), 34–45. <https://doi.org/10.35468/jlb-01-2022-03>



- Kulgemeyer, C. & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, 55(10), 1393–1418. <https://doi.org/10.1002/tea.21457>
- Kunter, M. & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H.-G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Waxmann-E-Books Didaktik, Schule und Unterricht: Band 1. Bedingungen und Effekte guten Unterrichts* (S. 9–31). Waxmann.
- Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften: Methodische Herausforderungen. *Unterrichtswissenschaft*, 38(1), 68–86.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *The journal of educational psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Kunz, H. & Uhl, S. (2022). Der Vorbereitungsdienst: Desiderate und Perspektiven der Forschung. *Seminar*, 28(2), 22–35. <https://elibrary.utb.de/doi/epdf/10.3278/SEM2202W004>;
- Künzli, R. (1996). Schul- und Erfahrungsnahe in der Lehrerbildung. Elemente des Konzeptes für die Ausbildung der Sekundar- und Reallehrer und -lehrerinnen am Didaktikum in Aarau. *Beiträge zur Lehrerbildung*, 14, 291–301. <https://doi.org/10.25656/01:13337>
- Küth, S., Scholl, D. & Schüle, C. (2021). Entscheidungstendenzen als psychoemotionale Einflussfaktoren auf das selbsteingeschätzte unterrichtliche Planungsverhalten angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 24(5), 1165–1182. <https://doi.org/10.1007/s11618-021-01029-0>
- Landis, J. R. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174.
- Leat, D. (Hrsg.). (1998). *Thinking through geography*. Chris Kington. <https://permalink.obvsg.at/AC15157182>
- Leijen, Ä., Malva, L., Pedaste, M. & Mikser, R. (2022). What constitutes teachers' general pedagogical knowledge and how it can be assessed: A literature review. *Teachers and Teaching*, 28(2), 206–225. <https://doi.org/10.1080/13540602.2022.2062710>
- Lenke, G., Thillmann, H., Wirth, J., Dicke, T. & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProWiN-Tests. *Zeitschrift für Erziehungswissenschaft*, 18(2), 225–245. <https://doi.org/10.1007/s11618-015-0627-5>
- Lin, J.-W. (2016). Do Skilled Elementary Teachers Hold Scientific Conceptions and Can They Accurately Predict the Type and Source of Students' Preconceptions of Electric Circuits? *International Journal of Science and Mathematics Education*, 14, 287–307. <https://doi.org/10.1007/s10763-015-9635-4>
- Lipowsky, F., Drollinger-Vetter, Klieme, E., Pauli, C. & Reusser, K. (2018). Generische und fachdidaktische Dimensionen von Unterrichtsqualität - zwei Seiten einer Medaille? In M. Martens, K. Rabenstein, K. Bräu, M. Fetzer, H. Gresch, I. Hardy & C. Schelle (Hrsg.), *Konstruktionen von Fachlichkeit: Ansätze, Erträge und Diskussionen in der empirischen Unterrichtsforschung* (S. 183–202). Julius Klinkhardt.
- Livingston, C. & Borko, H. (1989). Expert-novice differences in teaching: A cognitive analysis and implications for teacher education. *Journal of Teacher Education*, 40(4), 36–42.
- Lübeck, M. (2020). *'Basiskonzepte' der Erkenntnisgewinnung im Biologieunterricht: Ein Nachschlagewerk mit Aufgabenbeispielen* (1. Auflage). Waxmann.

- Magnusson, S. J., Borko, H. & Krajcik, J. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Hrsg.), *Examining pedagogical content knowledge* (S. 95–132). Kluwer Academic Publishers.
- Mahler, D. & Arnold, J. (2022). MaSter-Bio – Messinstrument für das akademische Selbstkonzept zum technologiebezogenen Professionswissen von angehenden Biologielehrpersonen. *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1), 37–52. <https://doi.org/10.1007/s40573-022-00137-6>
- Mahler, D., Großschedl, J. & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, 39(2), 213–237. <https://doi.org/10.1080/09500693.2016.1276641>
- Maier, U., Bohl, T., Drücke-Noe, C., Hoppe, H., Kleinknecht, M. P. & Metz, K. (2014). Das kognitive Anforderungsniveau von Aufgaben analysieren und modifizieren können: Eine wichtige Fähigkeit von Lehrkräften bei der Planung eines kompetenzorientierten Unterrichts. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32(2), 340–358. [http://fox.leuphana.de/portal/de/publications/das-kognitive-anforderungsniveau-von-aufgaben-analysieren-und-modifizieren-koennen\(4c6b035f-3b67-4fe8-bb94-28692e82e98a\).html](http://fox.leuphana.de/portal/de/publications/das-kognitive-anforderungsniveau-von-aufgaben-analysieren-und-modifizieren-koennen(4c6b035f-3b67-4fe8-bb94-28692e82e98a).html)
- Mavhunga, E. (2020). Revealing the structural complexity of component interactions of topic-specific PCK when planning to teach. *Research in Science Education*, 50(3), 965–986. <https://doi.org/10.1007/s11165-018-9719-6>
- Mavhunga, E. & Ndlovu, B. P. (2023). Defining Science Content Knowledge for Teaching as a Base for Teacher Development: A Case for Organic Chemistry. *African Journal of Research in Mathematics, Science and Technology Education*, 1–15. <https://doi.org/10.1080/18117295.2023.2247711>
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Springer-Lehrbuch. Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 177–186). Springer.
- Mayer, J. & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Lehrbuch. Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 19–29). Springer. [https://doi.org/10.1007/978-3-642-37827-0\\_2](https://doi.org/10.1007/978-3-642-37827-0_2)
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12., überarb. Aufl.). Beltz Pädagogik. Beltz. [http://content-select.com/index.php?id=bib\\_view&ean=9783407293930](http://content-select.com/index.php?id=bib_view&ean=9783407293930)
- Mayring, P. (2019). Qualitative Inhaltsanalyse – Abgrenzungen, Spielarten, Weiterentwicklungen. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3). <https://doi.org/10.17169/fqs-20.3.3343>
- Meinhardt, C., Rabe, T. & Krey, O. (2018). Formulierung eines evidenzbasierten Validitätsarguments am Beispiel der Erfassung physikdidaktischer Selbstwirksamkeitserwartungen mit einem neu entwickelten Instrument. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 131–150. <https://doi.org/10.1007/s40573-018-0079-6>
- Meisert, A. (2016). Eine Langzeitplanung anlegen. In C. Schaper (Hrsg.), *Werkzeugkoffer Pädagogisches: Ein Handbuch für den Start in den Lehrberuf* (S. 85–103). Vandenhoeck & Ruprecht.

- Meisert, A. (2021). Wie kann Biologieunterricht geplant werden? In U. Spörhase (Hrsg.), *Biologie-Didaktik: Praxishandbuch für die Sekundarstufen I und II* (8. Aufl., S. 241–272). Cornelsen.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2. Aufl.). Sage.
- Ministère de l'Éducation Nationale. (2013). *Arrêté du 1er juillet 2013 relatif au référentiel des compétences professionnelles des métiers du professorat et de l'éducation*.  
<https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027721614/>
- Moodley, K. & Gaigher, E. (2019). Teaching Electric Circuits: Teachers' Perceptions and Learners' Misconceptions. *Research in Science Education*, 49(1), 73–89.  
<https://doi.org/10.1007/s11165-017-9615-5>
- Morine-Dershimer, G. (1979). *Teacher plan and classroom reality: The South Bay Study, part IV*. Research series No. 60. Institute for Research on Teaching College of Education Michigan State University.
- Morrison, J. A. & Lederman, N. G. (2003). Science teachers' diagnosis and understanding of students' preconceptions. *Science Education*, 87(6), 849–867.  
<https://doi.org/10.1002/sce.10092>
- Moskal, B. M. (2000). Scoring Rubrics: What, When and How? *Practical Assessment, Research, and Evaluation*, 7(10), 1–6. <https://doi.org/10.7275/A5VQ-7Q66>
- Moskal, B. M. & Leydens, J. A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research, and Evaluation*, 7(7), 1–5. <https://doi.org/10.7275/q7rm-gg74>
- Mülhausen, J. & Pütz, N. (Hrsg.). (2013). *Mysteries: 9 rätselhafte Fälle für den Biologieunterricht*. Aulis-Verlag.
- Müller, S.-L. & Mahler, D. (in Druck). Charakteristika der Biologie – Bio-NOS als ein Impuls für die Lehrkräftebildung. In B. Reinisch, D. Krüger & D. Mahler (Hrsg.), *Biologiedidaktische Nature-of-Science-Forschung: Zukunftsweisende Praxis*. Springer.
- Mundfrom, D. J., Shaw, D. G. & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159–168.  
[https://doi.org/10.1207/s15327574ijt0502\\_4](https://doi.org/10.1207/s15327574ijt0502_4)
- Munthe, E. & Conway, P. F. (2017). Evolution of research on teachers' planning: implications for teacher education. In J. Hunsu & D. J. Clandinin (Hrsg.), *Evolution of research on teachers' planning: Implications for teacher education* (S. 836–849). SAGE Publications Ltd.
- Mutton, T., Hagger, H. & Burn, K. (2011). Learning to plan, planning to learn: the developing expertise of beginning teachers. *Teachers and Teaching*, 17(4), 399–416.  
<https://doi.org/10.1080/13540602.2011.580516>
- National Council for Teacher Education. (2016). *National guidelines for the primary and lower secondary teacher education programme for years 1-7*. [https://www.uhr.no/\\_f/p-1/i9667e583-aa3b-4f25-a8fe-64af8b199072/national\\_guidelines\\_for\\_the\\_primary\\_and\\_lower\\_secondary\\_teacher\\_education\\_programme\\_for\\_years\\_1\\_7.pdf](https://www.uhr.no/_f/p-1/i9667e583-aa3b-4f25-a8fe-64af8b199072/national_guidelines_for_the_primary_and_lower_secondary_teacher_education_programme_for_years_1_7.pdf)
- Nauck, J. (1996). Zur Planungs- und Reflexionskompetenz von Studierenden. Inhaltsanalyse didaktischer Akten zum Allgemeinen Schulpraktikum. In K. H. Sander (Hrsg.),

- Schulpraktische Studien. Erfahrungen mit dem Braunschweiger Modell der Lehrerbildung* (S. 181–210). Schmidt.
- Navy, S. L., Luft, J. A. & Msimanga, A. (2022). The learning opportunities of newly hired teachers of science. In J. A. Luft & M. G. Jones (Hrsg.), *Handbook of Research on Science Teacher Education* (S. 245–256). Routledge.
- Ndihokubwayo, K., Byukusenge, C., Byusa, E., Habiyaremye, H. T., Mboniyirivuze, A. & Mukagihana, J. (2022). Lesson plan analysis protocol (LPAP): A useful tool for researchers and educational evaluators. *Heliyon*, 8(1), 1-7.  
<https://doi.org/10.1016/j.heliyon.2022.e08730>
- Nerdel, C. (2017). Planung von naturwissenschaftlichem Unterricht. In C. Nerdel (Hrsg.), *Grundlagen der Naturwissenschaftsdidaktik* (S. 213–245). Springer.
- Neuhaus, B. J. (2021). Unterrichtsqualität aus der Perspektive der Biologiedidaktik. *Unterrichtswissenschaft*, 49(2), 273–283. <https://doi.org/10.1007/s42010-021-00114-x>
- Neumann, K., Kind, V. & Harms, U. (2019). Probing the amalgam: the relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, 41(7), 847-861.  
<https://doi.org/10.1080/09500693.2018.1497217>
- Neuweg, G. H. (2018). *Können als Magd des Wissens? Skeptische Anmerkungen zur Lehrerkompetenzforschung* [Youtube-Video].  
[https://www.youtube.com/watch?v=B4U3Go9Q\\_4s](https://www.youtube.com/watch?v=B4U3Go9Q_4s)
- Neuweg, G. H. (2022). *Lehrerbildung: Zwölf Denkfiguren im Spannungsfeld von Wissen und Können*. Waxmann.
- Norman, G. R. & Streiner, D. L. (2014). *Biostatistics: The bare essentials* (Fourth edition). People's Medical Publishing House - USA.
- Norville, K. & Park, S. (2021). The impact of the cooperating teacher on master of arts in teaching preservice science teachers' pedagogical content knowledge. *Journal of Science Teacher Education*, 32(4), 444–468. <https://doi.org/10.1080/1046560X.2020.1850614>
- Oehlschläger, H.-J. (1978). *Zur Praxisrelevanz pädagogischer Literatur: Strukturen und Trends der Literaturrezeption praktizierender Lehrer* (1. Aufl.). Klett-Cotta.  
<https://permalink.obvsg.at/AC01783421>
- Oh, P. S. (2019). Features of Modeling-Based Abductive Reasoning as a Disciplinary Practice of Inquiry in Earth Science. *Science & Education*, 28(6-7), 731–757.  
<https://doi.org/10.1007/s11191-019-00058-w>
- Oser, F. K. (1997). Standards in der Lehrerbildung. Teil 1: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerbildung*, 15.  
<https://doi.org/10.25656/01:13343> (Beiträge zur Lehrerbildung 15 (1997) 1, S. 26-37).
- Oser, F. K. & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Hrsg.), *Handbook of research on teaching* (S. 1031–1065). American Educational Research Association.
- Otero, V. K. & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. *Journal of Research in Science Teaching*, 45(4), 497–523.  
<https://doi.org/10.1002/tea.20229>
- Ozogul, G., Olina, Z. & Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development*, 56(2), 181–201. <https://doi.org/10.1007/s11423-006-9012-7>

- Pacaci, C., Ustun, U. & Ozdemir, O. F. (2023). Effectiveness of conceptual change strategies in science education: A meta-analysis. *Journal of Research in Science Teaching*, 1–63. 10.1002/tea.21887
- Panadero, E., Andrade, H. & Brookhart, S. M. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher*, 45(1), 13–31. <https://doi.org/10.1007/s13384-018-0258-y>
- Panadero, E. & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: a review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E. & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30, 100329. <https://doi.org/10.1016/j.edurev.2020.100329>
- Park, S. & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941. <https://doi.org/10.1002/tea.21022>
- Park, S., Jang, J.-Y., Chen, Y.-C. & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching? Evidence from an empirical study. *Research in Science Education*, 41(2), 245–260. <https://doi.org/10.1007/s11165-009-9163-8>
- Park, S. & Oliver, J. S. (2008). National Board Certification (NBC) as a catalyst for teachers' learning about teaching: the effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45(7), 812–834. <https://doi.org/10.1002/tea.20234>
- Park, S. & Suh, J. K. (2019). The PCK map approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 185–197). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_8](https://doi.org/10.1007/978-981-13-5898-2_8)
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2. ed. [Nachdr.]). Sage.
- Pearsall, N. R., Skipper, J. E. J. & Mintzes, J. J. (1997). Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. *Science Education*, 81(2), 193–215. [https://doi.org/10.1002/\(SICI\)1098-237X\(199704\)81:2%3C193::AID-SCE5%3E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-237X(199704)81:2%3C193::AID-SCE5%3E3.0.CO;2-A)
- Peter, E. (2007). Was brauchen Kressesamen zum Keimen? Experimente als Schiedsrichter. In R. Duit, H. Gropengießer & L. Stäudel (Hrsg.), *Naturwissenschaftliches Arbeiten: Unterricht und Material 5-10* (S. 64–69). Friedrich.
- Peterson, P. L., Marx, R. W. & Clark, C. M. (1978). Teacher planning, teacher behavior, and student achievement. *American Educational Research Journal*, 15(3), 417–432.
- Petitmengin, C. (2006). Describing one's subjective experience in the second person: An interview method for the science of consciousness. *Phenomenology and the Cognitive Sciences*, 5(3-4), 229–269. <https://doi.org/10.1007/s11097-006-9022-2>
- Pohlmann, M. (2019). *Förderung ethischer Bewertungskompetenz: der Einfluss ausgewählter Lerngelegenheiten auf die inhaltliche Ausdifferenzierung und die Kohärenz der Komponenten des fachdidaktischen Wissens von Biologielehrkräften* [Dissertation]. Universität Oldenburg. <https://nbn-resolving.org/urn:nbn:de:gbv:715-oops-41077>; <https://d-nb.info/1182534066/34>
- Popper, K. R. (2005). *Gesammelte Werke: Bd. 3. Logik der Forschung* (H. Keuth, Hg.). Mohr Siebeck.

- Post, M. (2018). *Die Bergmannsche Regel in der Ökologie auf dem didaktischen Prüfstand* [MNU Journal]. [https://www.mnu.de/images/didaktischerpruefstand/5BergmannscheRegel/BergmannscheRegel\\_finaleLangfassung\\_2018-12-24.pdf](https://www.mnu.de/images/didaktischerpruefstand/5BergmannscheRegel/BergmannscheRegel_finaleLangfassung_2018-12-24.pdf)
- Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, 48(3), 409–446. <https://doi.org/10.1007/s42010-020-00082-8>
- Reddy, Y. M. & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. <https://doi.org/10.1080/02602930902862859>
- Reinisch, B. & Fricke, K. (2022). Broadening a nature of science conceptualization: Using school biology textbooks to differentiate the family resemblance approach. *Science Education*, 106(6), 1375–1407. <https://doi.org/10.1002/sce.21729>
- Reynolds, R., Macqueen, S. & Ferguson-Patrick, K. (2020). Teacher educators and pre-service teachers work towards knowledge integration. In T. Lehmann (Hrsg.), *International perspectives on knowledge integration: Theory, Research, and good practice in pre-service teacher and higher education* (S. 389–401). Brill.
- Reynolds, W. M. & Park, S. (2021). Examining the relationship between the educative teacher performance assessment and preservice teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, 58, 721–748. <https://doi.org/10.1002/tea.21676>
- Riemeier, T. (2007). Moderater Konstruktivismus. In D. Krüger & H. Vogt (Hrsg.), *Springer-Lehrbuch. Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 69–79). Springer.
- Riese, J., Vogelsang, C., Schröder, J., Borowski, A., Kulgemeyer, C., Reinhold, P. & Schecker, H. (2022). Entwicklung von Unterrichtsplanungsfähigkeit im Fach Physik: Welchen Einfluss hat Professionswissen? *Zeitschrift für Erziehungswissenschaft*(25), 843–867. <https://doi.org/10.1007/s11618-022-01112-0>
- Rios, J. & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108–116. <https://doi.org/10.7334/psicothema2013.260>
- Rohrman, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*(9), 222–245.
- Rohrman, B. (2007). *Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data*. Project Report University of Melbourne/Australia. <http://www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf>
- Rollnick, M. (2017). Learning about semi conductors for teaching - the role played by content knowledge in pedagogical content knowledge (PCK) development. *Research in Science Education*, 47(4), 833–868. <https://doi.org/10.1007/s11165-016-9530-1>
- Roloff, J., Klusmann, U., Lüdtke, O. & Trautwein, U. (2020). The predictive validity of teachers' personality, cognitive and academic abilities at the end of high school on instructional quality in Germany: a longitudinal study. *AERA Open*, 6(1), 1-17. <https://doi.org/10.1177/2332858419897884>
- Rönnebeck, S., Bernholt, S. & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>

- Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schulle, K. & Wickler, N. I. Z. (2011). Videobased Lesson Analysis: Effective Science PD for Teacher and Student Learning. *Journal of Research in Science Teaching*, 48(2), 117-148. <https://doi.org/10.1002/tea.20408>
- Rothland, M. (2021). Anmerkungen zur Modellierung und Operationalisierung (allgemeindidaktischer) Unterrichtsplanungskompetenz. *Unterrichtswissenschaft*, 50, 347-372. <https://doi.org/10.1007/s42010-021-00111-0>
- Rothland, M., Cramer, C. & Terhart, E. (2018). Forschung zum Lehrerberuf und zur Lehrerbildung. In R. Tippelt & B. Schmidt-Hertha (Hrsg.), *Handbuch Bildungsforschung* (S. 1011–1034). Springer.
- Ruys, I., van Keer, H. & Aelterman, A. (2012). Examining pre-service teacher competence in lesson planning pertaining to collaborative learning. *Journal of Curriculum Studies*, 44(3), 349–379. <https://doi.org/10.1080/00220272.2012.675355>
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159–179. <https://doi.org/10.1080/02602930801956059>
- Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, 67(3), 273–288. <https://doi.org/10.1007/s10734-013-9649-1>
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N. & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020–1049. <https://doi.org/10.3102/0002831213477680>
- Sandmann, A. (2014). Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Lehrbuch. Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179–188). Springer. [https://doi.org/10.1007/978-3-642-37827-0\\_15](https://doi.org/10.1007/978-3-642-37827-0_15)
- Schecker, H. & Parchmann, I [I.] (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Schiering, D., Sorge, S., Keller, M. M. & Neumann, K. (2023). A proficiency model for pre-service physics teachers' pedagogical content knowledge (PCK): What constitutes high-level PCK? *Journal of Research in Science Teaching*, 60(1), 136–163. <https://doi.org/10.1002/tea.21793>
- Schlesinger, L., Jentsch, A., Kaiser, G., König, J. & Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM*, 50(3), 475–490. <https://doi.org/10.1007/s11858-018-0917-5>
- Schmelzing, S., Driel, H., Jüttner, M., Brandenbusch, S., Sandmann, A. & Neuhaus, J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the cardiovascular system. *International Journal of Science and Mathematics Education*, 11(6), 1369–1390. <https://doi.org/10.1007/s10763-012-9384-6>
- Schmiemann, P. & Lücken, M. (2014). Validität – Misst mein Test, was er soll? In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Lehrbuch. Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 107–118). Springer. [https://doi.org/10.1007/978-3-642-37827-0\\_9](https://doi.org/10.1007/978-3-642-37827-0_9)
- Schnebel, S., Kreis, A. & Musow, S. (2017). Wie schätzen Studierende ihre Planungskompetenz und den Nutzen kooperativer Unterrichtsplanung ein? In A. Kreis & S. Schnebel (Hrsg.), *Lehrerbildung auf dem Prüfstand: 2017, Sonderheft. Peer Coaching in der praxissituierten Ausbildung von Lehrpersonen* (S. 107–128). Verlag Empirische Pädagogik.

- Schneider, R. M. & Plasman, K. (2011). Science teacher learning progressions. *Review of Educational Research*, 81(4), 530–565. <https://doi.org/10.3102/0034654311423382>
- Scholl, D. (2018). *Metatheorie der allgemeinen Didaktik: Ein systemtheoretisch begründeter Vorschlag*. Klinkhardt Forschung. Julius Klinkhardt. <https://elibrary.utb.de/doi/book/10.35468/9783781556539>
- Scholl, D., Küth, S., Flath, M., Lathan, H., Schwarz, B., Wolters, P., Rheinländer, K. & Schüle, C. (2019). Zum Konstrukt der Planungskompetenz in allgemein- und fachdidaktischen Ansätzen. In D. Scholl, S. Wernke & D. Behrens (Hrsg.), *Jahrbuch für Allgemeine Didaktik* (S. 75–94). Schneider Verlag Hohengehren.
- Scholl, D., Küth, S. & Schüle, C. (2022). Interdependentes Entscheiden in der Unterrichtsplanung – Entwicklung eines generischen Rahmenmodells und eines vignettenbasierten Fähigkeitstests. *Zeitschrift für Erziehungswissenschaft*, 25(4), 895–916. <https://doi.org/10.1007/s11618-022-01117-9>
- Scholl, D. & Plöger, W. (2020). Unterricht als Gegenstand der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 21–30). Julius Klinkhardt.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books. <http://www.loc.gov/catdir/enhancements/fy0832/82070855-d.html>
- Schrader, J. & Schöb, S. (2016). Die Planung von Lehr-Lern-Einheiten mit digitalen Medien: Konzepte und Befunde. *Zeitschrift für Weiterbildungsforschung*, 39(3), 331–347. <https://doi.org/10.1007/s40955-016-0078-5>
- Schreiber, L. M., Paul, G. D. & Shibley, L. R. (2012). The development and test of the public speaking competence rubric. *Communication Education*, 61(3), 205–233. <https://doi.org/10.1080/03634523.2012.670709>
- Schreiber, N. & Gut, C. (2022). Die Validierungspraxis bei hands-on Experimentiertests in der naturwissenschaftsdidaktischen Forschung. *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1). <https://doi.org/10.1007/s40573-022-00145-6>
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 15(1). <https://doi.org/10.17169/FQS-15.1.2043>
- Schröder, J., Riese, J., Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Kulgemeyer, C., Reinhold, P. & Schecker, H. (2020). Die Messung der Fähigkeit zur Unterrichtsplanung im Fach Physik mit Hilfe eines standardisierten Performanztests. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 103–122. <https://doi.org/10.1007/s40573-020-00115-w>
- Schulz, W. (1975). Unterricht - Analyse und Planung. In P. Heimann, G. Otto & W. Schulz (Hrsg.), *Unterricht: Analyse und Planung* (7. Aufl., S. 13–47). Hermann Schroedel Verlag.
- Schulz, W. (1980). Die lerntheoretische Didaktik. *Westermanns pädagogische Beiträge*, 32(2), 80–85.
- Schurz, G. (2008). Patterns of abduction. *Synthese*, 164(2), 201–234. <https://doi.org/10.1007/s11229-007-9223-4>
- Seel, A. (2011). Wie angehende Lehrer/innen das Planen lernen. In K.-H. Arnold, T. Bohl & K. Zierer (Hrsg.), *Jahrbuch für Allgemeine Didaktik 2011: Thementeil: Entwicklung und Weiterentwicklung allgemeindidaktischer Modelle der Unterrichtsplanung* (S. 31–45). Schneider Verlag Hohengehren.



- Seidel, T., Rimmele, R. & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction*, 15(6), 539–556.  
<https://doi.org/10.1016/j.learninstruc.2005.08.004>
- Şen, M. (2023). Suggestions for the Analysis of Science Teachers' Pedagogical Content Knowledge Components and Their Interactions. *Research in Science Education*, 1–15.  
<https://doi.org/10.1007/s11165-023-10124-7>
- Şen, M., Demirdöğen, B. & Öztekin, C. (2022). Interactions among topic-specific pedagogical content knowledge components for science teachers: the impact of content knowledge. *Journal of Science Teacher Education*, 33(8), 860–887.  
<https://doi.org/10.1080/1046560X.2021.2012630>
- SenBJF. (2015). *Rahmenlehrplan für die Jahrgangsstufen 7-10: Teil C (Biologie)*.  
[https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/rahmenlehrplaene/Rahmenlehrplanprojekt/amtliche\\_Fassung/Teil\\_C\\_Biologie\\_2015\\_11\\_10\\_WEB.pdf](https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/rahmenlehrplaene/Rahmenlehrplanprojekt/amtliche_Fassung/Teil_C_Biologie_2015_11_10_WEB.pdf)
- SenBJF. (2017). *Handbuch Vorbereitungsdienst: Materialien für den reformierten Berliner Vorbereitungsdienst* (6. Auflage).  
[https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch\\_vorbereitungsdienst.pdf](https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch_vorbereitungsdienst.pdf)
- Verordnung über den Vorbereitungsdienst und die Staatsprüfung für Lehrämter,  
<https://gesetze.berlin.de/perma?d=jlr-LehrVorbDStPrVBEV5IVZ> (2022 & i.d.F.v. 05.08.2022). <https://gesetze.berlin.de/perma?d=jlr-LehrVorbDStPrVBEV5IVZ>
- SenBJF & MBJS. (2021). *Rahmenlehrplan für die gymnasiale Oberstufe: Teil C (Berlin und Brandenburg)*.  
[https://www.berlin.de/sen/bildung/unterricht/faecher-rahmenlehrplaene/rahmenlehrplaene/rfp\\_go\\_teil\\_c\\_biologie\\_2021.pdf?ts=1684146726](https://www.berlin.de/sen/bildung/unterricht/faecher-rahmenlehrplaene/rahmenlehrplaene/rfp_go_teil_c_biologie_2021.pdf?ts=1684146726)
- Shavelson, R. J. (1972). Some Aspects of the Correspondence between Content Structure and Cognitive Structure in Physics Instruction. *Journal of Educational Psychology*, 63(3).
- Shavelson, R. J. & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51(4), 455–498.  
<https://doi.org/10.3102/00346543051004455>
- Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard educational review*, 57(1), 1–22.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21(2), 141–157.  
[https://doi.org/10.1016/0883-0355\(94\)90029-9](https://doi.org/10.1016/0883-0355(94)90029-9)
- Spörhase, U. (Hrsg.). (2021). *Biologie-Didaktik: Praxishandbuch für die Sekundarstufen I und II* (8. Aufl.). Cornelsen.
- Stamann, C., Janssen, M. & Schreier, M. (2016). Qualitative Inhaltsanalyse - Versuch einer Begriffsbestimmung und Systematisierung. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 17(3), 1–16.
- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2022). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 07.10.2022*.  
[https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf)

- Stein, M. K., Grover, B. W. & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455. <https://doi.org/10.2307/1163292>
- Stein, M. K. & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: an analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, 2(1), 50–80. <https://doi.org/10.1080/1380361960020103>
- Stender, A. (2014). *Unterrichtsplanung: Vom Wissen zum Handeln: theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung. Studien zum Physik- und Chemielernen: Bd. 170*. Logos Verlag.
- Stender, A. & Brückmann, M. (2020). Processes of knowledge integration between realms of pedagogical content knowledge and how to capture them. In T. Lehmann (Hrsg.), *International perspectives on knowledge integration: Theory, Research, and good practice in pre-service teacher and higher education* (S. 111–132). Brill.
- Stender, A., Brückmann, M. & Neumann, K. (2017). Transformation of Topic-Specific Professional Knowledge into Personal Pedagogical Content Knowledge through Lesson Planning. *International Journal of Science Education*, 39(12), 1690-1714. <https://doi.org/10.1080/09500693.2017.1351645>
- Strietholt, R. & Terhart, E. (2009). Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung. *Zeitschrift für Pädagogik*, 55, S. 622–645. <http://www.content-select.com/10.3262/ZP0904622>
- Tardent Kuster, J. (2020). *Unterrichtsplanungen von angehenden Lehrpersonen zum experimentellen Handeln: Eine videografegestützte Analyse von Unterrichtsplanungen*. Pädagogische Hochschule Heidelberg.
- Taut, S. & Rakoczy, K. (2016). Observing instructional quality in the context of school evaluation. *Learning and Instruction*, 46, 45–60. <https://doi.org/10.1016/j.learninstruc.2016.08.003>
- Teo, T., Chai, C. S., Hung, D. & Lee, C. B. (2008). Beliefs about teaching and uses of technology among pre-service teachers. *Asia-Pacific Journal of Teacher Education*, 36(2), 163–174. <https://doi.org/10.1080/13598660801971641>
- Terhart, E. (2019). *Didaktik: Eine Einführung*. Reclam.
- Tierney, R. & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research, and Evaluation*(9), Artikel 2, 1–7. <https://doi.org/10.7275/jtvt-wg68>
- Tomas, C., Whitt, E., Lavelle-Hill, R. & Severn, K. (2019). Modeling Holistic Marks With Analytic Rubrics. *Frontiers in Education*, 4, Artikel 89, 1–19. <https://doi.org/10.3389/educ.2019.00089>
- Turley, E. D. & Gallagher, C. W. (2008). On the "uses" of rubrics: reframing the great rubric debate. *The English Journal*, 97(4), 87–92.
- Upmeier zu Belzen, A., Alonzo, A. C., Krell, M. & Krüger, D. (2019). Learning Progressions and Competence Models: A Comparative Analysis. In E. McLoughlin, O. E. Finlayson, S. Erduran & P. E. Childs (Hrsg.), *Contributions from Science Education Research. Bridging Research and Practice in Science Education* (Bd. 6, S. 257–271). Springer International Publishing. [https://doi.org/10.1007/978-3-030-17219-0\\_16](https://doi.org/10.1007/978-3-030-17219-0_16)
- Upmeier zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.

- van der Valk, T. A. E. & Broekman, H. H. G. B. (1999). The lesson preparation method: a Way of Investigating Pre-Service Teachers' Pedagogical Content Knowledge. *European Journal of Teacher Education*, 22(1), 11–22.
- van Driel, J. H., Berry, A. & Meirink, J. A. (2014). Research on science teacher knowledge. In N. G. Lederman & S. K. Abell (Hrsg.), *Handbook of research on science education* (2. Aufl., S. 848–870). Routledge.
- Veal, W. R. & MaKinster, J. (1999). Pedagogical content knowledge taxonomies. *Electronic Journal of Science Education*, 3(4), 1–18.
- Vogelsang, C., Kulgemeyer, C. & Riese, J. (2022). Learning to plan by learning to reflect? - Exploring relations between professional knowledge, reflection skills, and Planning skills of preservice physics teachers in a one-semester field experience. *Education Sciences*, 12(7), 479. <https://doi.org/10.3390/educsci12070479>
- Vogelsang, C. & Riese, J. (2017). Wann ist eine Unterrichtsplanung ‚gut‘? - Planungsperformanz in Praxisratgebern zur Unterrichtsplanung. In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 47–61). Julius Klinkhardt.
- Vorholzer, A. & Aufschnaiter, C. von (2020). Dimensionen und Ausprägungen fachinhaltlicher Kompetenz in den Naturwissenschaften – ein Systematisierungsversuch. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 1–18. <https://doi.org/10.1007/s40573-020-00107-w>
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969. <https://doi.org/10.1037/a0025125>
- Wallace, A. R. (1855). On the law which has regulated the introduction of new species. *Annals and Magazine of Natural History, including Zoology, Botany, and Geology*, 16, 184–196.
- Watkins, M. W. (2018). Exploratory factor analysis: a guide to best practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Weber, K. E. & Czerwenka, K. (2021). Anschlussfähigkeit und Kooperation der ersten und zweiten Phase der Lehrkräftebildung. In J. Peitz & M. Harring (Hrsg.), *Das Referendariat: Ein systematischer Blick auf den schulpraktischen Vorbereitungsdienst* (S. 255–264). Waxmann.
- Wegener, D. T. & Blankenship, K. L. (2007). Ecological validity. In R. F. Baumeister & K. D. Vohs (Hrsg.), *Encyclopedia of social psychology* (S. 275–277). Sage Publications.
- Weinert, F. E. (2001). Vergleichende Leistungsmessungen an Schulen. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Beltz.
- Weingarten, J. (2019). *Wie planen angehende Lehrkräfte ihren Unterricht? Empirische Analysen zur kompetenzorientierten Gestaltung von Lernangeboten*. Waxmann.
- Weitzel, H. & Blank, R. (2020). Pedagogical Content Knowledge in peer dialogues between pre-service biology teachers in the planning of science lessons. Results of an intervention Study. *Journal of Science Teacher Education*, 31(1), 75–93. <https://doi.org/10.1080/1046560X.2019.1664874>
- Wellnitz, N. (2013). Frucht oder Nicht-Frucht: Kriteriengeleitetes Vergleichen zur Systematik von essbaren Pflanzenteilen. In P. Schmiemann & J. Mayer (Hrsg.), *Experimentieren Sie! Biologieunterricht mit Aba-Effekt: Selbstständiges, kompetenzorientiertes Erarbeiten von Lehrplaninhalten*. (S. 22–24). Cornelsen.

- Wellnitz, N. & Mayer, J. (2013). Erkenntnismethoden in der Biologie: Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 315–345.
- Wengert, H. G. (1989). *Untersuchungen zur alltaeglichen Unterrichtsplanung von Mathematiklehrern: Eine kognitionspsychologische Studie*. Peter Lang.
- Wernke, S. & Zierer, K. (2017). Die Unterrichtsplanung- Ein in Vergessenheit geratener Kompetenzbereich?! In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 7–16). Julius Klinkhardt.
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292–305.
- Whittington, K., Southerland, S. A. & Tekkumru-Kisa, M. (2021). Examining relevance in pre-service science teachers' lesson plans. *International Journal of Science Education*, 1–23. <https://doi.org/10.1080/09500693.2021.2012850>
- Wiggins, G. P. & McTighe, J. (2005). *Understanding by design* (2. Aufl.). Association for Supervision and Curriculum Development.
- Wittwer, J., Kratschmayr, L. & Voss, T. (2020). Wie gut erkennen Lehrkräfte typische Fehler in der Formulierung von Lernzielen? *Unterrichtswissenschaft*, 48(1), 113–128. <https://doi.org/10.1007/s42010-019-00056-5>
- Wolf, J. (2022). *(Wie) Wirkt Lehrerbildung? Ein Beitrag zur geschichtsdidaktischen Kompetenz- und Lehrerbildungsforschung am Beispiel des Unterrichtsplanungswissens von angehenden Geschichtslehrerinnen und -lehrern*. Ruhr-Universität Bochum. <https://doi.org/86599>
- Yinger, R. J. (1980). A Study of teacher planning. *Elementary School Journal*, 80(3), 107–127.
- Zabel, J. (2007). Was tut das Tier? Beobachten und Deuten lernen anhand von Verhaltensprotokollen. In R. Duit, H. Gropengießer & L. Stäudel (Hrsg.), *Naturwissenschaftliches Arbeiten: Unterricht und Material 5-10* (S. 12–17). Friedrich.
- Zahorik, J. A. (1975). Teachers' planning models. *Educational Leadership*, 33, 134–139.
- Zaragoza, A., Seidel, T. & Santagata, R. (2023). Lesson analysis and plan template: scaffolding preservice teachers' application of professional knowledge to lesson planning. *Journal of Curriculum Studies*, 1–15. <https://doi.org/10.1080/00220272.2023.2182650>
- Ziepprecht, K. & Meier, M. (2019). Biologische Rätsel. *Unterricht Biologie kompakt*, 43(450).
- Zierer, K., Werner, J. & Wernke, S. (2015). Besser planen? Mit Modell! Empirisch basierte Überlegungen zur Entwicklung eines Planungskompetenzmodells. *Die deutsche Schule*, 107(4), 376–396.

## Anhang

### Anhang 1: Kodierleitfaden *PCK mapping* (Beiträge 2 und 3)

#### A. Problemaufriss und Fragestellungen

Obwohl die Planung von Unterricht genuiner Bestandteil der professionellen Kompetenzen von Lehrkräften ist (BAUMERT & KUNTER, 2006) und als ihre „Kernaufgabe“ gilt (KMK, 2005), liegt bislang weder ein Instrument zur standardisierten Erfassung von Planungskompetenz (KÖNIG, BUCHOLTZ & DOHMEN, 2015) noch ein etabliertes allgemein- oder fachdidaktisches Kompetenzmodell der Planungskompetenz (WERNKE & ZIERER, 2017) vor.

Als Teil der professionellen Handlungskompetenzen von Lehrkräften bildet neben dem Fachwissen (*content knowledge*, kurz CK) und dem pädagogischen Wissen (*pedagogical knowledge*, kurz PK) das fachdidaktische Wissen (*pedagogical content knowledge*, kurz PCK) den für die Profession von Lehrkräften konstitutiven Wissensbereich (BAUMERT & KUNTER, 2006).

Da die Expertise von Lehrkräften u.a. elaboriertes fachdidaktisches Wissen voraussetzt, soll im Rahmen dieser Studie mittels eines Kategoriensystems das individuelle, sich in einer konkreten Situation manifestierende fachdidaktische Wissen (enacted PCK, kurz ePCK; CARLSON & DAEHLER, 2019) von Biologie-Referendar\*innen erhoben werden. Aus der Untersuchung von schriftlichen Unterrichtsplanungen lassen sich für jede\*n Referendar\*in die fachdidaktischen Wissensbestände ermitteln, die für die Planung des Unterrichtsvorhabens relevant sind.

Das fachdidaktische Wissen von Lehrkräften naturwissenschaftlicher Fächer lässt sich als fünfdimensionales Konstrukt verstehen (PARK & OLIVER, 2008; SUH & PARK, 2017), das folgende Wissensfacetten umfasst:

1. Wissen über kontextuelle Rahmenbedingungen
2. Wissen über das didaktische Potential des Themas
3. Wissen über Schülerkognition
4. Wissen über fachspezifische Instruktionsstrategien
5. Wissen über Diagnostik

Werden alle diese Facetten bei der Planung von Unterricht berücksichtigt, ließe sich das Ausmaß der Lehr-Lern-Situationen per sé innewohnenden Kontingenz (ROSENBERGER 2018) auf elaborierte Weise reduzieren. Die dazu notwendigen Planungsentscheidungen lassen sich den fünf Facetten zuordnen und dienen in ihrer Wechselwirkung dazu, Lernprozesse anzuregen und zu steuern. Sie werden in einem schriftlichen Unterrichtsentwurf dargelegt. Dabei sollte es sich jedoch nicht bloß um Deskriptionen oder die Schilderung subjektiver Erfahrungen handeln, sondern im Sinne des *pedagogical reasoning* (LOUGHRAN, KEAST & COOPER, 2016) um eine Analyse und Begründung von Planungsentscheidungen. Das Kategoriensystem enthält aus diesem Grund Kategorien, die auf Deskription abzielen, und solche, die auf Begründungen abzielen. Durch die Inhaltsanalyse schriftlicher Unterrichtsplanungen lassen sich die Vernetzungen des fachdidaktischen Wissens ermitteln und in Form von PCK-Maps visualisiert darstellen (PARK, 2019). Die erste und derzeit einzige Anwendung dieser Methode in der deutschsprachigen biologiepädagogischen Forschung hat POHLMANN (2019) anhand von Interview-Transkripten mit Lehrkräften im Bereich der Bewertungskompetenz vorgenommen.

Beim vorliegenden Datenmaterial handelt es sich um Unterrichtsentwürfe der unterrichtspraktischen Prüfung im Staatsexamen im Fach Biologie im Land Berlin. Gemäß der „Verordnung über den Vorbereitungsdienst und die Staatsprüfung für Lehrämter“ (SENBJF, 2014) müssen die Referendar\*innen in ihren beiden Fächern jeweils eine Unterrichtsstunde durchführen (ebd. §22, Abs. 1). Die Note für die unterrichtspraktische Prüfung ergibt sich für beide Fächer jeweils aus der Planung, der Durchführung und der Reflexion im Analysegespräch mit dem Prüfungsausschuss (ebd., §22, Abs. 3).

Da es sich im Rahmen der Lehrkräftebildung um die finale Prüfungsleistung handelt, kann davon ausgegangen werden, dass vielen der schriftlichen Unterrichtsplanungen elaborierte Planungsentscheidungen zugrunde liegen und sich die verschiedenen Wissensbestände in den Entwürfen wiederfinden lassen. Es lässt sich annehmen, dass

die Planungen am Ende der zweiten Phase der Lehrkräftebildung insofern elaborierter sind als beispielsweise Planungen von Master-Studierenden, als die Vernetzung der Wissensbestände komplexer und die Qualität des *pedagogical reasoning* höher ist. Bislang gibt es nur wenige Untersuchungen der Unterrichtsplanungen von Referendar\*innen. So untersucht beispielsweise WEINGARTEN (2019) mithilfe eines allgemein-didaktischen Kategoriensystems die grundsätzliche Planungspraxis, KÖNIG, BUCHHOLTZ & DOHMEN (2015) nehmen speziell die didaktische Adaptivität in den Blick. Eine biologiespezifische Untersuchung der Planungskompetenz angehender Lehrkräfte liegt bislang nicht vor.

Dieses Desiderat der biologiedidaktischen Forschung soll durch eine empirisch fundierte Analyse schriftlicher Unterrichtsplanungen geschlossen werden. Aus diesen grundsätzlichen Überlegungen ergeben sich die folgenden forschungsleitenden Fragestellungen:

- F<sub>1</sub>: *Inwiefern lässt sich in den Staatsexamensentwürfen verknüpftes fachdidaktisches Wissen (PCK) identifizieren?*
- F<sub>2</sub>: *Wie werden Planungsentscheidungen begründet?*
- F<sub>3</sub>: *Lassen sich kompetenzbereichsspezifische Qualitätsmerkmale beschreiben, die Lehramts-studierenden oder Referendar\*innen bei der Planung von Biologieunterricht als Handlungsanleitung helfen können?*

## B. Hinweise für die Durchführung der Codierung

### B.1. Definition der Auswahleinheit

Die Auswahleinheit „beinhaltet die physisch vorliegenden Materialien, die aus dem gesamten Spektrum verfügbaren Medienmaterials für die Untersuchung ausgewählt werden“ (RÖSSLER 2005, S.39). Demnach wird systematisch festgelegt, welche Teile des Medienmaterials mittels qualitativer Inhaltsanalyse untersucht werden sollen. Im Rahmen dieser Studie wird eine **Stichprobe von 111 schriftlichen Unterrichtsentwürfen** im Fach Biologie der Sekundarstufen I und II im Rahmen der Staatsexamensprüfung für das Lehramt im Land Berlin aus den Jahren 2018 und 2019 untersucht. Es handelt sich nicht um eine gezogene Stichprobe aus einer größeren Menge an Entwürfen, sondern um sämtliche Entwürfe aus den letzten beiden Jahren. Die Entwürfe verteilen sich folgendermaßen auf die vier Kompetenzbereiche des Fachs Biologie (KMK 2004, SENBJF 2015):

	Fachwissen	Erkenntnisgewinnung	Kommunikation	Bewertung/Reflexion
Sek. I	6	7	23	2
Sek. II	14	16	36	7

Da die Entwürfe vor der Übergabe durch die Senatsverwaltung für Bildung, Jugend und Familie des Landes Berlin vollständig anonymisiert worden sind, können bei der Stichprobenziehung keine Kontextvariablen (Geschlecht, Alter, Nebenfächer, Examensnote o.ä.) berücksichtigt werden.

### B.2. Definition der Analyseeinheit

Als Analyseeinheit werden „jene Elemente aus dem Untersuchungsmaterial [definiert], für die im Rahmen der Codierung eine Klassifizierung vorgenommen wird“ (RÖSSLER 2005, S.40). Sie bildet die Grundlage für die Codierung und damit für die Analyse des vorliegenden Medienmaterials. Im Rahmen dieser Studie bildet die Analyseeinheit **ein schriftlicher Unterrichtsentwurf ohne Anhang**. Codiert wird neben den verbindlichen Bestandteilen eines Unterrichtsentwurfs (Sachstrukturanalyse, Analyse der Lehr-Lern-Struktur, Kompetenzschwerpunkt usw.) in jedem Fall auch die schriftliche Verlaufsplanung, auch wenn diese in einigen Entwürfen dem Anhang zugeordnet wird.

### B.3. Definition der Codiereinheit

Die Codiereinheit „benennt diejenigen Aspekte, die an dem Medienmaterial interessant sind, um die Forschungsfrage zu beantworten und durch die Kategorien adressiert werden“ (RÖSSLER 2005, S.41). Demnach werden alle Stellen innerhalb des Medienmaterials codiert, denen eine Kategorie aus dem Kategoriensystem zugeordnet werden kann. Im Rahmen dieser Studie wird die Codiereinheit formal definiert als **mindestens vollständiger Teilsatz**, d. h. bis zum Komma in Satzgefügen oder bis zum Punkt in Hauptsätzen bzw. bis zur Konjunktion in Satzreihen.

### B.4. Definition der Kontexteinheit

Die Kontexteinheit „hilft dem Codierer, den korrekten Zusammenhang der Analyseeinheit zu erfassen, indem sie in Zweifelsfällen den Rückgriff auf einen größeren Berichtskontext erlaubt“ (RÖSSLER 2005, S.42). Im Rahmen dieser Studie bildet die Kontexteinheit **der Anhang eines schriftlichen Unterrichtsentwurfs**. Dies ist beispielsweise dann hilfreich, wenn in der Analyse der Lehr-Lern-Struktur Aufgabenstellungen beschrieben und begründet werden, die im Wortlaut nur auf den Arbeitsblättern zu finden sind, die sich im Anhang befinden.

## C. Einfach- und Mehrfachcodierungen

### C.1. Generelle Regelungen zur Codierung

- (1) Bitte lesen Sie zunächst den gesamten Unterrichtsentwurf aufmerksam durch.
- (2) Je Codiereinheit wird ein Code auf der Ebene der Subkategorie vergeben. Dabei sind intuitive Entscheidungen zu vermeiden und stattdessen die Codierhinweise in der Codieranleitung zu beachten.
- (3) Innerhalb eines Unterrichtsentwurfs kann eine Codiereinheit mehreren Subkategorien zugeordnet werden.
- (4) Notieren Sie Unklarheiten (z. B. aufgrund fehlender Trennschärfe der Codebeschreibungen), so dass das Kategoriensystem gegebenenfalls optimiert werden kann.

### C.2. Spezielle Regelungen zur Codierung

- (1) Die Qualität der Ausführungen wird nicht bewertet. So wird eine Codiereinheit in jedem Fall einer Subkategorie zugeordnet, auch wenn sie nur beiläufig und wenig elaboriert dargestellt wird (z. B. muss eine tabellarische Sequenzplanung, in denen Standards und Inhaltsbeschreibungen fehlen, dennoch als 1.2.2.1. codiert werden).
- (2) Nicht jeder Satz innerhalb eines Unterrichtsentwurfs muss codiert werden. Folgende Elemente bleiben uncodiert:
  - Deckblatt und Anhang (exkl. tabellarischer Stundenverlaufsplan)
  - allg. Lerngruppenanalyse (z. B. Geschlechterverhältnis)
  - räumliche Ausstattung
  - Beschreibung von Problemen einzelner SuS (sonderpädagogischer Förderbedarf,
- (3) Da das Kategoriensystem neben deskriptiven Elementen auch im Sinne des *pedagogical reasoning* Begründungszusammenhänge erfasst, ist aufmerksam zu unterscheiden, ob ein Teilsatz eine Planungsentscheidung beschreibt oder sie begründet. Unter Begründungen werden neben kausalen Fügungen (weil, da, aufgrund, wegen, o.ä.) auch finale Fügungen (damit, um...zu, o.ä.) oder nur implizite Formulierungen verstanden:

- i. **kausal:** [...] da wir die Fortbewegung an aktiven Tieren beobachten wollen.
- ii. **final:** [...] um eine überschaubare Auswahl zu gewährleisten und unterschiedliche Ergebnisse der SuS zu erhalten.
- iii. **implizit:** Das Tafelbild soll die SuS zu einer Positionierung anregen.  
Auf eine Zwischensicherung wird aus Zeitgründen verzichtet.



## D. Kategoriensystem

### D.1. Wissen über kontextuelle Rahmenbedingungen

Lehrkräfte müssen bei ihrer Unterrichtsplanung die nationalen Bildungsstandards (KMK 2004) bzw. die sich daraus ergebenden Kompetenzanforderungen in den Rahmenlehrplänen (SENBJF 2015; SENBJS 2006) beachten. Wissen über derartige Richtlinien und die Prüfungsaufgaben ist für die Planung von Unterricht essentiell, um den SuS einen handelnden Umgang mit Wissen zu ermöglichen (WEINERT 2001) und sie langfristig und kumulativ auf die auf den normativen Referenzsystemen basierenden Abschlussprüfungen vorzubereiten.

### 1.1. normative Orientierung

Referenzsysteme für die Planung von Unterricht sind zunächst die Steuerungsdokumente, die nationale (Bildungsstandards) und länderspezifische (Rahmenlehrpläne) Vorgaben machen. Zudem werden die Inhalte und Kompetenzen von jeder Schule in ein schulinternes Curriculum eingebunden und ausgestaltet. Insofern stellt die normative Orientierung eine wichtige Wissenskomponente dar, um den eigenen Unterricht anschlussfähig zu machen (z. B. schulübergreifend, jahrgangübergreifend) und die SuS somit verbindlich auf Abschlussprüfungen vorzubereiten.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
1.1.1. Steuerungsdokumente	Eine Orientierung an den Richtlinien der KMK und den Lehrplänen.	<i>Es werden Bezüge auf die KMK und die Rahmenlehrpläne codiert, wenn diese oder Teile aus ihnen explizit genannt bzw. gar direkt zitiert werden. Kompetenzbezüge werden nur codiert, wenn auf sie in der allgemeinen Fassung (RLP, KMK) Bezug genommen wird. Nicht codiert werden Standardkonkretisierungen (vgl.1.2.1.1.).</i>	<i>Die SuS können Experimente zur Überprüfung von Hypothesen nach Vorgaben durchführen. (StEx_S1_E_NU_Ö_01)</i>  <i>Im Sinne einer <b>Gesundheitsförderung und -erziehung</b> (StEx_S1_B_Hd_Sex_01)</i>
1.1.2. Schulinternes Curriculum	<i>Informationen zum schulinternen Curriculum</i>	<i>Es werden Bezüge auf das SchiC codiert, wenn dieses explizit genannt bzw. gar direkt zitiert wird.</i>	<i>Die vorliegende Unterrichtsreihe orientiert sich sowohl an dem Rahmenlehrplan der gymnasialen Oberstufe des Landes Berlin als auch am <b>Schulinternen Curriculum des Kant-Gymnasiums</b>. (StEx_S2_K_Ö_04)</i>

### 1.2. Kompetenzorientierung

Laut Weinert (2001) lassen sich Kompetenzen als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und

Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und

Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ verstehen. Diese Kompetenzen sollten Ausgangspunkt der Unterrichtsplanung sein, so dass langfristig ein Kompetenzzuwachs nachweisbar wird. Insofern muss unterschieden werden zwischen der stundenspezifischen

Kompetenzorientierung (*Sind alle Planungsentscheidungen funktional, um den in der Einzelstunde zu fördernden Standard zu bedienen?*) und der langfristigen Kompetenzorientierung (*Ermöglicht die Gestaltung einer Unterrichtssequenz eine schrittweise Kompetenzentwicklung?*).

### 1.2.1. Unterrichtsstunde

Auch wenn eine Kompetenzentwicklung der SuS langfristig angelegt und nicht in einer 45-Minuten-Stunde abzuhandeln ist, bildet eine Einzelstunde einen Baustein, in dem ein Standard aus dem Rahmenlehrplan gefördert werden soll. Dieses Unterrichtsziel wird in Form einer Standardkonkretisierung formuliert (MAYER 2019). Anhand von Indikatoren, muss erkennbar werden, ob der angestrebte Kompetenzzuwachs erreicht worden ist.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
1.2.1.1. Beschreibung des Kompetenzzuwachses	Es werden Informationen zu den Kompetenzen, welche die SuS in dem geplanten Lernvorhaben verbessern oder neu erwerben sollen, erteilt.	Es werden alle Bezüge auf die in der Einzelstunde zu fördernden Kompetenzen samt Erklärungen zur Relevanz codiert. Dabei werden nicht nur "Standard-konkretisierungen", sondern auch die konkreten Kompetenzbezüge im Fließtext, Ausführungen für den "Mindest-", "Regel-" und "Maximalstandard" sowie die sogenannte "Hauptlernaktion" codiert.	Die Schülerinnen <b>nennen</b> die Vorteile, welche Zooxanthellen und Korallenpolyp der Steinkoralle aus ihrer Symbiose ziehen, <b>stellen</b> auf der Grundlage dieses Wissens <b>Hypothesen</b> zur Erklärung [...] und de[s] damit verbundenen Korallensterben[s] <b>auf</b> und <b>erklären</b> diese Phänomene im Anschluss auf der Grundlage eines Infotextes. (StEx_S2_F_Ö_04)  (...) ermöglicht eine <b>Progression innerhalb der Kommunikationskompetenz</b> . (StEx_S2_K_Ö_04)
1.2.1.2. Indikatoren zum Nachweis eines Kompetenzzuwachses	Es werden Indikatoren genannt, anhand derer ein Kompetenzzuwachs nachgewiesen werden kann.	Es werden äußerlich wahrnehmbare Indikatoren in Bezug auf den konkretisierten Standard codiert, keine allgemeinen Beschreibungen dessen, was die SuS erarbeiten (Vergleiche: "Die SuS markieren in einem Fachtext Schlüsselwörter und fassen ihn mündlich zusammen." versus "Die SuS lesen einen Fachtext.").	Die Schüler*innen <b>können</b> die Fortbewegung des Regenwurms unter Einbeziehung seines strukturellen Aufbaus <b>erklären</b> . (StEx_S1_E_NU_Ö_01)

### 1.2.2. Unterrichtssequenz

Kompetenzentwicklung kann nicht in 45-Minuten-Einheiten gedacht werden, sondern wird im Sinne des kumulativen Lernens (FREIMAN 2001) in einer zumeist acht bis zwölf Stunden umfassenden Sequenz progressiv gestaltet. Die Reihenfolge der Lerninhalte sowie die einzelnen Schritte des Kompetenzerwerbs (LEISEN 2014) werden in einer Sequenzplanung (oft auch Reihenplanung) beschrieben und im Fließtext erläutert und begründet.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
1.2.2.1. Beschreibung der Struktur der Unterrichtssequenz	<i>Der Kompetenzerwerb wird in Teilschritten dargelegt, so dass deutlich wird, welche Kompetenzen die SuS im Verlauf der Unterrichtssequenz erwerben sollen.</i>	<i>Es werden Stellen codiert, in denen auf den über die Einzelstunde hinausreichenden Kompetenzerwerb (retrospektiv und prospektiv) eingegangen wird. Auch Verweise und Bezüge auf einschlägige Kompetenzmodelle werden als Indiz für einen nachhaltigen Kompetenzaufbau in Stufen codiert. Nicht codiert werden Stellen, die sich ausschließlich auf die Einzelstunde beziehen (vgl. 1.2.2.1.). Nicht codiert werden Stellen zum Vorwissen (vgl. 3.1.1.).</i>	<i>"Die Kompetenzschwerpunkte der gesamten Unterrichtsreihe unterliegen hauptsächlich einem alternierenden Wechsel von Fachwissen und Kommunikation. (StEx_S2_K_P_04)</i>  <i>Die Reflexion von Handlungsalternativen sollte dabei stets das Ziel für die SuS sein. (StEx_S2_F_Ö_04)"</i>  <i>Der <b>Schwerpunkt der Unterrichtsreihe</b> ist neben der Vermittlung biologischen Fachwissens zum Thema Zellteilung [...], vor allem die Verwendung biologischer Fachsprache. Die SuS erklären <b>während der Unterrichtsreihe</b> Genetik Fachbegriffe und <b>wenden sie vermehrt</b> im Unterrichtsalltag <b>an</b>. (StEx_S1_B_Hr_G_01)</i>
1.2.2.2. Begründung der Struktur der Unterrichtssequenz	<i>Der angestrebte Kompetenzzuwachs sowie die Anordnung der Lernschritte in der Sequenzplanung wird begründet, so dass ersichtlich wird, warum die zeitliche Abfolge der Einzelstunden so gewählt worden ist.</i>	<i>Es werden Stellen codiert, in denen der langfristige Kompetenzerwerb und die daraus resultierende Sequenzstruktur begründet wird. Dabei muss der Kompetenzbereich/Standard nicht explizit genannt werden. Auch codiert werden Alternativen, die vorgeschlagen werden, so wie die Gründe dafür, warum man sich dagegen entschieden hat.</i>	<i>"Ich habe mich für die regelmäßigen Kommunikationsstunden entschieden, weil die Lernenden Defizite aufweisen im Umgang mit Informationsmaterialien, wie z. B. Diagrammen, Tabellen oder Lehrbuchtexten sowie in der Textproduktion. (StEx_S2_K_P_04)</i>  <i>[Die Unterrichtssequenz soll] die SuS dazu befähigen [...] gesellschaftliche Meinungsbildungsprozesse und Entscheidungen, welche eben dieses Know-how benötigen, mitgestalten zu können und verschiedene Handlungsmöglichkeiten zum Schutz unseres Planeten entwickeln zu können. (StEx_S2_F_Ö_04)"</i>

### 1.2.3. Unterrichtsprinzipien

Um eine angestrebte Kompetenzentwicklung zu fördern, kann auf Unterrichtsprinzipien zurückgegriffen werden, die Lehr-Lern-Prozesse im Biologieunterricht fördern (KÖHLER 2004):

- *Prinzip der Anschaulichkeit*: Einsatz von Anschauungsmitteln, die Lerninteresse der SuS verstärken und das Verstehen/Behalten durch Konkretisierung erleichtern
- *Prinzip des Exemplarischen*: Wahl eines interessanten, repräsentativen und überschaubaren Themas innerhalb eines größeren Bereichs; Einbezug von Wissenschaftspropädeutik
- *Prinzip der Handlungsorientierung*: schülerzentrierter Unterricht, indem Lernprodukte hergestellt und reflektiert werden; keine einseitige Fokussierung auf Kognition
- *Prinzip der Situationsorientierung*: Herstellung von Anschlussfähigkeit des biologischen Fachwissens an schüler- bzw. gesellschaftsrelevante Themen und Erfahrungen
- *Prinzip der Problemorientierung*: Gestaltung des Lernprozesses von einer biologischen Problemstellung ausgehend, die die SuS eigenständig lösen
- *Prinzip der Wissenschaftsorientierung*: Anwendung und Reflexion über fachgemäße Arbeits- und Denkweisen und Einbezug wissenschaftspropädeutischer Aspekte
- *Prinzip des Pflegerischen & formenkundlichen Prinzip*: Erlernen eines verantwortungsvollen Umgangs mit Lebewesen & Erkennen von Vielfalt und Komplexität der Lebewesen

Einige dieser Prinzipien können implizit zur Begründung von Planungsentscheidungen herangezogen werden, ohne dass diese explizit genannt werden.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
1.2.3. Unterrichtsprinzipien	Bezugnahme auf Unterrichtsprinzipien, die für den Unterricht eine Leitfunktion darstellen oder das Unterrichtsprofil schärfen: Anschaulichkeit, Genetisches und Exemplarisches Lehren und Lernen, Handlungsorientierung, Situationsorientierung, Problemorientierung, Wissenschaftsorientierung, Prinzip des Pflegerischen, Formenkundliches Prinzip)	Es werden Stellen codiert, in denen für den Biologieunterricht im Allgemeinen oder für den ausgewählten Kompetenzbereich im Besonderen charakteristische Unterrichtsprinzipien zum Ausdruck kommen.	Der Unterrichtsgang folgt <b>dem problemorientierten Unterrichtsverfahren</b> ohne Hypothesenbildung. (StEx_S2_K_P_04)  Im Sinne einer Gesundheitsförderung und -erziehung sollen die SuS durch die eingehende Thematisierung <b>exemplarischer Verfahren der PND</b> und deren Bewertung auf deskriptiver sowie normativer Ebene, dazu ermächtigt werden, in Zukunft Informationen über Verfahren der PND oder anderer medizinischer Verfahren, selbstständig ordnen und bewerten zu können und selbstbestimmte Entscheidungen zu fällen. (StEx_S1_B_Hd_Sex_01)

## D.2. Wissen über das didaktische Potential des Themas

Um ein geeignetes Thema für den Unterricht auswählen zu können, bedarf es Wissens über das didaktische Potential von Fachinhalten. Abhängig vom gewählten Kompetenzschwerpunkt muss ein dafür

geeignetes Thema gewählt (POHLMANN 2019, S. 259), fachwissenschaftlich durchdrungen und didaktisch in angemessener Weise aufbereitet werden.

### 2.1. Lerninhalt

Für die Auswahl von Fachinhalten sollten die Relevanzkriterien *Schülerrelevanz*, *Gesellschaftsrelevanz* oder *Wissenschaftsrelevanz* (KATTMANN 2019) maßgebend sein. Die sich daraus ergebenden (Teil-)Themen

einer Sequenz werden verknüpft und über die Basiskonzepte in den Einheitlichen Prüfungsanforderungen (KMK 2004) bzw. die acht großen Ideen der Biologie strukturiert (GROPENGBER 2010).

### 2.1.1. Themenauswahl

Als Thema wird einzig der biologische Fachinhalt in Bezug auf die zu unterrichtende Einzelstunde, nicht etwa das Thema der Sequenz oder ein Kompetenzbereich verstanden. Sollen beispielsweise im Rahmen eines Experiments zur Photosynthese Hypothesen formuliert und geprüft werden, wird unter Themenauswahl nur all jenes verstanden, was das Thema „Photosynthese“ im Allgemeinen bzw. das ausgewählte Experiment im Besonderen beschreibt und dessen Relevanz begründet.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
2.1.1.1. Beschreibung der Themenauswahl	Das spezifische (Teil-)Thema der Unterrichtsstunde wird benannt und seine Stellung innerhalb eines größeren Themenkomplexes beschrieben.	Es werden nur Stellen codiert, in denen die Auswahl des konkreten Stunden-themas, nicht des Kompetenzbereichs, beschrieben wird. Nicht codiert werden Schilderungen des Auswahlprozesses.	Von den Muskelfasern über die Zellorganellen (Mitochondrien) wird an die sich daran anschließende Stoffwechselfysiologie angeknüpft und auf der Ebene der Makromoleküle und Moleküle diskutiert. Der Querbrückenzyklus, welcher in dieser Unterrichtsstunde angesprochen wird, stellt die molekulare Erklärung der Muskelkontraktion dar. (StEx_S2_K_P_04)
2.1.1.2. Begründung der Themenauswahl	Es werden Gründe für die Auswahl des Themas genannt, ggf. unter explizitem oder implizitem Verweis auf Relevanzkriterien (Schülerrelevanz, Gesellschaftsrelevanz, Wissenschaftsrelevanz) oder das Exemplarische Prinzip.	Es werden Stellen codiert, in denen die Adäquatheit des ausgewählten Themas erläutert und begründet wird. Nicht codiert werden organisatorisch-pragmatische Gründe (z. B. "Thema ist im RLP/SchiC vorgeschrieben").	Die <b>Alltags- und Schülerrelevanz der Thematik</b> ergibt sich zum einen aus dem <b>Interesse der Lernenden</b> in Bezug auf den eigenen Körper. Pubertierende Lernende beobachten ihren Körper sehr genau und möchten dessen Funktionsweise verstehen. Zum anderen stellt auch der Tod aus biologischer Sicht ein für Lernende <b>relevantes Thema</b> dar. Zudem wird die <b>Neugier der Lernenden</b> angesprochen einen Kriminalfall aufzulösen. (StEx_S2_K_P_04)

### 2.1.2. Fachliche Klärung

Als analytischer Teil des Modells der Didaktischen Rekonstruktion (GROPENGBER & KATTMANN 2019) umfasst die fachliche Klärung die Exegese der für das gewählte Thema relevanten Erkenntnisse, Methoden, Theorien und Begriffe sowie die ihnen innewohnenden Zusammenhänge aus fachwissenschaftlicher Perspektive. Unter Beachtung unterrichtsrelevanter Aspekte (Lerngruppenanalyse, Materialausstattung, Zeitaufwand, Lehrplan usw.) muss die fachliche Komplexität reduziert werden – entweder durch strukturelle

Reduktion (Weglassen von Daten, die die Kernaussage nicht berühren, Überführen von verbalen Daten in einfache Visualisierungen, Zusammenfassen von Elementen zu größeren Einheiten) (STAECK 2010; GROPEGIEBER & KATTMANN 2019) oder durch sektorale Reduktion (Thematisierung nur eines Ausschnitts eines Systems oder Phänomens) (GROPEGIEBER & KATTMANN 2019).

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
2.1.2.1. Sachstrukturanalyse	Beschreibung und Analyse des Fachinhalts unter Bezugnahme auf Genese, Funktion und Bedeutung der themenspezifischen wissenschaftlichen Vorstellungen und Konzepte.	Es werden die Stellen/Abschnitte codiert, in denen der Lerninhalt aus fachwissenschaftlicher Perspektive betrachtet und analysiert wird. Nicht codiert werden Stellen, in denen die unterrichtspraktische Umsetzung beschrieben wird.	Die Frage nach einer angemessenen Pränataldiagnostik ist eng mit verschiedenen gesellschaftlichen Standpunkten hinsichtlich eines Schwangerschaftsabbruchs verwoben. (StEx_S1_B_Hd_Sex_01)
2.1.2.2. Didaktische Reduktion	Begründung der Vereinfachung des Lerninhalts durch Maßnahmen der strukturellen und der sektoralen Reduktion (GROPEGIEBER & KATTMANN 2019)	Es werden die Stellen codiert, in denen eine begründete Reduktion des Fachinhalts dargelegt wird.	[...] und aus diesem Grund habe ich mich dafür entschieden <b>die neuesten Forschungsergebnisse in reduzierter Form</b> anzugeben. (StEx_S2_F_Ö_04)

## 2.2. Vernetzung

Vernetzung kann auf zweierlei Weisen betrachtet werden: Versteht man erstens die Biologie als Brückenfach zwischen den Natur- und den Sozialwissenschaften (KATTMANN 2019b), dann ergeben sich neben den fachspezifischen Kompetenzanforderungen (SENBJF 2015b) auch fächerübergreifende Bildungsaufgaben (SenBJF 2015a, z. B. Gesundheitsförderung, Sexualerziehung, Nachhaltige Entwicklung). Zum Wissen über das fachdidaktische Potential von naturwissenschaftlichen Themen gehört somit auch Wissen über

Aspekte, wie die Fächer Physik, Chemie und Biologie verknüpft werden können, z. B. fachüberschreitend oder fächerverbindend (HARMS 2008). Zweitens kann im Sinne des Spiralcurriculums innerhalb des Fachs Biologie Bezug auf vorhergehende Jahrgangsstufen bzw. noch folgende Jahrgangsstufen genommen und auf dieser Grundlage die Ausgestaltung eines Themas (z. B. hinsichtlich Umfangs, Komplexität o.ä.) beschrieben und begründet werden.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
2.2.1. Verknüpfung mit horizontalem Curriculum	Bezugnahme auf fachübergreifende, fächer-verbundene und fachunabhängige Aspekte	Codiert werden Aussagen, die in Bezug auf schulischen Unterricht Querverbindungen zwischen dem Thema im Fach Biologie und einem anderen Fach bzw. mehreren anderen Fächern herstellen. Nicht codiert werden wissenschaftliche Verbindungen (z. B. DNA zwischen Zellbiologie und Biochemie).	
2.2.2. Verknüpfung mit vertikalem Curriculum	Bezugnahme auf das Spiralcurriculum und bereits erworbene Kenntnisse und Fähigkeiten in vorhergehenden Jahrgangsstufen bzw. noch zu erwerbende Kenntnisse und Fähigkeiten in folgenden Jahrgangsstufen	Codiert werden Aussagen, die das Thema und Bedingungen der Vermittlung in vorhergehenden und folgenden Jahrgangsstufen behandeln.	"Im Hinblick auf die Kursphase erachte ich es als notwendig den Lernenden geeignete Strategien zur Formulierung von aufgabenzentrierten Fließtexten an die Hand zu geben. (StEx_S2_K_P_04)"

### D.3. Wissen über Schülerkognition

Eine sorgsame Analyse der Lernvoraussetzungen der SuS vor dem Unterricht ist für sämtliche Planungsentscheidungen relevant, da kein Standard, keine Methode und keine Aufgabenstellung an sich gut oder schlecht geeignet ist, sondern nur im Sinne der didaktischen Adaptivität (KÖNIG, BUCHHOLTZ & DOHMEN 2015) für die jeweilige Lerngruppe funktional oder dysfunktional sein kann.

### 3.1. Didaktische Orientierung

Im Sinne der konstruktivistischen Lerntheorie (DUIT & TREAGUST 1998; RIEMEIER 2007) ist die Erhebung der Alltagsvorstellungen und Präkonzepte der SuS elementarer Ausgangspunkt für die Initiation von Lehr-Lern-Prozessen (ZABEL 2019). Insofern müssen themenbezogen typische Schülervorstellungen, das fachliche Vorwissen sowie mögliche Lernschwierigkeiten vorab mitgedacht werden (HAMMANN & ASSHOFF, 2014)

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
3.1.1. Lernpotential-Diagnose	Informationen zu den fachlichen Voraussetzungen der Lerngruppe (z. B. Vorwissen, Präkonzepte, Alltagsvorstellungen, Schwierigkeiten).	Es werden Stellen codiert, in denen auf fachliches Vorwissen, zu erwartende Lernschwierigkeiten oder auf Schülervorstellungen eingegangen wird.	"Dem Regenwurm sind die Schüler*innen <b>bereits</b> als Vertreter der Destruenten <b>begegnet</b> . (StEx_S1_E_NU_Ö_01)

			<p>Die Rolle von ATP in dem Querbrückenzyklus <b>wurde bisher</b> im Unterricht <b>nicht thematisiert.</b> (StEx_S2_K_P_04)</p> <p>da ich in den letzten Wochen feststellte, dass die Lernenden <b>Defizite bei der Formulierung</b> eines Fließtextes aufwiesen (StEx_S2_K_P_04)</p> <p>Durch Prominente wie Mark Benecke oder Krimi-Serien ist das <b>Berufsbild</b> eines Gerichtsmediziners <b>bekannt</b> (StEx_S2_K_P_04)</p>
3.1.2.	Methodische Vorkenntnisse und Fähigkeiten	Informationen zu den methodischen Voraussetzungen der Lerngruppe (Meyer 1994): Handlungsmuster (z. B. Protokollieren, Debattieren, Szenisches Spiel), Unterrichtsschritte (Einstieg, Erarbeitung, Sicherung, Transfer, Problemformulierung, Übungsphase), Sozialformen (EA, PA, GA).	<p>Es wird nur codiert, wenn explizit auf methodische Vorkenntnisse eingegangen wird.</p> <p>"Die SuS <b>können in 2er- und 3er-Gruppen arbeiten.</b>" (StEx_S1_E_NU_Ö_01)</p> <p>Die <b>Methode des Lerntempo-Duetts</b> wurde bisher ein Mal durchgeführt. (StEx_S1_E_NU_G_01)</p>

### 3.2. Pädagogische Orientierung

Neben den inhaltlichen Wissensbeständen der SuS müssen weitere Aspekte beachtet werden, die einen Einfluss auf Lehr-Lern-Prozesse haben und die daher in der Planung zu berücksichtigen sind (vgl. Angebot-Nutzungs-Modell; HELMKE 2015, S. 71). Zum einen können soziale Voraussetzungen relevant sein (z. B. Klassenklima,

Gruppendynamik), die durch didaktisch-methodische Entscheidungen beeinflusst werden können. Zum anderen sollte die Lernmotivation der SuS als „Tendenz [...], intellektuelle Betätigung als sinnvoll und lohnend zu erleben“ (BROPHY 1988, S.205) berücksichtigt werden.



Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
3.2.1. Soziale Voraussetzungen	Informationen zu den sozialen Voraussetzungen der Lerngruppe (z. B. Gruppenstruktur, Kooperationsfähigkeit, Teamfähigkeit, Konflikte, Lernatmosphäre, Arbeitsverhalten).	Es wird nur codiert, wenn explizit auf soziale Bedingungen eingegangen wird.	Die Klasse <b>neigt</b> in Gruppenphasen <b>zur Unruhe</b> und privaten Unterhaltungen, was einige SuS teilweise vom konzentrierten Arbeiten abhält. [...] Teilweise werden <b>Antworten ohne Meldung gerufen, auf</b> entsprechende <b>Hinweise reagieren</b> die SuS jedoch. (StEx_S1_B_Hd_Sex_01)  Das <b>Klassenklima ist</b> , wenn auch nicht auf den ersten Blick sichtbar, sehr <b>angespannt und belastet</b> . (StEx_S1_B_Hr_G_01)
3.2.2. Motivationale Voraussetzungen	Informationen zu den motivationalen Voraussetzungen der Lerngruppe (z. B. Motivation, Interesse, Mitarbeit).	Es wird nur codiert, wenn explizit auf motivationale Voraussetzungen eingegangen wird.	"Im letzten Unterrichtsbesuch vor dem Examen, [...], <b>arbeiteten</b> die Schüler*innen <b>besonders gut und schnell mit</b> . Es ist im Examen damit zu rechnen, dass sie wieder besondere Leistungen erbringen wollen und wegen der großen Zuschauerzahl und der besonderen Situation <b>sehr gut</b> und <b>motiviert</b> mitarbeiten werden. (StEx_S1_E_NU_Ö_01)  In der Klasse sind einige Schülerinnen und Schüler mit <b>großem Interesse am Fach Biologie</b> , [...]. (StEx_S1_B_Hr_G_01)  In dieser Gruppe handelt es sich um SuS, die <b>nicht sehr viel Interesse an biologischen Themen</b> und Aktivitäten zeigen und daher eine <b>sehr geringe Arbeitsmotivation</b> aufweisen (StEx_S1_B_Hr_G_01)

#### D.4. Wissen über fachspezifische Instruktionsstrategien

Wie POHLMANN (2019, S.228) richtigerweise feststellt, haben zahlreiche Studien zum PCK, die u.a. auch den theoretischen Rahmen dieser Untersuchung bilden (MAGNUSSON et al., 1999; PARK & OLIVER, 2008), das Wissen über Strategien zur Gestaltung wirksamer Lernumgebungen als eine der wichtigsten Facetten des PCK nachgewiesen. Hierunter werden alle Entscheidungen verstanden, die die Gestaltung der Lernumgebung und die Steuerung des Lernprozesses betreffen.

#### 4.1. Gestaltung der Lernumgebung

Alle Planungsentscheidungen statischen Charakters, die also im Vorhinein getroffen und in der Regel im Unterricht nicht verändert werden (Medieneinsatz, Sozialformen, Aufgabenstellungen, Differenzierungsmaßnahmen), werden als Gestaltung der Lernumgebung verstanden. Im engeren Sinne werden hierunter alle Entscheidungen verstanden, die zur Gestaltung des Lehr-Lern-Arrangements beitragen.

#### 4.1.1. Didaktische Strukturierung

Im Sinne des Modells der Didaktischen Rekonstruktion (KATTMANN 2019) werden als didaktische Strukturierung diejenigen Entscheidungen verstanden, mithilfe derer Lernwege ermöglicht werden, um von den Lernpotentialen der SuS zu den fachlich geklärten Begriffen zu gelangen.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
4.1.1.1 Beschreibung der didaktischen Strukturierung	Didaktisch-methodische Entscheidungen (Phasenstruktur, Methodik, Medien-einsatz, Gestaltung von Arbeitsmaterialien) werden beschrieben.	Es werden nur die didaktisch-methodischen Entscheidungen codiert, die den Lernprozess strukturieren und begleiten, nicht Aufgabenstellungen, Differenzierungsmaßnahmen, Impulse	Die Erarbeitung des Ablaufs der Vorgänge, welche zur Korallenbleiche und zum Tod der Korallenpolypen durch die Bleiche führen, erfolgt aufgabengeleitet und in Anlehnung an das <b>Think-Pair-Share-Format</b> . In einem <b>ersten Schritt</b> gilt es den Infotext in Einzelarbeit [...] zu lesen. <b>Im Anschluss</b> soll die konkrete zeitliche Abfolge der Teilschritte in Form eines [...] dargestellt werden. (StEx_S2_F_Ö_04)  Dabei dienen die <b>Moderationskarten</b> als Hilfestellung und Erinnerungstütze. (StEx_S1_B_Hr_G_01)  Während dieser Phase findet die Vorstellung der Arbeitsergebnisse aus den <b>Gruppenphase</b> statt und die Gruppen positionieren sich zu der Stundenfrage. (StEx_S1_B_Hr_G_01)  zwecks Transparenz wird die <b>Frage an der Tafel fixiert</b> . (StEx_S1_K_R_G_01)
4.1.1.2 Begründung der didaktischen Strukturierung	Begründung von didaktisch-methodischen Entscheidungen zur Gestaltung der Lernumgebung unter Einbeziehung möglicher Alternativen.	Es wird nur codiert, wenn didaktisch-methodische Entscheidungen in ihrer Funktionalität für die angestrebte Kompetenzentwicklung begründet werden. Auch codiert werden vorgeschlagene Alternativen Nicht codiert werden Begründungen, die sich aus dem Wissen über die Schülerkognition (Facette 3) ergeben.	Aus diesem Grund wird in dieser Unterrichtsstunde auf die Formulierung von Hypothesen verzichtet und der Kompetenzschwerpunkt auf [...] Textproduktion, gelegt. (StEx_S2_K_P_04)  Ich habe mich für diese kooperative Sozialform entschieden, um den Schreibprozess zu entlasten. (StEx_S2_K_P_04)  <b>zwecks Transparenz</b> wird die Frage an der Tafel fixiert. (StEx_S1_K_R_G_01)  Das <b>Tafelbild stellt</b> zum einen die Stundenfrage <b>vor</b> zum anderen <b>soll</b> es die SuS zu einer Positionierung <b>anregen</b> . (StEx_S1_B_Hr_G_01)

**Sollten SuS Schwierigkeiten haben**, sich darunter etwas vorzustellen [, dann...] (StEx\_S1\_K\_R\_G\_01)

**Sollte es** den Schülerinnen und Schülern dennoch **schwerfallen**, eine Vermutung zu äußern [, dann...] (StEx\_S2\_K\_Ö\_04)

#### 4.1.2. Aufgabenstellungen

Ganz gleich ob Lernaufgaben (HÖBLE & JAHNKE, 2010), Leistungsaufgaben (HARMS 2019) oder Diagnoseaufgaben (FRIEDRICH, FELLER, JOOS & SPÖRHASE-EICHMANN 2016) – da Aufgabenstellungen, aus Schülerperspektive betrachtet, der Kern der Unterrichtsstunde („Erarbeitungsphase“) betreffen, sollten sie und ihr Potential, als „Aufforderung zum Lern-Handeln“ (BRUDER 2008) wirken und somit kognitiv, affektiv und psychomotorisch aktivieren (ANDERSON & KRATWOHL 2001), beschrieben und ihr Einsatz sowie ihre Ausgestaltung begründet werden. Als „Aufgabenstellung“ werden im Rahmen dieser Studie handlungsinitiierende Arbeitsaufträge in der Erarbeitungsphase definiert, die oft auf einem Arbeitsblatt für die SuS fixiert sind.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
4.1.2.1 Beschreibung von Aufgabenstellungen	Beschreibung der Funktion der Aufgabenstellung für den Lernprozess	Es wird nur codiert, wenn die Aufgabenstellung als solche beschrieben wird, nicht der Einsatz (vgl. König, Bucholtz & Dohmen 2015)	Die <b>erste Aufgabe</b> entspricht dem ersten Schritt (Planung) der von Leisen vorgeschlagenen drei Schritte zur Forderung der Schreibkompetenz, denn dabei werden bereits Fachbegriffe für den Fließtext gesammelt und zugeordnet. (StEx_S2_K_P_04)  Die <b>Aufgaben</b> sind progressiv aufgebaut. (StEx_S1_K_R_G_01)  Zudem wird eine <b>Zusatzaufgabe für schnelle und leistungsstarke</b> SuS angeboten. Hier sollen die gewonnenen Einsichten in die Komplexität des Trüffelbaus grafisch veranschaulicht werden (Wechsel der Darstellungsform). (StEx_S2_F_Ö_05)
4.1.2.2 Begründung von Aufgabenstellungen	Begründung des Einsatzes der Aufgabenstellung für den Lernprozess, ggf. Bezugnahme auf kognitives, affektives oder	Es wird nur codiert, wenn Gründe für den Einsatz bzw. die Gestaltung der Aufgabenstellung genannt werden. (-> Analyse der Lehr-Lern-Struktur) Auch codiert werden Alternativen, die	Die Lernenden sollen <b>mit dieser Aufgabe</b> und den folgenden Aufgaben auch <b>erkennen</b> , dass der Vaterschaftsnachweis mit Hilfe der Blutgruppen sehr fehlerbehaftet ist und nicht mit 100 Sicherheit als

	psychomotorisches Aktivierungspotential	vorgeschlagen werden, so wie die Gründe dafür, warum man sich dagegen entschieden hat.	Vaterschaftsnachweis dienen kann. (StEx_S1_F_BKS_G_01)  Das Fazit in <b>Aufgabe 9</b> soll das Bild der Stunde <b>abrunden</b> (StEx_S2_F_Ö_05)
--	-----------------------------------------	----------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------

#### 4.1.3. Binnendifferenzierung

Unter Binnendifferenzierung (oder auch innerer Differenzierung) werden Bemühungen verstanden, den heterogenen Lernvoraussetzungen der SuS (Kognition, Vorwissen, Emotionalität, Lerntempo, Sozialverhalten usw.) durch didaktische Maßnahmen zu begegnen und allen SuS einen Kompetenzzuwachs zu ermöglichen, z. B. durch Ermöglichung verschiedener Lernzugänge, durch Lernhilfen, durch verschiedene Schwierigkeitsniveaus und Zusatzaufgaben oder durch die leistungshomogene bzw. –heterogene Zusammenstellung von Arbeitsgruppen (KRÜGER & MEYFARTH, 2009).

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
4.1.3.1. Beschreibung von Differenzierungsmaßnahmen	Die ergriffenen Differenzierungsmaßnahmen (horizontal: Differenzierung nach Interesse, Inhalt, sozialen Gruppen, Lernstil, Lernmethode; vertikal: Differenzierung nach Leistungsniveau) werden beschrieben.	Es wird nur codiert, wenn die Differenzierungsmaßnahme als solche beschrieben wird, nicht der Einsatz. (-> Lerngruppenanalyse, Analyse der Lehr-Lern-Struktur)	SuS, die <b>sehr schnell</b> zu einem Ergebnis in ihrer Gruppe kommen <b>erhalten ein weiteres Fallbeispiel</b> . (StEx_S1_B_Hd_Sex_01)  Dem versuche ich <b>mit diversen Differenzierungsmaßnahmen</b> gerecht zu werden. Dabei haben sich besonders die <b>Vorgabe unsortierter Stichworte</b> , aber auch Hilfen in Form von <b>Leitfragen</b> , welche die Bearbeitung des Arbeitsauftrags kleinschnittig anleiten, bewährt. Diese werden auch in dieser Stunde bereitgehalten. (StEx_S2_F_Ö_04)
4.1.3.2. Begründung von Differenzierungsmaßnahmen	Begründung des Einsatzes ausgewählter Differenzierungsmaßnahmen	Es wird nur codiert, wenn explizit auf die Lerngruppe bzw. einzelne SuS Bezug genommen wird (-> Lerngruppenanalyse, Analyse der Lehr-Lern-Struktur). Auch codiert werden Alternativen, die vorgeschlagen werden, so wie die Gründe dafür, warum man sich dagegen entschieden hat.	Die nummerierten Leitfragen <b>helfen</b> Olivia <b>dabei, konzentriert</b> und vor allen Dingen <b>im eigenen Tempo</b> Schrittfür Schritt die im Text dargestellte Abfolge <b>nachzuvollziehen</b> . (StEx_S2_F_Ö_04)  Dieser Leitfaden ist nicht starr und kann von den Lernenden so angepasst werden, dass sie ihre <b>eigene Schreibstrategie entwickeln</b> (Reflexion der eigenen Schreibstrategie). (StEx_S2_K_P_04)

## 4.2. Unterrichtssteuerung

Alle Planungsentscheidungen dynamischen Charakters, die also zwar im Vorhinein getroffen werden, sich aber auf konkrete Schüler-Lehrer-Interaktionen beziehen, die nicht im Vorhinein planbar sind, werden als Unterrichtssteuerung verstanden. Im engeren Sinne werden hierunter alle Entscheidungen verstanden, die zur Gestaltung des Lehr-Lern-Arrangements beitragen. Während sich laut LEISEN (2014) klassische

### 4.2.1. Impulse

Während Fragestellungen (*Wer kann etwas zu dem Bild sagen?*) oft geschlossenen Charakter haben, von den SuS nur wenige Worte als Reaktion auslösen und eine starke Lenkung durch die Lehrkraft erfordern, kann mithilfe von Impulsen (*Beim Betrachten des Bildes gebt jedem von euch eine Menge durch den Kopf. Schildert bitte eure Assoziationen.*) eine größere Gruppe von SuS angesprochen werden, die ihre vielfältigen Ideen in einem diskursiven, multiperspektivischen Gespräch einbringen können, so dass der Redeanteil der Lehrkraft im Falle gelungener Gesprächsführung zugunsten der SuS reduziert werden kann.

Gesprächsführung vor allem auf den Lerngegenstand kapriziert, geht es bei einem gelingenden Unterrichtsgespräch im diskursiven Sinne um den Austausch von Ideen, die Diagnose von Vorstellungen, die Reflexion von Lernprodukten, die Einordnung von Missverständnissen usw., was durch planbare Moderationstechniken zu ermöglichen sei.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
4.2.1.1. Beschreibung von Impulsen	Es werden operationalisierte Impulse formuliert, die in jeder Phase und ggf. in den Phasenübergängen den Unterrichtsgang zielgerichtet auf den ausgewählten Kompetenzzuwachs hin steuern.	Es wird codiert, wenn Impulse explizit ausformuliert werden. (--> Stundenverlaufsplanung) oder paraphrastisch beschrieben werden (--> Analyse der Lehr-Lern-Struktur)	<p>Anschließend werden sie <b>gebeten</b> Vermutungen zu äußern, wie sich der Regenwurm fortbewegt. (StEx_S1_E_NU_Ö_01)</p> <p><b>Erklären</b> Sie den vorgefundenen Zustand der Leiche. (StEx_S2_K_P_04)</p> <p>Durch den <b>Impuls</b> „<b>Was könnte die Frage von Manuel und Lisa sein?</b>“ werden die SuS schnell die naheliegende Problemfrage formulieren können: (StEx_S1_K_R_G_01)</p> <p><b>Für den Einstieg</b> wird eine <b>Aussage</b> gewählt, die zunächst keinen Sinnzusammenhang besitzt. Es handelt sich hier um einen sogenannten <b>stillen Impuls</b>. (StEx_S2_K_Ö_04)</p>

4.2.1.2. Begründung von Impulsen	Begründung des Einsatzes von Impulsen unter Berücksichtigung der damit verbundenen Steuerungsfunktion	Es wird nur codiert, wenn der Zweck eines Impulses in seiner Funktionalität für die angestrebte Kompetenzentwicklung begründet wird. Auch codiert werden Alternativen, die vorgeschlagen werden, so wie die Gründe dafür, warum man sich dagegen entschieden hat.	<p><b>Durch den Impuls</b> „Was könnte die Frage von Manuel und Lisa sein?“ <b>werden die SuS</b> schnell die naheliegende <b>Problemfrage formulieren können:</b> (StEx_S1_K_R_G_01)</p> <p><b>Die Steuerung</b> durch die Lehrkraft <b>hilft</b> der z.T. unruhigen Gruppe sich auf den Lerngegenstand zu <b>fokussieren</b> — Klarheit und Übersichtlichkeit werden als Mittel für den Umgang mit der offenen Gesprächskultur eingesetzt. (StEx_S1_B_Hr_G_01)</p>
----------------------------------	-------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

#### 4.2.2. Antizipiertes Schülerverhalten

Eine wichtige Fähigkeit von Lehrkräften ist es, im Vorhinein begründete Vermutungen darüber anzustellen, welche Auswirkungen einzelne Planungsentscheidungen in der Unterrichtspraxis haben könnten und wie sie sich auf den Lernprozess der SuS auswirken. In dieser Kategorie werden nur diejenigen Antizipationen verstanden, die sich auf die Kontingenz unterrichtlichen Handelns beziehen, d. h. nicht auf die vorab festlegbaren und bestimmbareren Unterrichtssituationen (z. B. SuS nennen die drei Mendel'schen Regeln; SuS lesen AB), sondern auf die offenen Situationen, in denen ein Spektrum an möglichen Antworten der SuS erwartbar ist (z. B. als Reaktion auf einen offenen Impuls).

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
4.2.2. Antizipiertes Schülerverhalten	Das Verhalten der SuS und konkrete Äußerungen als verbale Reaktion auf Impulse der Lehrkraft bzw. als Beiträge von Unterrichtsgesprächen werden antizipiert. Auch codiert werden antizipierte Arbeitsergebnisse, Vermutungen über den Verlauf von Unterrichtsgesprächen und möglicherweise entstehende Schwierigkeiten	Es wird nur codiert, wenn die Antizipation über reine Deskription äußerlich wahrnehmbaren Verhaltens ("SuS hören zu", "SuS füllen AB aus") hinausgeht. Nicht codiert werden Musterlösungen.	<p>Die Einigung innerhalb der Gruppen <b>könnte Schwierigkeiten bergen</b>, da die SuS die Argumente evtl. sehr unterschiedlich bewerten. (StEx_S1_B_Hd_Sex_01)</p> <p>Ich <b>nehme an</b>, die Mehrzahl der <b>SuS würden</b> spontan einen rezessiven Erbgang <b>vermuten</b>. (StEx_S1_K_R_G_01)</p> <p>Obwohl für Aufgabe 5 TIPP's auf den Tischen bereitliegen, <b>könnten manche SuS</b> dennoch <b>Hemmungen haben</b>, diese zu nutzen. Darauf werde ich mein Augenmerk lenken und sie ggf. ermuntern. (StEx_S1_K_R_G_01)</p>

## D.5. Wissen über Diagnostik

Spätestens die moderne Feedbackforschung hat nachgewiesen, dass die Ein-beziehung von Feedbackelementen einen signifikanten Einfluss auf den Lern-erfolg haben kann (HATTIE, 2015; HATTIE & TIMPERLEY 2007). Die im Unterricht entwickelten Lernprodukte, gezeigte Lernfortschritte oder auftretende Lern-schwierigkeiten können und sollten berücksichtigt werden, unter anderem um Entwicklungspotentiale transparent zu machen und Lernerfolge rückzumelden (WEINGARTEN 2019).

## 5.1. Schülerfeedback

Als Schülerfeedback werden alle Elemente einer Unterrichtsplanung verstanden, in denen die SuS entweder einander bzw. der Lehrkraft eine Rückmeldung zum Lernprozess, zu Lernprodukten oder Lernschwierigkeiten geben, wobei feed-up (Was ist mein/unser Ziel?), feed-back (Wie geht es voran?) und feed-forward (Was kommt als Nächstes?) inkludiert sind (HATTIE & TIMPERLEY 2007). Oft werden Phasen, in denen sich die SuS untereinander austauschen oder korrigieren sollen geplant, ohne dass dies als „Schülerfeedback“ bezeichnet wird.

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
5.1.1. Beschreibung des Schülerfeedbacks	Das in der geplanten Lernumgebung vorgesehene Feedback der SuS zur Aufgabenstellung, zum Lernprozess und/oder zur Selbstregulation wird beschrieben, wobei feed-up (Was ist mein/unser Ziel?), feed-back (Wie geht es voran?) und feed-forward (Was kommt als Nächstes?) inkludiert sind.	Es wird nur codiert, wenn die Art und Durchführungsweise des Feedbacks beschrieben wird.	Die anderen Gruppen können die Präsentationen mit ihren <b>Ergebnissen</b> und somit durch neue oder alternative Argumente <b>ergänzen</b> . (StEx_S1_B_Hd_Sex_01)  [...] und damit sich in der Sicherung jeweils vier Lernende einer Tischreihe gegenseitig eine <b>schriftliche Rückmeldung</b> geben können. (StEx_S2_K_P_04)  Die restlichen SuS erhalten einen <b>Beurteilungsbogen</b> mit dem sie die Planung <b>einschätzen und reflektieren</b> sollen. (StEx_S2_E_Ö_05)  Außerdem sollen sie <b>schauen ob</b> dieser Weg ihrem <b>eigenen entspricht</b> und wenn nicht, ob dieser dennoch nachvollziehbar ist (StEx_S2_K_Ö_04)

5.1.2. Begründung des Schülerfeedbacks	Die Berücksichtigung von Schülerfeedback zur Aufgabenstellung, zum Lernprozess oder zur Selbstregulation wird hinsichtlich ihres Zwecks/Nutzens begründet.	Es wird codiert, wenn der Zweck des Schülerfeedbacks beschrieben wird. Auch codiert werden Alternativen, die vorgeschlagen werden, so wie die Gründe dafür, warum man sich dagegen entschieden hat.	Das kurze Feedback <b>regt</b> die SuS <b>dazu an</b> über das Erreichen der Lernziele <b>zu reflektieren</b> und sich <b>selbst einzuschätzen</b> . Es <b>dient</b> aber in dieser Form <b>vor allem der Lehrkraft als Rückmeldung</b> wie die Schüler das Erlernte einschätzen, um weiteren <b>Übungsbedarf abzuschätzen</b> . (StEx_S1_B_Hr_G_01)  <b>Damit alle</b> Lernprodukte <b>gewürdigt</b> werden, (StEx_S2_E_Ö_05)
----------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## 5.2. Lehrerfeedback

Als Lehrerfeedback werden alle Elemente einer Unterrichtsplanung verstanden, in denen die Lehrkraft den SuS eine Rückmeldung zum Lernprozess, zu Lernprodukten oder Lernschwierigkeiten gibt. Dieses Feedback kann auf drei Ebenen stattfinden (HATTIE 2015), die in dieser Kategorie allesamt gemeint sind:

1. FB zur Aufgabe (*Haben die SuS die Aufgabe korrekt gelöst?*)
2. FB zum Lösungsprozess (*War der Lösungsweg der SuS geeignet?*)
3. FB zur Selbstregulation (*Inwiefern haben sich die SuS beim Lernen unterstützt?*)

Subkategorie	Beschreibung	Codierhinweise	Ankerbeispiel
5.2.1. Beschreibung des Lehrerfeedbacks	Das in der geplanten Lernumgebung vorgesehene Feedback der Lehrkraft zum Lernprozess bzw. Lernprodukten wird beschrieben.	Es wird nur codiert, wenn die Art und Durchführungsweise des Feedbacks beschrieben wird.	Anhand eines <b>Lösungswortes</b> können die Lernenden sich <b>selbst überprüfen</b> . (StEx_S2_K_P_04)  [...] werde ich <b>Musterlösungen</b> ausgeben. (StEx_S1_K_R_G_01)
5.2.2. Begründung des Lehrerfeedbacks	Die Berücksichtigung von Lehrerfeedback zur Aufgabenstellung, zum Lernprozess der SuS oder zur Selbstregulation der SuS wird hinsichtlich ihres Zwecks/Nutzens begründet.	Es wird codiert, wenn der Zweck des Lehrerfeedbacks beschrieben wird. Auch codiert werden Alternativen, die vorgeschlagen werden, so wie die Gründe dafür, warum man sich dagegen entschieden hat.	Es besteht natürlich die Gefahr, dass die Lernenden die Zuordnung der Stichworte nur anhand des Lösungswortes durchführen, aber diese <b>Zuordnung</b> müssen sie <b>später zur Beschreibung</b> der Vorgänge bei der intakten Muskelkontraktion zur



Formulierung des Gutachtens  
**überprüfen.** (StEx\_S2\_K\_P\_04)

Die **Selbstkontrolle dient dazu** nicht vorhandenes oder fehlerhaft erinnertes Wissen **zu korrigieren**, und nebenbei die **Selbständigkeit** der SuS durch Eigenverantwortung im Kontrollprozess **zu stärken.** (StEx\_S1\_B\_Hr\_G\_01)

**Da** eine schriftliche Sicherung **zu lange dauern würde** (StEx\_S1\_K\_R\_G\_01)

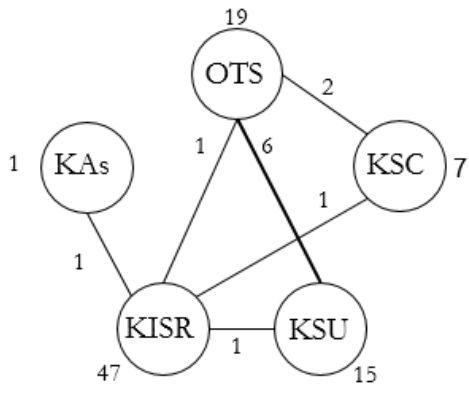
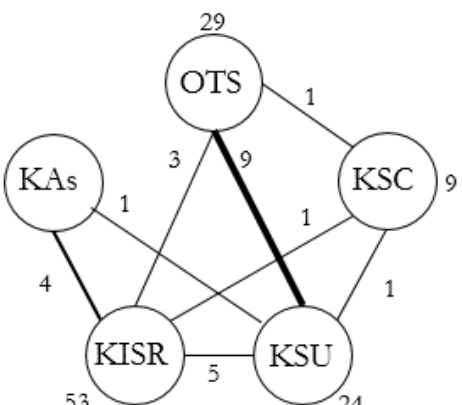
## E. Quellen

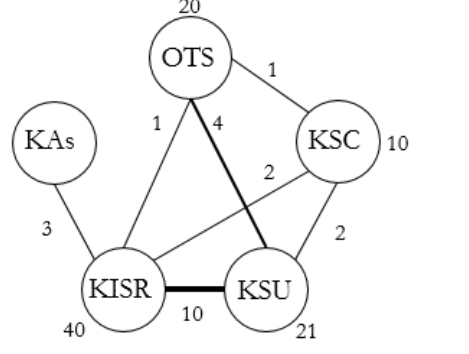
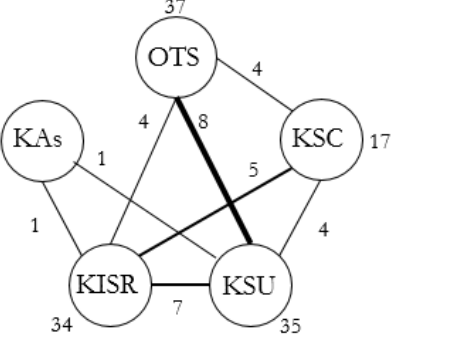
- ANDERSON, L.W. & KRATHWOHL, D. (2001): A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. New York.
- BAUMERT, J., & KUNTER, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. Zeitschrift für Erziehungswissenschaft, 9(4), S.469-520.
- BROPHY, J.E. (1988): On motivating students. In: Berliner, D. & Rosenshine, B. (Hrsg.): Talks to teachers, New York, S.201-245.
- BRUDER, R. (2008): Vielseitig mit Aufgaben arbeiten. Mathematische Kompetenzen nachhaltig entwickeln und sichern. In R. BRUDER, T. LEUDERS, & A. BÜCHTER (Hrsg.), Mathematikunterricht entwickeln. Bausteine für kompetenzorientiertes Unterrichten, Berlin, S.18-52.
- CARLSON, J., & DAEHLER, K. R. (2019): The refined consensus model of pedagogical content knowledge in science education. In: Hume, A., Cooper, R., & Borowski, A. (Hrsg.): Repositioning pedagogical content knowledge in teachers' knowledge for teaching science, Singapore, S.77-92.
- DUIT, R., & TREAGUST, D. F. (1998). Learning in science - From behaviourism towards social constructivism and beyond. In: B. J. FRASER & K. TOBIN (Hrsg.), International handbook of Science Education, Part 1, Dordrecht, S. 3-25.
- FREIMAN, T. (2001): Kumulatives Lernen im Biologieunterricht, in; Praxis der Naturwissenschaften – Zeitschrift für den experimentellen Unterricht, Nr.50/7, S.1-2.
- FRIEDRICH, I., FELLER, W., JOOS, A. & SPÖRHASE-EICHMANN, U. (2016): Diagnosekompetenz erwerben. Lernaufgaben analysieren und konstruieren, in: Unterricht Biologie, Nr.417, S.8-14.
- GESS-NEWSOME, J. & LEDERMAN N.G. (Hrsg.) (1999): Examining pedagogical content knowledge. The Construct and its Implications for Science Education, Dordrecht.
- GROPENGIEßER, H. & KATTMANN, U. (2019): Didaktische Rekonstruktion. In: GROPENGIEßER, H., HARMS, U. & KATTMANN, U.: Fachdidaktik Biologie, 11. Auflage, Hallbergmoos 2019, S.16-23.
- HAMMANN, M., & ASSHOFF, R. (2014): Schülervorstellungen im Biologieunterricht. Ursachen für Lernschwierigkeiten, Stuttgart.
- HARMS, U. (2019): Didaktische Rekonstruktion. In: GROPENGIEßER, H., HARMS, U. & KATTMANN, U.: Fachdidaktik Biologie, 11. Auflage, Hallbergmoos 2019, S.251-265.
- HATTIE, J., & TIMPERLEY, H. (2007). The power of feedback. *Review of educational research*, 77(1), S.81-112.
- HATTIE, John A. C. (2015): Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von „Visible learning“, besorgt von Wolfgang Beywl und Klaus Zierer (3. Aufl.), Hohengehren.
- HELMKE, A. (2015). Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts, (6. aktual. Aufl). Seelze.

- HÖBLE, C. & JAHNKE, L. (2010): Gute Lernaufgaben für den Biunterricht? – Eine große Herausforderung, in: KIPER, H., MEINTS, W., PETERS, S., SCHLUMP, S. & SCHMIT, S. (Hrsg.): Lernaufgaben und Lernmaterialien im kompetenzorientierten Unterricht, Stuttgart.
- KATTMANN, U. (2019): Auswahl und Verknüpfung der Lerninhalte. In: GROPEGIEBER, H., HARMS, U. & KATTMANN, U.: Fachdidaktik Biologie, 11. Auflage, Hallbergmoos, S. 29-39.
- KATTMANN, U. (2019): Brückenfach Biologie. In: GROPEGIEBER, H., HARMS, U. & KATTMANN, U.: Fachdidaktik Biologie, 11. Auflage, Hallbergmoos, S. 143-148.
- KMK (2005): Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss, Beschluss der Kultusministerkonferenz vom 16.12.2004, München. Zugriff am 21.12.2019 unter: [https://www.kmk.org/fi\\_leadadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Bildungsstandards-Biologie.pdf](https://www.kmk.org/fi_leadadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf).
- KÖNIG, J., BUCHHOLTZ, C., & DOHMEN, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18(2), S.375-404.
- LEISEN, J. (2014): Wie soll ich denn meinen Unterricht planen? Lehr-Lern-Prozesse planen am Beispiel Elektrizitätslehre in Physik. In: MAIER, U. (Hrsg.): Lehr-Lernprozesse in der Schule: Referendariat. Praxiswissen für den Vorbereitungsdienst, Bad Heilbrunn, S.102-117.
- LEISEN, J. (2015): Lernumgebung und Lernschritte durch Moderation steuern. Wie man anders im Unterricht kommunizieren kann. In: *Pädagogik* 15 (11), S. 14-17.
- LOUGHRAN, J., KEAST, S., & COOPER, R. (2016). Pedagogical reasoning in teacher education. In: LOUGHRAN, J. & HAMILTON, M.L. (Hrsg.): *International handbook of teacher education*, Vol.1, Singapore, S.387-421.
- MAGNUSSON, S., KRAJCIK, J. & BORKO, H. (1999): Nature, sources, and development of pedagogical content knowledge for science teaching. In: GESS-NEWSOME, J. & LEDERMAN N.G. (Hrsg.): *Examining pedagogical content knowledge. The Construct and its Implications for Science Education*, Dordrecht, S.95-132.
- MAIER, U. (Hrsg.) (2012): *Lehr-Lernprozesse in der Schule: Referendariat. Praxiswissen für den Vorbereitungsdienst*, Bad Heilbrunn.
- MAYER, J. (2019): Unterrichtsziele formulieren. In: GROPEGIEBER, H., HARMS, U. & KATTMANN, U. (Hrsg.): *Fachdidaktik Biologie*, 11. Auflage, Hallbergmoos, S. 220-227.
- PARK, S. (2019): Reconciliation Between the Refined Consensus Model of PCK and Extant PCK Models for Advancing PCK Research in Science. In: Hume, A., Cooper, R., & Borowski, A. (Hrsg.): *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*, Singapore, S.117-128.
- PARK, S., & OLIVER, J. S. (2008). National board certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45(7), S.812–834.
- POHLMANN, M. (2019): Förderung ethischer Bewertungskompetenz: der Einfluss ausgewählter Lerngelegenheiten auf die inhaltliche Ausdifferenzierung und die Kohärenz der Komponenten des fachdidaktischen Wissens von Biologielehrkräften, Oldenburg.

- RIEMEIER, T. (2007): Moderater Konstruktivismus. In: KRÜGER, D., VOGT, H. (Hrsg.): Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden. S. 69-79.
- ROSENBERGER, K. (2018): Unterrichten: Handeln in kontingenten Situationen, Weinheim.
- RÖSSLER, P. (2005): Inhaltsanalyse, Stuttgart.
- (SENBJF 2014a) SENATSVERWALTUNG FÜR BILDUNG, JUGEND UND FAMILIE (Hrsg.) (2014): Handbuch für den Vorbereitungsdienst, Berlin.
- (SENBJF 2014b) SENATSVERWALTUNG FÜR BILDUNG, JUGEND UND FAMILIE (Hrsg.) (2014): Verordnung über den Vorbereitungsdienst und die Staatsprüfung für Lehrämter, Berlin.
- (SENBJF 2015a) SENATSVERWALTUNG FÜR BILDUNG, JUGEND UND FAMILIE (Hrsg.) (2015): Rahmenlehrplan für die Sekundarstufe I (Teil B) – Fächerübergreifende Kompetenzentwicklung, Berlin.
- (SENBJF 2015b) SENATSVERWALTUNG FÜR BILDUNG, JUGEND UND FAMILIE (Hrsg.) (2015): Rahmenlehrplan für die Sekundarstufe I (Teil C) – Biologie, Berlin.
- (SENBS) SENATSVERWALTUNG FÜR BILDUNG, JUGEND UND SPORT (Hrsg.) (2006): Rahmenlehrplan die gymnasiale Oberstufe – Biologie, Berlin.
- STAECK, L. (2010): Zeitgemäßer Biologieunterricht. Eine Didaktik für die Neue Schulbiologie, Baltmannsweiler.
- SUH, J. K., & PARK, S. (2017). Exploring the relationship between pedagogical content knowledge (PCK) and sustainability of an innovative science teaching approach. *Teaching and Teacher Education*, 64, S.246-259.
- STRELLER, S., BOLTE, C., DIETZ, D., NOTO LA DIEGA, R. (2019): *Chemiedidaktik an Fallbeispielen. Anregungen für die Unterrichtspraxis*, Berlin/Heidelberg.
- WEINERT, F. (2001): *Leistungsmessung an Schulen*, Weinheim.
- WEINGARTEN, J. (2019): *Wie planen angehende Lehrkräfte ihren Unterricht? Empirische Analysen zur kompetenzorientierten Gestaltung von Lernangeboten*, Münster/New York.
- WERNKE, S., ZIERER, K. (2017): *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung*, Bad Heilbrunn.
- ZABEL, J. (2019): Erkenntnistheorie und Lernen. In: GROPPENGIEBER, H., HARMS, U. & KATTMANN, U.: *Fachdidaktik Biologie*, 11. Auflage, Hallbergmoos, S. 98-105.

**Anhang 2: PCK maps der Staatsexamensentwürfe (Beiträge 2 und 3)**

StEx_S2_B_Ö_01	StEx_S2_B_Ö_04
	
<b>Note des Fachseminarleiters:</b> 1	<b>Note des Fachseminarleiters:</b> 1
<b>Begründung:</b> alle Planungsentscheidungen sinnvoll und sehr gut begründet; sehr stimmig	<b>Begründung:</b> alle wichtigen Aspekte, etwas unübersichtlich, aber sehr klar

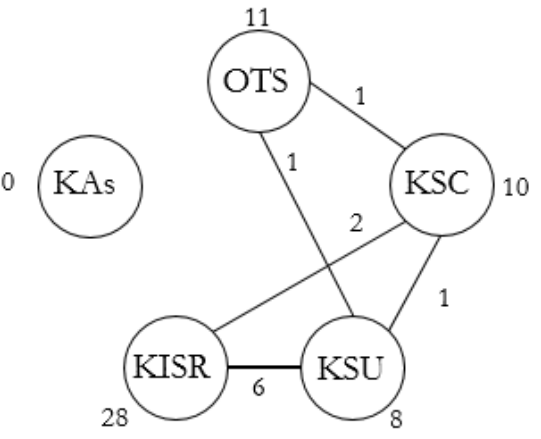
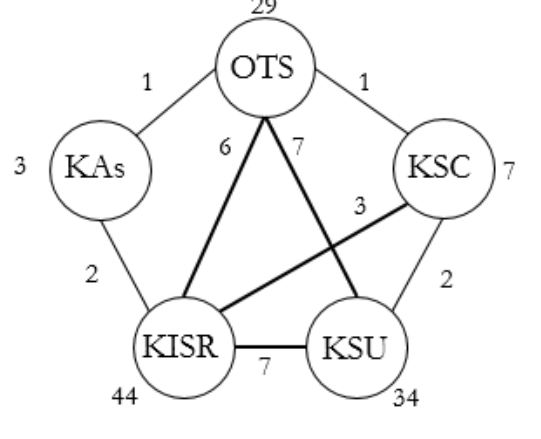
StEx_S2_E_E_01	StEx_S2_F_Ö_04
	
<b>Note des Fachseminarleiters:</b> 1	<b>Note des Fachseminarleiters:</b> 1
<b>Begründung:</b> eher 1- wegen zu vieler konkretisierter Standards; Stand der Kompetenzentwicklung soll sein, was SuS vor der Stunde schon können (--> hier: Präjudizierung "Der schafft sowieso xy nicht."); Standards aus RLP sollen am Ende der Doppeljahrgangsstufe erreicht werden, nicht am Ende einer 45-Minuten-Stunde (--> Standard in Einzelschritte zerlegen (muss ins SchiC)	<b>Begründung:</b> Reihenplanung sehr gut (immer mit erwähnt, was die SuS konkret gemacht haben); langfristige Reihenplanung in Worten begründet; Indikatoren sehr wichtig; Lerngruppenanalyse schlägt sich in meht.-did. Analyse wieder; Fachliche Klärung fokussiert das, was vermittelt werden soll; was ist davon für SuS wichtig, wie bringe es näher?

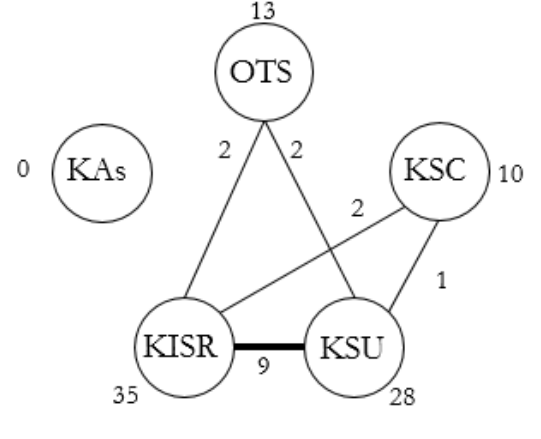
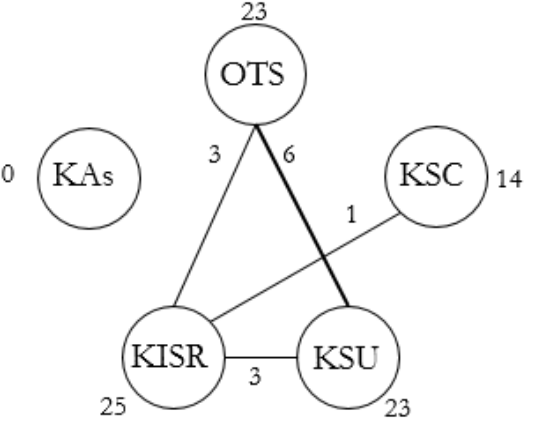
StEx_S2_K_Ö_04	StEx_S2_B_Ö_06
<p>Network diagram for StEx_S2_K_Ö_04. Nodes: KAs (1), KISR (44), KSU (31), KSC (13), OTS (13). Connections: KAs-KISR (1), KISR-KSU (6), KISR-KSC (2), KISR-OTS (3), KSU-KSC (4), KSU-OTS (5), KSC-OTS (2).</p>	<p>Network diagram for StEx_S2_B_Ö_06. Nodes: KAs (0), KISR (28), KSU (24), KSC (14), OTS (30). Connections: KAs-KISR (0), KISR-KSU (9), KISR-KSC (1), KISR-OTS (7), KSU-KSC (5), KSU-OTS (11), KSC-OTS (2).</p>
<b>Note des Fachseminarleiters:</b> 1	<b>Note des Fachseminarleiters:</b> 1-
<b>Begründung:</b> sehr gute Entscheidungen, oft schwache Begründungen, sprachlich sehr ungenau, langfristiges Konzept zur Kompetenzförderung fehlt	<b>Begründung:</b> sehr hohes inhaltliches Niveau, passt gut zur Lerngruppe (keine riesige Heterogenität); ausführliches Konzept zur Kompetenzentwicklung, wenig Lerngruppenanalyse, aber war auch nicht mehr nötig

StEx_S2_E_Ö_04	StEx_S2_E_Ö_09
<p>Network diagram for StEx_S2_E_Ö_04. Nodes: KAs (3), KISR (35), KSU (44), KSC (24), OTS (32). Connections: KAs-KISR (1), KISR-KSU (9), KISR-KSC (4), KISR-OTS (3), KSU-KSC (4), KSU-OTS (2), KSC-OTS (4).</p>	<p>Network diagram for StEx_S2_E_Ö_09. Nodes: KAs (2), KISR (47), KSU (15), KSC (7), OTS (19). Connections: KAs-KISR (2), KISR-KSU (3), KISR-KSC (1), KISR-OTS (1), KSU-KSC (1), KSU-OTS (2), KSC-OTS (1).</p>
<b>Note des Fachseminarleiters:</b> 1-	<b>Note des Fachseminarleiters:</b> 1-
<b>Begründung:</b> methodischer Teil: Aufgabenanalyse, Materialien hervorragend; Antizipation von Schwierigkeiten sehr gut, didaktischer Teil eher wie ausgefülltes Formular; kein langfristiges Konzept; Schwerpunkt der Stunde sinnvoll gewählt; keine Diskrepanz zwischen Stand der Entwicklung und dem Standard (-> die können das also offenbar schon)	<b>Begründung:</b> alles sehr gut, außer Sicherung; zu wenig auf Kompetenzschwerpunkt bezogen

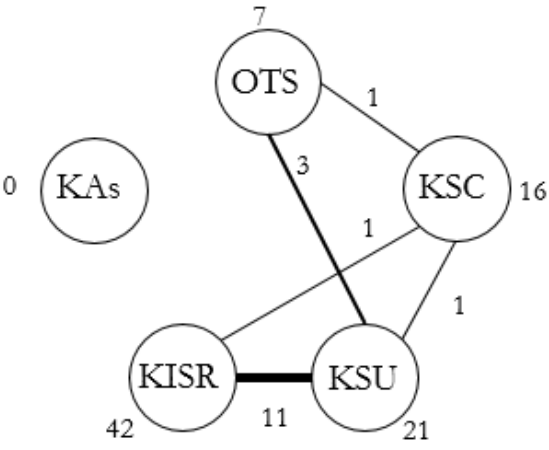
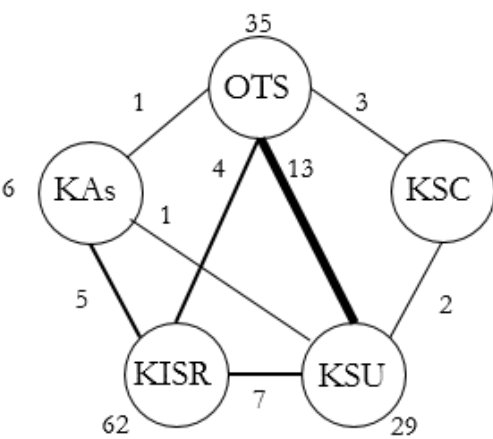
StEx_S2_K_P_08	StEx_S2_K_P_09
<p>Note des Fachseminarleiters: 1-</p>	<p>Note des Fachseminarleiters: 1-</p>
<p><b>Begründung:</b> viele Dinge, die bei anderen fehlen; sehr lang, fundierte Begründungen; zielgerichtet auf Standard; einzige Einschränkung: mehrere Standards</p>	<p><b>Begründung:</b> did. Begründungen nicht so gut --&gt; gerade bei so komplexen Thema wäre didaktische Reduktion wichtig; Sicherung: Kompetenzschwerpunkt stärker miteinbeziehen; Niveaustufen bei Stand der Kompetenzentwicklung stehen oft nur die Ziele, aber nicht die Ausgangslage; langfristige Kompetenzentwicklung nicht sehr überzeugend; INSGESAMT Entwurf zu gut, um schlechter als 2+ zu geben</p>

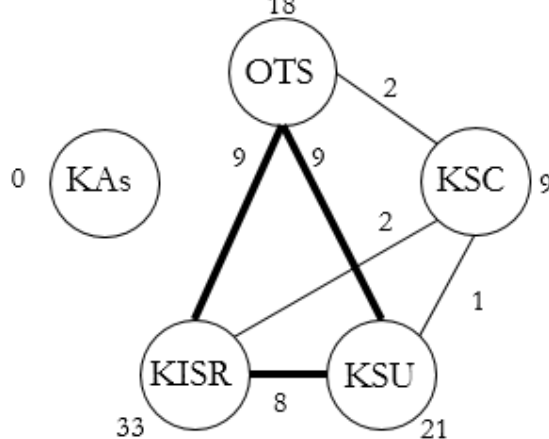
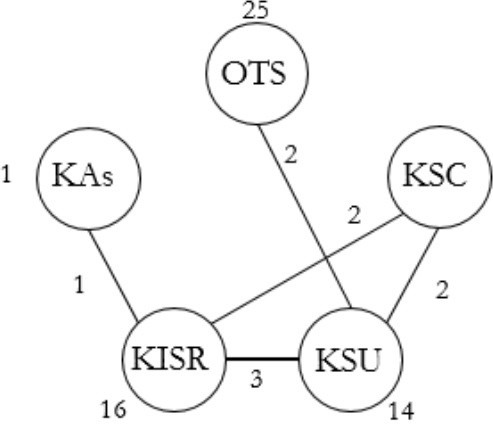
StEx_S2_B_Ö_03	StEx_S1_E_NU_Ö_01
<p>Note des Fachseminarleiters: 2+</p>	<p>Note des Fachseminarleiters: 2</p>
<p><b>Begründung:</b> methodischer Teil: Aufgabenanalyse, Materialien sehr gut; Antizipation von Schwierigkeiten sehr gut, kein langfristiges Konzept; Schwerpunkt der Stunde sinnvoll gewählt; keine Diskrepanz zwischen Stand der Entwicklung und dem Standard (-&gt; die können das also offenbar schon)</p>	<p><b>Begründung:</b> Begründungen schwach; Material gut; didaktisch 4 - methodisch 2</p>

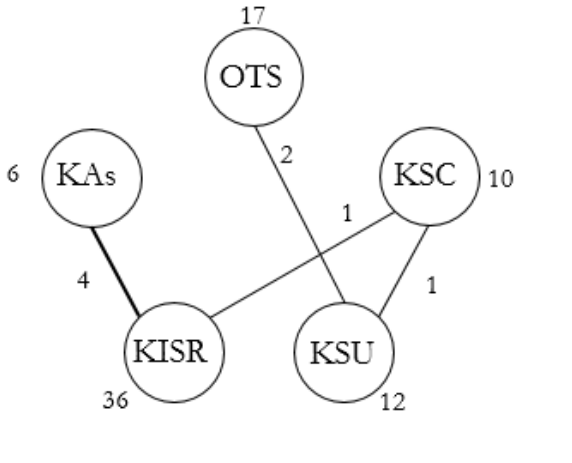
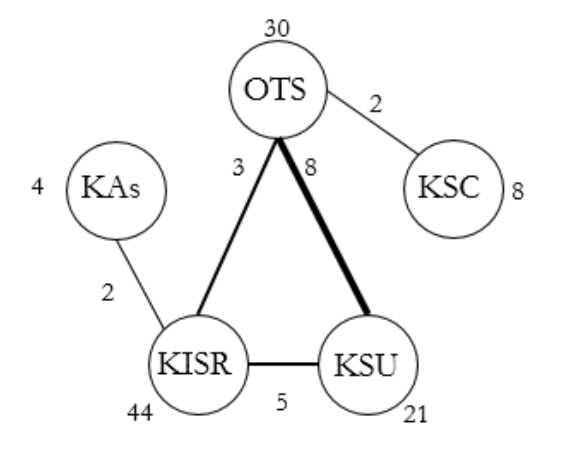
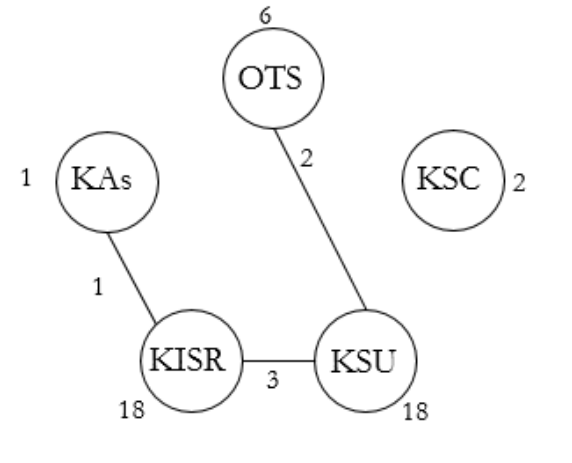
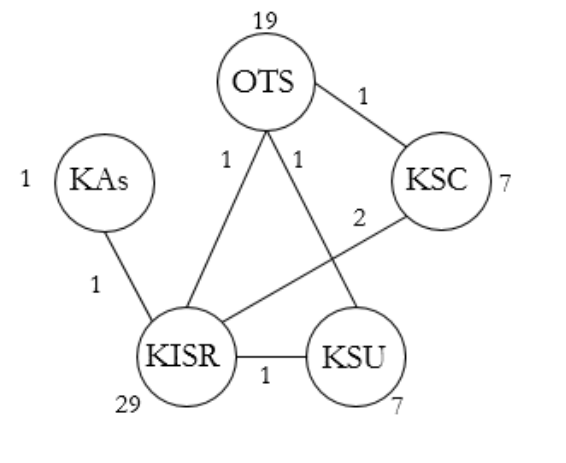
StEx_S1_F_BKS_G_01	StEx_S2_E_MU_01
	
Note des Fachseminarleiters: 2	Note des Fachseminarleiters: 2
<p><b>Begründung:</b> Schwächen bei Kompetenzen &amp; Standards (passt of nicht zusammen zur Methodik; mehrere Standards pro Stunde nicht förderbar; der eine muss sich dannin allem widerspiegeln; Mit welcher konkreten Aktion fördere ich das? (=Standardkonkretisierung)</p>	<p><b>Begründung:</b> etwas oberflächlich (Formular); Struktur schwierig (Text passt z.T. nicht zur Kapitelüberschrift; Funktion des Entwurfs nicht verstanden; Kompetenzen unklar; unklarer zeitlicher Verlauf der Reihe; Ausdrucksproblem (wesentliches Qualitätsmerkmal); Indikatoren z.T. ungeeignet; methodischer Teil in Ordnung</p>

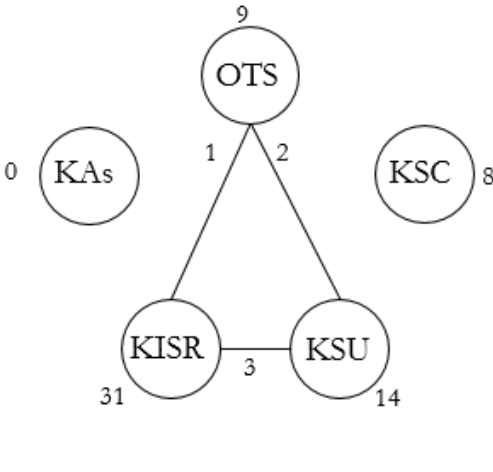
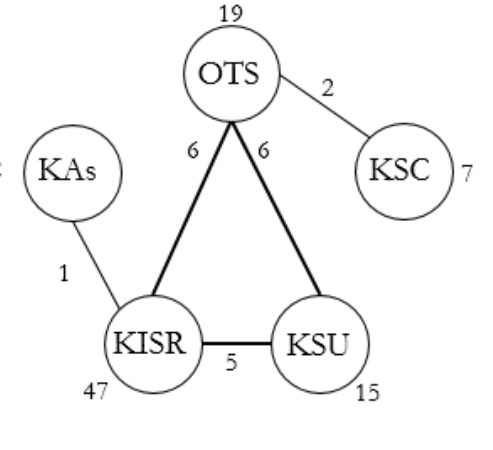
StEx_S2_F_Ö_05	StEx_S1_K_P_N_01
	
Note des Fachseminarleiters: 2	Note des Fachseminarleiters: 2-
<p><b>Begründung:</b> viel zu viel Belangloses, sehr wenig didaktisch-meth. Begründungen, fehlender Bezug zur Lerngruppe</p>	<p><b>Begründung:</b> Bezug im Unterricht zum Kompetenzschwerpunkt zu gering/oberflächlich; Kompetenzschwerpunkt zeigt sich nicht durchgehend in Aufgabenstellungen und Sicherung (Aufgaben trotzdem solide begründet, auch Instruktionsstrategien)</p>

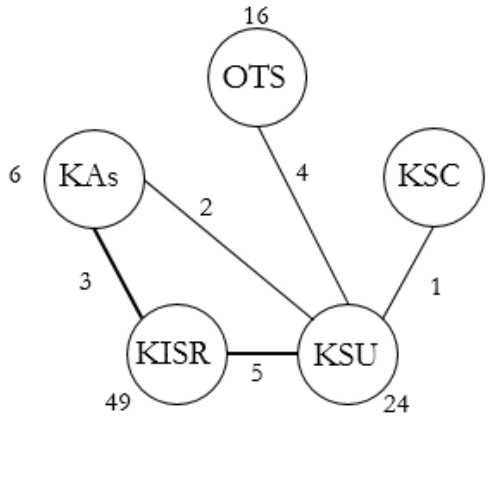
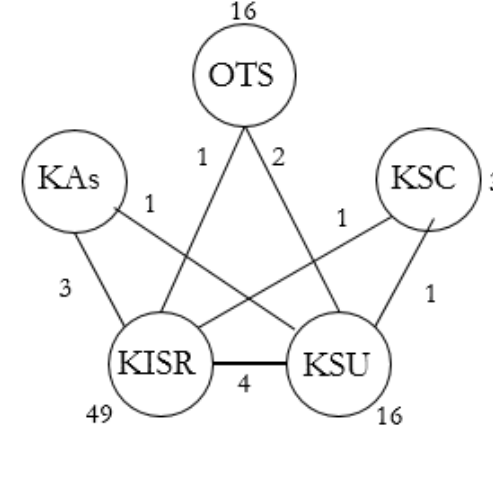


StEx_S2_B_Ö_02	StEx_S2_E_Ö_02
	
Note des Fachseminarleiters: 2-	Note des Fachseminarleiters: 2-
<p><b>Begründung:</b> meth. Teils sehr gut; auch Begründungen für Material, Methoden usw. (sehr gute Abstimmung zu Lerngruppenanalyse); Kompetenzschwerpunkt nicht leicht erkennbar/z.T. widersprüchlich; did. Reduktion fehlt; Sachstrukturanalyse ohne Bezug zur Unterrichtssituation; Unterrichtsziele nicht gut begründet; Darstellung der Standards nicht auf Kompetenzschwerpunkt Reflexion abgestimmt --&gt; Konkretisierung inkonsistent; Indikatoren fehlen völlig</p>	<p><b>Begründung:</b> sehr dicht, überfrachtet, wichtige Stellen zu wenig fokussiert; M unübersichtlich; Kompetenzförderung zu allgemein; kein Fokus, kein langfristiges Konzept; Hypothesenbildung in Sicherung nicht mehr aufgegriffen</p>

StEx_S2_E_Ö_03	StEx_S2_E_P_04
	
Note des Fachseminarleiters: 2-	Note des Fachseminarleiters: 2-
<p><b>Begründung:</b> kein langfristiges Konzept; sehr gute Abstimmung auf Lerngruppe; Entwicklung der Kompetenzstufen zu allgemein; Lerngruppenanalyse fehlt Bezug zum Standard; wenig Sachstrukturanalyse; SuS-Vorstellungen fehlen</p>	<p><b>Begründung:</b> Kompetenzentwicklung in Konkretisierung zu allgemein/oberflächlich, übersichtliche Rechenplanung; Synapsenmodelle fehlen im Anhang, Erfolgsaussichten nur drei, weil VIER Beispiele in einer Stunde viel zu viel sind</p>

StEx_S2_K_P_02	StEx_S2_K_P_10
	
<p><b>Note des Fachseminarleiters:</b> 2-</p>	<p><b>Note des Fachseminarleiters:</b> 2-</p>
<p><b>Begründung:</b> geringes Niveau; klare Aufgabenstellungen (daher noch 3); transparent für SuS Schwerpunkt der Stunde; nur ansatzweise langfristiges Konzept; Bezug zu Standards nicht sehr überzeugend; kaum didaktische Reduktion/didaktische Begründungen/Relevanzanalyse; gute Binnendifferenzierung</p>	<p><b>Begründung:</b> sehr fundiert; zu viele Standards, die zwar bedient werden, aber nicht sinnvoll fokussiert; Aufgabenstellung gut; keine Transparenz, was heute für SuS fokussiert werden soll (Inhalt oder Kommunikation)</p>
StEx_S1_E_Mo_S_01	StEx_S1_E_NU_G_01
	
<p><b>Note des Fachseminarleiters:</b> 3</p>	<p><b>Note des Fachseminarleiters:</b> 3</p>
<p><b>Begründung:</b> sprachliche Probleme (nicht Stil, sondern Ausdruck, Präzision, Sprachrichtigkeit); didaktischer Teil gefällt wenig, keine langfristige Kompetenzentwicklung (sehr schwache Begründung für Zielsetzungen); fachl. Fehler/Ungenauigkeit (Differenzierung Serums-/Plasmaspender)</p>	<p><b>Begründung:</b> methodisch ganz i Ordnung, didaktischer Teil passt aber nicht dazu; Kompetenzschwerpunkt passt nicht dazu (= Formular); nicht konsistent; Lerngruppenvoraussetzungen fehlen</p>

StEx_S1_E_NU_S_01	StEx_S1_F_BKSF_N_01
	
<p><b>Note des Fachseminarleiters:</b> 3</p>	<p><b>Note des Fachseminarleiters:</b> 3</p>
<p><b>Begründung:</b>          sprachl. Schwächen; keine langfristige Kompetenzentwicklung, unklare Darstellung Standards/Konkretisierung; Unterrichtsvoraussetzungen finden sich in Planung nicht wieder (z. B. Abstimmung von Sozialformen auf Lerngruppe); keine fundierten Begründungen; gemessen an Ansprüchen des LAA gut, der setzt das Niveau aber viel zu gering an</p>	<p><b>Begründung:</b>          evtl Überforderung (ISS); zu wenig fokussiert mit Blick auf Kompetenzschwerpunkt (nur oberflächlich angedeutet); keine sinnvolle langfristige Kompetenzentwicklung (dann auch keine sinnvolle Einzelstunde); sehr gute Lerngruppenanalyse; Sozialformen/Aufgaben gut auf Gruppe abgestimmt; überdimensioniert,/unübersichtlich</p>

StEx_S1_K_P_E_01	StEx_S2_E_Ö_05
	
<p><b>Note des Fachseminarleiters:</b> 3</p>	<p><b>Note des Fachseminarleiters:</b> 3</p>
<p><b>Begründung:</b>          Einbettung in Unterrichtssequenz nicht klar; ansonsten sehr gut</p>	<p><b>Begründung:</b>          viel zu oberflächlich, sprachlich ungenau, geringes Anforderungsniveau; didaktische Begründungen fehlen völlig; für Grundkurs zu niedriges Niveau; kein Konzept für langfristige Entwicklung</p>

<p>StEx_S1_F_BKS_E_01</p>	<p>StEx_S2_E_Ö_01</p>
<p>Note des Fachseminarleiters: 3-</p>	<p>Note des Fachseminarleiters: 3-</p>
<p><b>Begründung:</b> Voraussetzungen Lerngruppe sehr oberflächlich, langfristige Kompetenzentwicklung nicht erkennbar; Standardbezug unklar; didaktische Analyse (Sachstruktur) fehlt; immerhin sinnvolles Arbeitsmaterial und umsetzbar</p>	<p><b>Begründung:</b> kein eindeutiger Kompetenzschwerpunkt; Erkenntnisgewinnung nicht verstanden; zu wenig Fokus auf Standard; ausführliche Lerngruppenanalyse; zu wenig Fokus auf Aufgabenstellungen</p>

<p>StEx_S2_E_Ö_07</p>	<p>StEx_S2_F_P_01</p>
<p>Note des Fachseminarleiters: 3-</p>	<p>Note des Fachseminarleiters: 3-</p>
<p><b>Begründung:</b> Unterrichtsreihe nicht einleuchtend, sehr gelungene ausführliche Darstellung der Voraussetzungen; die gehen aber leider nicht in den Entwurf ein --&gt; keine Binnendiff., keine did. Reduktion</p>	<p><b>Begründung:</b> Ziele unklar formuliert, kein langfristiges Konzept, Niveaustufen nicht auf den Schwerpunkt bezogen --&gt; Kommunikation auch in Aufgabenstellungen nicht abgebildet; Sozialform + Sprachförderung sehr gut</p>

<p>StEx_S2_K_P_01</p>	<p>StEx_S2_K_P_12</p>
<p><b>Note des Fachseminarleiters:</b> 3-</p> <p><b>Begründung:</b> kein langfristiges Konzept; Eindruck der Zufälligkeit; wenig Bezug zum Standard; Aufgaben nicht passend zum Schwerpunkt, aber sonst in Ordnung</p>	<p><b>Note des Fachseminarleiters:</b> 3-</p> <p><b>Begründung:</b> Stand der Kompetenzentwicklung passt nicht zum Standard; z.T. sehr oberflächlich; Aufgaben ok --&gt; Fokus auf Kompetenzschwerpunkt fehlt jedoch; Prüfung von Lernzuwachs fehlt; Binnendiff./Sozialformen ok, aber nur oberflächlich begründet</p>

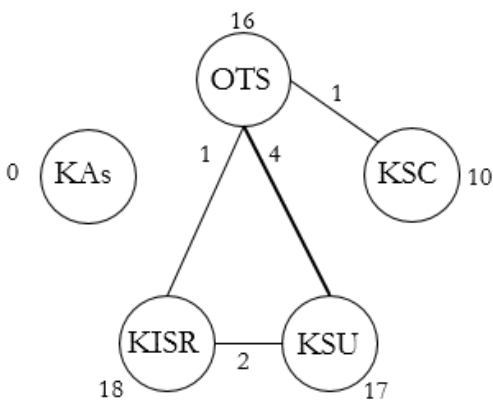
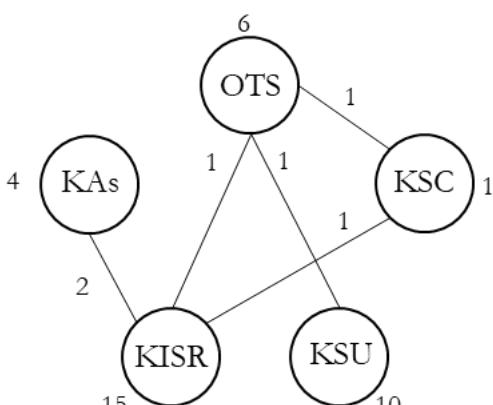
<p>StEx_S2_K_P_13</p>	<p>StEx_S1_B_Hd_Sex_01</p>
<p><b>Note des Fachseminarleiters:</b> 3-</p> <p><b>Begründung:</b> viel zu lange Reihenplanung, Stunde nicht eingebettet; nur ansatzweise langfristige Kompetenzentwicklung; Stand in Bezug auf Standard nicht erkennbar; drei Standards --&gt; Indikatoren passen nicht dazu; didaktische Reduktion akzeptabel, aber zu gering/fehlende Auseinandersetzung mit Komplexität des Themas; Aufgaben inhaltlich sehr gut!; Kontrolle des Lernzuwachses fehlt; sehr viel Material</p>	<p><b>Note des Fachseminarleiters:</b> 4</p> <p><b>Begründung:</b> Zeit nicht realistisch; Begründungen für Vorgehen fehlen, wenig fundiert; viele didaktische Überlegungen; Konzept einer langfristigen Kompetenzentwicklung fehlt (z. B. auch für Doppeljahrgangsstufe)</p>

StEx_S1_B_Hr_G_01	StEx_S1_F_BKSF_S_01
Note des Fachseminarleiters: 4	Note des Fachseminarleiters: 4
<p><b>Begründung:</b> wegen fehlenden Materials eigentlich 5; alle Ausführungen müssen sich im Material widerspiegeln; wenig fokussiert, Umsetzung fehlt; Aufgaben elementar</p>	<p><b>Begründung:</b> keine Lerngruppenanalyse; Unterrichtsreihe zeitlich nicht nachvollziehbar (kein Reihename; keine Unterscheidung in ES/DS); Kompetenzschwerpunkt nur im Ansatz erkennbar; did. Begründungen der Zielsetzungen und der Vbermittlungsstrategien fehlen</p>

StEx_S1_K_P_S_02	StEx_S1_K_R_G_01
Note des Fachseminarleiters: 4	Note des Fachseminarleiters: 4
<p><b>Begründung:</b> keine differenzierten Begründungen; anfängerhaft; keine sinnvolle Darstellung des Kompetenzschwerpunkts/keine langfristige Förderung; gewählter Schwerpunkt zwar sinnvoll, aber in Planung viel zu wenig berücksichtigt</p>	<p><b>Begründung:</b> muss ISS sein (wegen Coaching), daher sehr viele Differenzierungsmaßnahmen für sehr schwierige Lerngruppe, unrealistische Zeitplanung; Planeungsentscheidungen auf Rahmenbedingungen optimal abgestimmt</p>

StEx_S2_B_Ö_05	StEx_S2_K_P_04
<p>Network diagram for StEx_S2_B_Ö_05. Nodes: KAs (1), KISR (27), KSU (16), KSC (8), OTS (18). Edges: KAs-KISR (1), KISR-KSU (1), KISR-KSC (2), KSU-KSC (2), OTS-KSU (5).</p>	<p>Network diagram for StEx_S2_K_P_04. Nodes: KAs (13), KISR (114), KSU (36), KSC (14), OTS (37). Edges: KAs-KISR (7), KAs-KSU (5), KISR-KSU (15), KISR-KSC (8), KSU-KSC (2), OTS-KISR (8), OTS-KSU (6), OTS-KSC (4).</p>
Note des Fachseminarleiters: 4	Note des Fachseminarleiters: 4
<b>Begründung:</b> digitale Inhalte fehlen, unübersichtlich	<b>Begründung:</b> überdimensioniert, an alles gedacht, typischer Prüfungsentwurf, packt alles rein, ist viel zu viel (Formular, als Instrument ungeeignet, da Wichtiges nicht von Unwichtigem getrennt); sehr differenziert, aber nicht funktional; nicht realistisch: überfrachtet

StEx_S2_B_G_01	StEx_S2_K_P_06
<p>Network diagram for StEx_S2_B_G_01. Nodes: KAs (1), KISR (14), KSU (8), KSC (17), OTS (11). Edges: KAs-KISR (1), KISR-KSU (3), KISR-KSC (3), KSU-KSC (3), OTS-KISR (2), OTS-KSU (2), OTS-KSC (2).</p>	<p>Network diagram for StEx_S2_K_P_06. Nodes: KAs (1), KISR (25), KSU (23), KSC (7), OTS (10). Edges: KAs-KISR (1), KISR-KSU (7), OTS-KISR (2), OTS-KSU (4), KSC is isolated.</p>
Note des Fachseminarleiters: 4-	Note des Fachseminarleiters: 4-
<b>Begründung:</b> kein langfristiges Konzept; Begründungen zum Kompetenzzschwerpunkt fehlen; keine Lerngruppenanalyse; keine Reihenplanung; Standards/Konkretisierung fehlen; Aufgabenstellungen könnten erfolgreich sein	<b>Begründung:</b> Wahl Kompetenzzschwerpunkt nicht nachvollziehbar; keine Analyse des Kompetenzstandes, die mit Kommunikation zu tun hätte; Problemorientierung nicht geklärt; didaktische Begründungen nicht überzeugend; keine Progression in Aufgabenstellungen; Gruppenpuzzle nur damit begründet, dass man mehr Inhalt unterbringen kann, ist sehr schwach; schlechte Aufgabe nach Expertengruppe "Tauscht euch aus!"; Qualität der Aufgaben nicht sinnvoll; Begründungen der Sozialformen nicht sinnvoll, schlägt sich nämlich nicht in der Gestaltung der Aufgaben nieder

StEx_S2_E_P_03	StEx_S2_E_Ö_06
	
<p><b>Note des Fachseminarleiters:</b> 4-</p>	<p><b>Note des Fachseminarleiters:</b> 5</p>
<p><b>Begründung:</b> kein langfristiges Konzept; Begründungen zum Kompetenzschwerpunkt fehlen; keine Lerngruppenanalyse; keine Reihenplanung; Standards/Konkretisierung fehlen; Aufgabenstellungen könnten erfolgreich sein</p>	<p><b>Begründung:</b> Lerngruppenvoraussetzungen fehlen, Reihe nicht begründet und zeitlich unrealistisch, keine Niveaustufen differenziert beim Kompetenzschwerpunkt, Sachstrukturanalyse fehlt, Standard aus RLP nicht genannt, keine Indikatoren</p>



**Anhang 3: Kodierleitfaden Kriterienraster (Beitrag 7)**

*Leroy Großmann & Dirk Krüger*

**Biologieunterricht erfolgreich planen –  
Ein Kriterienraster zum Schreiben von Unterrichtsentwürfen**

# Kodierleitfaden

**Kontakt:**

Leroy Großmann

Freie Universität Berlin

Fachbereich Biologie, Chemie, Pharmazie

Institut für Biologie

Didaktik der Biologie

leroy.grossmann@fu-berlin.de

*K2teach wird im Rahmen der gemeinsamen Qualitätsinitiative Lehrerbildung von Bund und Ländern  
aus Mitteln des Bundesministeriums für Bildung und Forschung gefördert.*

## K Langfristige Kompetenzförderung in einer Unterrichtssequenz und Beitrag der Einzelstunde

Mit der Einführung der Bildungsstandards (KMK 2005; KMK 2020) richtete sich der Fokus nicht mehr bloß auf die Vermittlung von Inhalten, sondern auf die Entwicklung von Kompetenzen der Lernenden (Lersch 2010), also auf ihre Fähigkeiten und Fertigkeiten, Probleme in spezifischen Kontexten lösen zu können (Weinert 2001). Kompetenzorientierung kann sowohl als eine Einstellung von Lehrkräften und somit als Teil der Lehrerprofessionalität als auch als wichtiges Merkmal der Unterrichtsqualität aufgefasst werden (Weingarten 2019). Eine gute Planung würde sich entsprechend dadurch auszeichnen, dass im Rahmen einer Unterrichtssequenz eine Kompetenz langfristig zielgerichtet, systematisch und progressiv entwickelt wird (Lersch 2010), indem inhalts- und prozessbezogene Kompetenzen sinnvoll aufeinander abgestimmt werden (K1, K2). Die im Unterrichtsentwurf fokussierte Einzelstunde sollte sich in ein solches langfristiges Konzept zur Kompetenzförderung einfügen und mit einem klar formulierten Lernziel deutlich machen, inwiefern diese Einzelstunde einen Beitrag zur Kompetenzförderung leistet (K3) und woran sich dies nachweisen lässt (K4).

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
K1	Darstellung der Sequenzplanung		Die tabellarische Sequenzplanung enthält eine Auflistung aller Stunden, in denen jeweils das Thema und der ... genannt werden. ... zu fördernde Kompetenzbereich ...	... konkretisierte Standard als operationalisiertes und präzise formuliertes Lernziel ...
<b>Kodierhinweise:</b>		<i>Kodieren, wenn nur das Thema oder nur die zu fördernde Kompetenz angegeben wird oder wenn die Sequenzplanung vollständig fehlt</i>	<i>Auch kodieren, wenn zusätzlich der Standard aus dem Lehrplan wiedergegeben wird, ohne ihn zu konkretisieren.</i>	<i>Nicht kodieren, wenn nur der Standard aus dem Lehrplan wiedergegeben wird, ohne ihn auf das konkrete Unterrichtsthema hin anzupassen.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
K2	Progression der Kompetenzentwicklung in der Sequenzplanung		Die Sequenzplanung folgt einem erkennbaren Kompetenzschwerpunkt, der ... ... in einigen Stunden progressiv in steigender Komplexität, Abstraktion oder Eigenständigkeit gefördert werden soll.	... in der überwiegenden Zahl der Stunden progressiv in steigender Komplexität, Abstraktion oder Eigenständigkeit gefördert werden soll. Inhalts- und prozessbezogene Kompetenzen werden sinnvoll miteinander verknüpft, d. h. in den Stunden werden Zusammenhänge zu bereits erarbeiteten Wissens-elementen hergestellt und diese in plausiblen Anwendungssituationen genutzt.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn in der Sequenzplanung keine Kompetenzen wiederholt progressiv bedient werden, d. h. wenn beliebige Stunden austauschbar wären, ohne dass sich das auf die Sequenz auswirken würde und es keinen klaren Schwerpunkt gibt, was die SuS nach der Sequenz besser können sollen als vorher.</i>	<i>Kodieren, wenn zumindest eine Kompetenz (z. B. die Auswertung von Daten) mindestens zwei Mal in der Sequenz auftritt und anhand der Standardkonkretisierung eine Progression erkennbar ist.</i>	<i>Kodieren, wenn in mehr als der Hälfte der Stunden systematisch Aufbau eine Kompetenz (z. B. die Auswertung von Daten) entwickelt wird. Nicht kodieren, wenn sich nur ein Kompetenzbereich (z. B. Erkenntnisgewinnung) wiederholt, aber mit ganz unterschiedlichen Aspekten, die nicht systematisch aufeinander aufbauen (Modellieren, Hypothesen, Datenauswertung...)</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
K3	Standardkonkretisierung/ Lernziel		In der Standardkonkretisierung der Unterrichtsstunde wird ein operationalisiertes Lernziel formuliert, das auf einem Standard laut Rahmenlehrplan basiert. Es wird ... ... eher der Lernprozess als das Lernziel beschrieben.	... im Lernziel präzise beschrieben, was die SuS nach der Stunde konkret besser können als vorher.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn eine Standardkonkretisierung fehlt oder ein Ziel formuliert wird, das keinen Standard des Lehrplans aufgreift.</i>	<i>Kodieren, wenn das Lernziel nicht deutlich wird, d. h. wenn der "Die SuS..."-Satz so konkret auf den Lernprozess ("indem"-Satz) und damit das ausgewählte Thema angepasst ist, dass hier überwiegend der Verlauf der geplanten Stunde skizziert wird.</i>	<i>Kodieren, wenn deutlich wird, welche Kompetenz die SuS am Ende der Stunde (weiter-)entwickelt haben. Vom Umfang her kann der "indem"-Satz, der den Lernprozess beschreibt, länger sein als der "Die SuS..."-Satz (Lernziel), es muss aber deutlich werden, dass hier etwas gelernt wird, was auch in Folgestunden bei anderen Themen nutzbar ist.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
K4	Indikatoren zum Nachweis des Kompetenzzuwachses		Die Indikatoren zum Nachweis des Kompetenzzuwachses beziehen sich spezifisch auf die Standardkonkretisierung und beschreiben ... ... das Verhalten der Lerngruppe.	... ein beobachtbares Verhalten der Lerngruppe mit geeigneten Operatoren, so dass sich das Erreichen des Kompetenzzuwachses eindeutig nachweisen lässt.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn keine Indikatoren aufgeführt werden.</i>	<i>Kodieren, wenn das geplante Verhalten der SuS im Verlauf der Stunde aufgezählt wird (z. B. SuS lesen Text, SuS füllen AB aus, SuS präsentieren Ergebnisse).</i>	<i>Kodieren, wenn mindestens zwei der aufgeführten Indikatoren ein beobachtbares Verhalten beschreiben, das unmittelbar in Verbindung zur Standardkonkretisierung steht. Nicht kodieren, wenn der Bezug zur Standardkonkretisierung nicht in mind. zwei Fällen erkennbar ist.</i>

## I Begründete Auswahl und Sequenzierung der Inhalte zur Kompetenzförderung

Dass der Kompetenzbereich *Mit Fachwissen umgehen* nicht ohne biologische Inhalte gefördert werden kann, ist offensichtlich. Auch die drei prozessbezogenen Kompetenzbereiche *Erkenntnisse gewinnen*, *Kommunizieren* und *Bewerten* (KMK, 2005; KMK, 2020) kommen nicht ohne biologisches Fachwissen aus. Viele Planungsmodelle betonen die Bedeutung einer Analyse des Fachinhalts (z. B. *Fachliche Klärung im Modell der Didaktischen Rekonstruktion*, Gropengießer & Kattmann 2013). Eine gute Planung würde sich entsprechend dadurch auszeichnen, dass auf der Basis einer fachlich korrekten fachwissenschaftlichen Darstellung des Fachinhalts (I1) die Relevanz des ausgewählten Fachinhalts begründet (I2) und vor dem Hintergrund des bereits erworbenen Wissens der Lernenden (Lersch, 2010) reduziert wird (I3). Hierbei geht also nicht allein um eine fachwissenschaftliche, sondern auch um eine fachdidaktische Perspektive. Dazu gehört u.a. auch eine bewusste Entscheidung über die notwendigen Fachbegriffe (I4), mit denen die Lernenden konfrontiert werden und die zum Erreichen des angestrebten Kompetenzzuwachses (vgl. K3) notwendig sind.

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
I1	Sachstruktur-analyse		In der Sachstrukturanalyse wird der Lerngegenstand fachlich korrekt ... ... aus überwiegend fachwissenschaftlicher Perspektive analysiert.	... so knapp wie möglich und so ausführlich wie nötig aus überwiegend fachdidaktischer Perspektive mithilfe fachwissenschaftlicher und fachdidaktischer Literatur dargestellt, d. h. in seiner Struktur (innere Logik, Ebenen, Perspektiven) analysiert.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn die Sachstrukturanalyse fehlt oder sie mindestens einen fachlichen Fehler enthält.</i>	<i>Kodieren, wenn die Sachstrukturanalyse wie eine Darstellung aus einem Lehrbuch gestaltet ist und nur wenige Ansätze einer didaktischen Perspektive vorhanden sind (z. B. eine didaktische Reduktion).</i>	<i>Nur kodieren, wenn zusätzlich zur fachwissenschaftlichen Darstellung auch eindeutig eine fachdidaktische Perspektive eingenommen wird, d. h. wenn aus der Sachstruktur (weniger) wichtige Aspekte abgeleitet werden, die für den Lernprozess relevant sein könnten. Fachdidaktische Quellen können als Anhaltspunkte dienen.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
I2	Auswahl des Lerngegenstandes		Die Auswahl des Lerngegenstandes zur Förderung des angestrebten Kompetenzzuwachses wird ... ... mit pragmatischen Argumenten (z. B. Vorgaben des Lehrplans/schulinternen Curriculums, Zeitnot) begründet.	... mit fachdidaktischer Relevanz (Schülerrelevanz, Gesellschaftsrelevanz, Wissenschaftsrelevanz) und Prinzipien (z. B. Problemorientierung, Handlungsorientierung) begründet.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn die Relevanz des Lerngegenstandes nicht begründet wird.</i>	<i>Kodieren, wenn die Relevanz des Lerngegenstandes ausschließlich mit der Verortung im Lehrplan oder zeitlichen Aspekten im Verlauf des Schuljahres begründet wird.</i>	<i>Kodieren, wenn die Relevanz des Lerngegenstandes auch mit der Lebenswelt, den Erfahrungen, den Interessen der konkreten Lerngruppe begründet wird oder seine Zugänglichkeit bzw. Beispielhaftigkeit (exemplarisches Prinzip) betont wird.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
I3	Didaktische Rekonstruktion		Der Lerngegenstand wird ... ... didaktisch reduziert, indem ausgewählte Aspekte des Lerngegenstandes ausgelassen oder vereinfacht werden.	... didaktisch rekonstruiert, indem ausgewählte Aspekte des Lerngegenstandes angesichts der Lerngruppenvoraussetzungen didaktisch reduziert werden und ein aussichtsreicher Lernweg beschrieben und begründet wird.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn es keine didaktische Reduktion gibt.</i>	<i>Kodieren, wenn der Lerngegenstand in seinem Umfang reduziert wird, ggf. auch ohne Begründung. Aussagen, die erkennen lassen, dass der Lerngegenstand in der Komplexität oder im Umfang reduziert werden, genügen.</i>	<i>Nur kodieren, wenn auf der Grundlage einer didaktischen Reduktion beschrieben wird, wie das Thema gewinnbringend unterrichtet werden kann (Lernweg). Dies können z. B. Bezugnahmen auf die Reihenfolge sein, in der das Thema erarbeitet wird, oder auf Lernschwierigkeiten, die dem Thema innewohnen und die daher mit einem bestimmten Lehr-Lern-Konzept bearbeitet werden. Der Bezug zu den Lerngruppenvoraussetzungen muss nicht explizit sein, ein impliziter Hinweis kann ausreichen (z. B. dass das Thema in Kl. 10 nur oberflächlich behandelt wird, weil es in der Sek. II nochmal vertieft aufgegriffen wird).</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
I4	Auswahl relevanter Fachbegriffe		Die in der Stunde zu nutzenden Fachbegriffe werden auf ... ... einige relevante, aber auch weniger relevante Fachbegriffe beschränkt.	... wenige, besonders relevante Fachbegriffe beschränkt und die Auswahl wird begründet.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn keine Begriffe genannt werden, die die SuS in der geplanten Stunde unbedingt benötigen.</i>	<i>Kodieren, wenn an irgendeiner Stelle im Entwurf (Tabellen, Fließtext) ein Begriff oder mehrere Begriffe genannt werden, die die SuS in der geplanten Stunde unbedingt benötigen. Mindestens einer dieser Begriffe muss für die Standardkonkretisierung relevant sein.</i>	<i>Kodieren, wenn an irgendeiner Stelle im Entwurf (Tabellen, Fließtext) ein Begriff oder mehrere Begriffe genannt werden, die die SuS in der geplanten Stunde unbedingt benötigen. Nur kodieren, wenn diese Begriffe für die Standardkonkretisierung tatsächlich sehr relevant sind und wenn diese Auswahl ggf. auch sehr kurz begründet wird.</i>

## S Voraussetzungen der Schülerinnen und Schüler in Bezug auf Kompetenzen, Vorwissen und Methoden

Eine Unterrichtsstunde ist nicht per se gut oder schlecht geplant. Wichtig ist, dass der geplante Unterricht an die bereits entwickelten Kompetenzen der konkreten Lerngruppe anknüpft und sie weiterentwickelt (König, Buchholtz & Dohmen 2015). Zu diesem Zweck ist es nötig, vorab präzise zu analysieren, wie weit die Kompetenzentwicklung konkret in Bezug auf das angestrebte Ziel (vgl. K3) bereits fortgeschritten ist (S1) und was der nächste Schritt der Kompetenzförderung sein sollte (S2). Zudem ist es ratsam, in Bezug auf den biologischen Fachinhalt (vgl. I) zu berücksichtigen, welche Schülervorstellungen die Lernenden zum biologischen Fachinhalt mitbringen könnten, da dies ein wichtiger Ausgangspunkt für die Gestaltung des Lernprozesses sein könnte (S3) (Gropengießer & Kattmann 2013). Darüber hinaus gibt es themenspezifische Lernschwierigkeiten, die den Lernprozess erschweren könnten und die in der Planung berücksichtigt werden könnten, um die Chance zu erhöhen, dass der angestrebte Lernweg (vgl. I3) beschritten werden kann (S4). Bei alledem ist auch wichtig, die bereits entwickelten methodischen Fähigkeiten (z. B. Kenntnis einer Methode) zu analysieren, mit deren Hilfe das angestrebte Ziel erreicht werden soll (S5).

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
S1	Analyse des Kompetenzstandes		Für das in der Standardkonkretisierung formulierte Ziel wird der Kompetenzstand der Lerngruppe für prototypische SuS auf mindestens zwei Niveaustufen analysiert. Es werden ... ... allgemeine kompetenzbereichsspezifische Aspekte beschrieben, in denen die SuS bereits über Kompetenzen verfügen.	... konkret auf die in der Standardkonkretisierung beschriebenen Tätigkeiten bezogene Aspekte beschrieben, in denen die SuS bereits über Kompetenzen verfügen.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn eine Lerngruppenanalyse fehlt oder wenn sie nicht mindestens zwei Niveaustufen analysiert.</i>	<i>Nur kodieren, wenn mindestens für zwei Niveaustufen (oft "Minimal-/Maximalstandard" genannt) der Kompetenzstand analysiert wird. Nicht kodieren, wenn die Analyse zumindest teilweise konkret auf die Standardkonkretisierung bezogen ist (☐ "erweitert").</i>	<i>Kodieren, wenn für mindestens zwei Niveaustufen Teilaspekte der Standardkonkretisierung analysiert werden, in denen die SuS (noch nicht) über Kompetenzen verfügen. Es sollte jeweils mindestens ein Aspekt präzise analysiert werden, der eindeutig für das in der Standardkonkretisierung beschriebene Lernziel relevant ist. Oft wird zwischen Minimal-, Regel- und Maximalstandard unterschieden und es werden Vorhersagen gemacht, was die SuS in der Stunde leisten werden. Diese Prognosen sind nicht als Analysen des Kompetenzstandes aufzufassen.</i>
Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
S2	Progression der Kompetenzentwicklung in der Einzelstunde		Zwischen dem aktuellen Kompetenzstand der SuS und der Standardkonkretisierung für die Unterrichtsstunde ist eine Progression ... ... teilweise erkennbar, weil die Standardkonkretisierung in einigen wenigen Aspekten über die bereits entwickelten Kompetenzen der SuS hinausgeht.	... ... erkennbar, weil die Standardkonkretisierung in wesentlichen Aspekten über die bereits entwickelten Kompetenzen der SuS hinausgeht.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn zwischen dem aktuellen Kompetenzstand der SuS und dem angestrebten Kompetenzzuwachs keine Progression, d. h. keine Diskrepanz erkennbar ist.</i>	<i>Kodieren, wenn zwischen dem dargestellten Kompetenzstand der Lerngruppe (s. S1) und dem Ziel in der Standardkonkretisierung in Teilen eine Progression, d. h. eine Diskrepanz, erkennbar ist. Kodieren, wenn sich zumindest ein Teilaspekt der Analyse (S1) in der Standardkonkretisierung wiederfindet und erkennbar wird, dass dieser gefördert werden soll.</i>	<i>Kodieren, wenn zwischen dem dargestellten Kompetenzstand der Lerngruppe (s. S1) und dem Ziel in der Standardkonkretisierung eine Progression, d. h. eine Diskrepanz, erkennbar ist. Nur kodieren, wenn deutlich wird, dass die/ einige SuS etwas noch nicht können, was in der geplanten Stunde gefördert werden soll. Die Standardkonkretisierung muss daran direkt ansetzen. Kann nicht kodiert werden, wenn die Analyse (S1) ausschließlich beschreibt, dass alle SuS verschiedene Dinge bereits können.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
S3	Schüler- vorstellungen		Für die Lerngruppe werden themenspezifische Schülervorstellungen ... ... ohne Belege beschrieben.	... konkret für die Lerngruppe diagnostiziert und in der Lehr- Lern-Struktur mit Bezug zur fachdidaktischen Literatur berücksichtigt.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn kein Bezug zu Schülervorstellungen hergestellt wird.</i>	<i>Kodieren, wenn Schülervorstellungen allgemein beschrieben werden (z. B. "Jugendliche in dem Alter glauben ..."), ohne dass dies für die konkrete Lerngruppe diagnostiziert wurde. Dafür kann auch Fachliteratur (z. B. Kattmann 2015; Hammann &amp; Asshoff 2014) genutzt werden</i>	<i>Kodieren, wenn für die konkrete Lerngruppe mindestens eine Schülervorstellung beschrieben wird. Nur kodieren, wenn auf einschlägige Fachliteratur (z. B. Kattmann 2015; Hammann &amp; Asshoff 2014) Bezug genommen wird.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
S4	Lernschwierig- keiten		Es werden Lernschwierigkeiten beschrieben, die ... ... allgemeindidaktische Probleme (z. B. sprachliche Hürden, Förderbedarfe) adressieren	... themenspezifische Probleme (z. B. Abstraktheit des Lerngegenstandes) adressieren und in der Gestaltung der Lehr- Lern-Struktur berücksichtigt werden.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn kein Bezug zu Lernschwierigkeiten hergestellt wird.</i>	<i>Kodieren, wenn auch für einzelne SuS allgemeine Probleme beschrieben werden, oft in Bezug auf die deutsche Sprache oder besondere Lernschwierigkeiten (z. B. Lese-Rechtschreib-Schwäche, ADHS).</i>	<i>Nur kodieren, wenn mindestens auf eine themenspezifische Lernschwierigkeit verwiesen wird und diese in der Stundengestaltung aufgegriffen wird (z. B. im Einstieg, auf einem Arbeitsblatt, auf einer Karte)</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
S5	Methodische Fähigkeiten		Die methodischen Fähigkeiten der Lerngruppe, die zum Erreichen des in der Standardkonkretisierung formulierten Lernziels ... analysiert. ... notwendig sind, werden ...	... analysiert und die Beschreibung beschränkt sich auf diejenigen Aspekte, die zum Erreichen des Ziels notwendig sind.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn kein Bezug zu methodischen Fähigkeiten hergestellt wird.</i>	<i>Kodieren, wenn mindestens zwei für den Lernprozess relevante methodischen Fähigkeiten (z. B. Textlesekompetenz, Erfahrungen in Gruppenarbeit, Erfahrungen mit spez. Binnendifferenzierungsmaßnahmen) beschrieben werden</i>	<i>Nur kodieren, wenn die beschriebenen methodischen Fähigkeiten (z. B. Textlesekompetenz, Erfahrungen in Gruppenarbeit, Erfahrungen mit spez. Binnendifferenzierungsmaßnahmen) bis auf maximal zwei Ausnahmen eindeutig erkennbar in Beziehung zur Standardkonkretisierung und dem angestrebten Lernprozess stehen.</i>



## L Lehr-Lern-Struktur der Einzelstunde

Auf Grundlage der Überlegungen zur Kompetenzorientierung (vgl. K), zum Fachinhalt (vgl. I) und zu den Voraussetzungen, die die Schülerinnen und Schüler mitbringen (vgl. S) wird eine Lehr-Lern-Struktur entwickelt, mit der der Lehr-Lern-Prozess gesteuert wird (Leisen 2014). Dazu gehören sämtliche Aspekte, mit deren Hilfe die Kompetenzförderung gestaltet wird: Auf verschiedenen Ebenen methodischen Handelns (Meyer 2020) müssen Entscheidungen getroffen werden: So sollte der angestrebte Kompetenzzuwachs (vgl. K3) konsequent vom Beginn bis zum Ende der Stunde verfolgt (L1) und der Beitrag der Unterrichtsschritte zum Erreichen des Ziels begründet werden (L2). Die in den jeweiligen Phasen eingesetzten Methoden (z. B. Sozialformen, Impulse) sollten dazu beitragen, den angestrebten Kompetenzzuwachs zu erreichen (L3). Auf der Ebene der Handlungsmuster sind Aufgabenstellungen (L4) und damit in Verbindung stehende Medien und Arbeitsmaterialien (L5) inklusive eines Erwartungshorizonts (L6) notwendig, die ebenfalls zum Erreichen dieses Ziels beitragen sollten. Um dies allen Lernenden zu ermöglichen, wäre eine binnendifferenzierte Gestaltung des Lehr-Lern-Prozesses wünschenswert (L7). Mit Blick auf die Durchführung der geplanten Stunde ist es zu empfehlen, die Lehr-Lehr-Struktur in der Planung immer wieder darauf hin zu überprüfen, welche Hindernisse im Lehr-Lern-Prozess auftreten könnten und welche Alternativen zur Beseitigung dieser Hindernisse infrage kommen (L8). Auf der Ebene der Unterrichtsschritte ist es im Sinne einer guten Strukturierung (Helmke 2015) hilfreich, den Lernenden das Ziel transparent zu machen (L9) (Anderson & Krathwohl 2014).

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L1	Struktur des Kompetenzerwerbs		Die Phasen bauen aufeinander auf und führen zu einem klaren Ziel, wobei sich der ausgewählte Kompetenzschwerpunkt in ... den Aufgabenstellungen der Erarbeitungsphase und als Fokus in der Sicherungsphase zeigt.	... allen Phasen und Phasenübergängen zeigt.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn der angestrebte Kompetenzschwerpunkt in keiner Phase, nur in der Erarbeitung oder nur in der Sicherung erkennbar ist.</i>	<i>Kodieren, wenn die Aufgabenstellungen (siehe Arbeitsblätter) und die Sicherungsphase eindeutig das angestrebte Ziel adressieren. Auch kodieren, wenn von mehreren Aufgaben in der Erarbeitungsphase nicht alle, aber mehrere eindeutig zielführend sind. In der Sicherungsphase muss deutlich werden, dass das angestrebte Ziel (z. B. Planung eines Experiments □ Erkenntnisgewinnung) Gegenstand der Besprechung ist, nicht allein der Fachinhalt (□ Fachwissen).</i>	<i>Kodieren, wenn die einzelnen Phasen und Phasenübergänge eindeutig erkennbar zum angestrebten Ziel führen, das wird oft in den Begründungen der Lehr-Lern-Struktur explizit betont. Auch kodieren, wenn dies in den Begründungen nicht explizit benannt wird. Hier ist es notwendig, vor dem Hintergrund der Standardkonkretisierung zu prüfen, ob die Phasen jeweils einen Schritt hin zu diesem Ziel beitragen (z. B. ob ein Einstieg nur zum Motivieren/ Aktivieren genutzt wird, oder ob er bereits die angestrebte Kompetenzentwicklung anbahnt). Auch kodieren, wenn die Phasenübergänge neben einer kompetenzorientierten Überleitung eher organisatorischen Charakter haben (z. B. Verteilung von Materialien, Umbau von Tischen).</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L2	Funktion der Phasenstruktur		Der geplante Verlauf der Stunde wird in chronologischer Reihenfolge vorrangig ... ... beschrieben, indem ausgewählte Unterrichtsschritte der verschiedenen Phasen in ihrer Gestaltung und ihrem geplanten Verlauf dargestellt werden.	... begründet, indem der Beitrag ausgewählter Unterrichtsschritte der verschiedenen Phasen zum Erreichen des angestrebten Kompetenzzuwachses erläutert wird.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn der Verlauf der Stunde nicht oder nicht chronologisch dargestellt wird. Auch kodieren, wenn der Verlauf chronologisch dargestellt wird, aber mindestens eine geplante Phase nicht dargestellt wird.</i>	<i>Kodieren, wenn die geplante Unterrichtsstruktur überwiegend beschreibend, nicht begründend, dargestellt wird. Dazu muss auf alle Phasen (Einstieg bis Sicherung, ggf. Transfer) Bezug genommen werden. Auch kodieren, wenn an wenigen vereinzelten Stellen etwas begründet wird, der Gesamteindruck aber eher deskriptiv ist, weil der Beitrag der jeweiligen Phasen zum Erreichen des Ziels nicht begründet wird.</i>	<i>Nur kodieren, wenn die geplante Unterrichtsstruktur überwiegend begründet wird und wenn deutlich wird, dass die Gestaltung der Unterrichtsschritte bewusst getroffen worden ist und mit dem Lernziel begründet wird.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L3	Angemessenheit der Methoden		Die ausgewählten Methoden sind ... Erreichen des in der Standardkonkretisierung beschriebenen Ziels ermöglichen. ... teilweise geeignet, da sie ein inhalts- oder prozessbezogenes ...	... geeignet, da sie ein inhalts- und prozessbezogenes ...
<b>Kodierhinweise:</b>		<i>Kodieren, wenn die Methoden der Erarbeitungsphase (dazu zählen auch die Sozialformen) nicht geeignet sind, weil sie nicht zum Ziel führen, das in der Standardkonkretisierung adressiert wird. Wenn z. B. in einer Bewertungsstunde, in der eher Diskussions- und Debattenformate (z. B. Fishbowl, Podiumsdiskussion o.ä.) geeignet sind, um verschiedene Perspektiven zu diskutieren, eine Partnerarbeit durchgeführt wird, in der ein Arbeitsblatt gelesen und besprochen wird, dann ist diese Methode nicht geeignet, auch wenn z.T. zumindest ein inhaltsbezogenes Ziel erreicht werden kann: Da das Ziel Bewertungskompetenz ist, müsste die Methodenwahl dies eindeutig adressieren, daher wäre es in diesem Beispiel als nicht grundständig zu kodieren.</i>	<i>Kodieren, wenn die Methoden der Erarbeitungsphase (dazu zählen auch die Sozialformen) geeignet sind, entweder einen fachlichen (inhaltsbezogenen) oder einen kompetenzorientierten (prozessbezogenen) Kompetenzzuwachs zu ermöglichen. "Geeignet" bedeutet, dass die Methode Aktivitäten und Denkprozesse der SuS anregt, die unmittelbar zum Erreichen des Ziels führen. Wenn z. B. in einer Fachwissensstunde verschiedene Verhütungsmethoden anhand von Arbeitsblättern erarbeitet werden sollen, wäre ein Gruppenpuzzle (arbeitsteilige Erarbeitung, anschließender Austausch) geeigneter als eine Stationsarbeit (alle SuS durchlaufen alle Stationen), da es nicht zielführend ist, wenn alle SuS mehrfach nach demselben Muster Arbeitsblätter lesen. Oft gibt es nicht die "eine" geeignete Methode, also nur kodieren, wenn es eine offensichtlich bessere Alternative gegeben hätte.</i>	<i>Kodieren, wenn die Methoden der Erarbeitungsphase (dazu zählen auch die Sozialformen) geeignet sind, sowohl einen fachlichen (inhaltsbezogenen) als auch einen kompetenzorientierten (prozessbezogenen) Kompetenzzuwachs zu ermöglichen. In Stunden zu den Kompetenzbereichen Erkenntnisgewinnung, Kommunikation und Bewertung ermöglichen sie also eine Auseinandersetzung und Reflexion über den Fachinhalt und den Prozess (also die Erkenntnisgewinnung, die Kommunikation oder die Bewertung). Nicht kodieren, wenn z. B. in einer Erkenntnisgewinnungsstunde in einem Gruppenpuzzle nur der Fachinhalt erarbeitet wird (inhaltsbezogen), dies zur Erarbeitung des Erkenntnisgewinnungsaspekts (prozessbezogen) aber ungeeignet ist.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L4	Kompetenzfördernde Interaktion mit Aufgabencharakter		Die Aufgabenstellungen adressieren den angestrebten Kompetenzzuwachs überwiegend. Die Gestaltung der Aufgabenstellungen in der Erarbeitungsphase wird hinsichtlich ausgewählter fachdidaktischer Aspekte (z. B. Schwierigkeitsgrad, Offenheit, sprachliche Komplexität) ... ... analysiert.	... analysiert und mit Blick auf die Lerngruppenvoraussetzungen oder den zu fördernden Kompetenzbereich begründet.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn keine Aufgabenstellung in ihrer Gestaltung analysiert wird.</i>	<i>Kodieren, wenn die Gestaltung mindestens einer Aufgabenstellung unter den o.g. Aspekten analysiert wird. Nicht kodieren, wenn nur beschrieben wird, dass die SuS eine Aufgabe bearbeiten sollen.</i>	<i>Kodieren, wenn mindestens eine Aufgabenstellung der Erarbeitungsphase eindeutig das Lernziel adressiert, in ihrer Gestaltung unter den o.g. Aspekten analysiert wird und explizit begründet wird. Auch kodieren, wenn das für einige Aufgabenstellungen nicht getan wird.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L5	Kompetenzfördernde Arbeitsmaterialien		Die Medien (z. B. Arbeitsblätter, Filme, Hilfekärtchen) adressieren den angestrebten Kompetenzzuwachs ... ... teilweise.	... überwiegend. Die Gestaltung der Medien wird hinsichtlich ausgewählter fachdidaktischer Aspekte (z. B. Primär-/Sekundärerfahrung, Abstraktheit, Komplexität) analysiert und mit Blick auf die Lerngruppenvoraussetzungen oder den zu fördernden Kompetenzbereich begründet.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn die Medien den angestrebten Kompetenzzuwachs nicht adressieren.</i>	<i>Kodieren, wenn zumindest das Arbeitsblatt den angestrebten Kompetenzzuwachs adressiert, wenn auch nicht in allen Aufgaben, Grafiken o.ä., aber immerhin teilweise.</i>	<i>Kodieren, wenn mindestens ein Medium in der o.g. Weise analysiert und seine Gestaltung begründet wird. Nur kodieren, wenn dieses Medium auch tatsächlich den angestrebten Kompetenzzuwachs adressiert.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L6	Erwartungshorizont		Ein Erwartungshorizont für die Arbeitsmaterialien bzw. das Tafelbild ... ... wird bereitgestellt und ist fachlich korrekt.	... wird bereitgestellt, ist fachlich korrekt und ist dem Niveau der Lerngruppe angemessen.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn es für ein Arbeitsblatt oder für ein Tafelbild, das entwickelt werden soll, keinen Erwartungshorizont gibt. Auch kodieren, wenn nur für ein Arbeitsblatt bzw. nur für das Tafelbild ein Erwartungshorizont fehlt. Auch kodieren, wenn einer der Erwartungshorizonte einen fachlichen Fehler enthält.</i>	<i>Kodieren, wenn für alle Arbeitsmaterialien und für das Tafelbild (falls eines entwickelt werden soll) ein Erwartungshorizont angegeben wird, der keinen fachlichen Fehler enthält.</i>	<i>Kodieren, wenn für alle Arbeitsmaterialien und für das Tafelbild (falls eines entwickelt werden soll) ein Erwartungshorizont angegeben wird, der keinen fachlichen Fehler enthält. Nur kodieren, wenn die Erwartungshorizonte dem in der Standardkonkretisierung angestrebten Niveau entsprechen (z. B. keine jahrgangsstufenuntypischen Fachbegriffe; keine Abstraktheit, die dem aktuellen Kompetenzstand der Lerngruppe erkennbar übersteigt) und keine Lösung erwartet wird, die von SuS mithilfe des angebotenen Materials nicht erreichbar ist.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L7	Binnendifferenzierung		Die Maßnahmen der Binnendifferenzierung zielen darauf ab, ... einen Kompetenzzuwachs zu ermöglichen. ... einzelnen SuS (Förderbedarf, sprachliche Hürden) bzw. den leistungsschwächeren SuS ...	... sowohl den leistungsschwächeren als auch den leistungsstärkeren SuS ...
<b>Kodierhinweise:</b>		<i>Kodieren, wenn keinerlei Maßnahmen ergriffen werden, um SuS verschiedener Leistungsniveaus, Lerntempos usw. auf unterschiedliche Weise Kompetenzzuwachs zu ermöglichen.</i>	<i>Kodieren, wenn entweder für einzelne SuS oder für eine Gruppe leistungsschwächerer SuS Maßnahmen ergriffen werden. Maßnahmen können Zusatzaufgaben, weniger Aufgaben, unterschiedlich schwierige Aufgaben, Sozialformen (Gruppenarbeit, Partnerarbeit), Medien/Materialien o.ä. umfassen.</i>	<i>Kodieren, wenn explizit Maßnahmen für Schwächere und für Stärkere ergriffen werden. Darunter fällt auch eine Differenzierung bezüglich des Lerntempos (Schnellere, Langsamere). Bitte auch auf Powerpoint-Folien achten, falls vorhanden: Manchmal stehen Zusatzaufgaben o.ä. dort oder auf den Arbeitsblättern.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L8	Antizipierte Hindernisse im Lehr-Lern-Prozess und Alternativen		Es werden ... Hindernisse im Lehr-Lern-Prozess antizipiert. ... überwiegend pragmatische (z. B. technische Aspekte) oder allgemein-pädagogische (z. B. Störungen) ...	... fachliche oder fachdidaktische ... Darüber hinaus werden Handlungsalternativen entwickelt, die sicherstellen, dass der angestrebte Kompetenzzuwachs erreicht werden kann.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn keine möglichen Hindernisse (Bruchstellen) beschrieben werden.</i>	<i>Kodieren, wenn pragmatische oder allgemein-pädagogische Hindernisse (Bruchstellen) beschrieben werden. Auch kodieren, wenn zusätzlich Handlungsalternativen dafür angeboten werden</i>	<i>Kodieren, wenn mindestens zwei mögliche Alternativen angegeben werden, die genutzt werden, um die eigenen Planungsentscheidungen zu begründen. Nur kodieren, wenn die Alternativen fachlich oder fachdidaktisch sind (z. B. Auswahl eines Realobjekts, einer Sozialform, einer Methode).</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
L9	Transparenz der Unterrichtsstruktur		Die Struktur der Stunde wird den SuS durch eine Strukturierungshilfe transparent gemacht, die den Verlauf der Stunde ... ... im Sinne eines Programmablaufs mit für SuS irrelevanten Angaben (z. B. Zeit) zeigt.	... mit Bezug auf die zu fördernden Kompetenzen übersichtlich zeigt.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn es keinerlei Strukturierungshilfe gibt. Auch kodieren, wenn die Lehrkraft nur mündlich skizziert, was in der Stunde passieren wird (wenn vorhanden, dann meist als Impuls zu Beginn des Einstiegs im Stundenverlaufsplan).</i>	<i>Kodieren, wenn es eine schriftliche/grafische Übersicht über den Lernprozess gibt, die den Ablauf der Stunde zeigt. Meistens geschieht dies in Form von Stichpunkten und Angabe der ungefähren Dauer (z. B. 1. Partnerarbeit AB1, 15 Min.; Gruppenarbeit Besprechung, 8 Min....).</i>	<i>Kodieren, wenn es eine schriftliche/grafische Übersicht über den Lernprozess gibt, die eindeutig die zu fördernde Kompetenz in den Mittelpunkt stellt (z. B. in einer Erkenntnisgewinnungsstunde eine Übersicht über den naturwissenschaftlichen Erkenntnisweg von Fragestellung bis Diskussion, in der markiert ist, welcher Schritt Schwerpunkt der Stunde ist; oder in einer Fachwissensstunde ein Advance Organizer oder eine Concept-Map, in der die schon erarbeiteten und evtl. noch zu erarbeitenden Inhalte dargestellt werden).</i>

## B Beurteilung und Feedback zur Kompetenzentwicklung

Die Lehr-Lern-Struktur der geplanten Stunde (vgl. L.) zielt darauf ab, eine spezifische Kompetenz der Lernenden zu fördern (vgl. K). Insbesondere in Sicherungs- und Übungsphasen sollte dann diagnostiziert werden, ob und ggf. inwiefern die Lernenden tatsächlich Kompetenzzuwachs erfahren haben. Es ist daher ratsam, den Lernenden zunächst anhand von Kriterien die eigene Erwartungshaltung transparent zu machen (B1), damit die Lernenden eine Richtschnur haben, an der sie sich orientieren können. Die daraufhin entwickelten Lernprodukte sollten anhand dieser Kriterien diskutiert werden, so dass die Lernenden eine Rückmeldung dazu erhalten, inwiefern sie das angestrebte Ziel erreicht haben und welche nächsten Entwicklungsschritte noch notwendig sind (B2).

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
B1	Transparenz der Leistungserwartungen		Die Leistungserwartungen in Form von Kriterien für das Lernprodukt (Inhalt, Form und Umfang) werden der Lerngruppe ... ... teilweise transparent gemacht, indem einige <u>relevante</u> formale Kriterien genannt werden, an denen sich die SuS orientieren können.	... vollständig transparent gemacht, indem relevante, <u>auf die Standardkonkretisierung bezogene</u> Kriterien genannt werden, an denen sich die SuS orientieren können.
<b>Kodierhinweise:</b>		<i>Kodieren, wenn es keine Kriterien für das Lernprodukt gibt oder diese Kriterien erst nachträglich (z. B. zur Leistungsbewertung) bekannt gemacht werden.</i>	<i>Kodieren, wenn den SuS formale Erwartungen an das Lernprodukt (z. B. ein Diagramm, ein Poster, einen Vortrag, einen Sachtext, eine Hypothese) mitgeteilt werden, das die SuS in der Erarbeitungsphase entwickeln (z. B. schriftlich auf einer PPP-Folie, als Beispiel/Muster auf einem Arbeitsblatt). Auch Kodieren, wenn auf Kriterien verwiesen wird, die in einer vorherigen Stunde erarbeitet worden sind.</i>	<i>Kodieren, wenn den SuS die kompetenzbereichsspezifische Erwartungen an das Lernprodukt (z. B. ein Diagramm, ein Poster, einen Vortrag, einen Sachtext, eine Hypothese) mitgeteilt werden, das die SuS in der Erarbeitungsphase entwickeln (z. B. schriftlich auf einer PPP-Folie, als Beispiel/Muster auf einem Arbeitsblatt). Nur kodieren, wenn die Mehrzahl der Kriterien relevant ist, um das Lernprodukt entwickeln zu können. Auch Kodieren, wenn auf Kriterien verwiesen wird, die in einer vorherigen Stunde erarbeitet worden sind.</i>

Kriterium		nicht grundständig (0 Punkte)	grundständig (1 Punkt)	erweitert (2 Punkte)
B2	Lernprodukte in Sicherungsphasen		In Sicherungsphasen werden Lernprodukte ... ... präsentiert. Dabei sind die SuS teilweise aktiv, weil einige SuS Ergebnisse präsentieren und einige SuS eine Rückmeldung geben. <del>Von der Lehrkraft erhalten sie eine Rückmeldung zu allgemeinen pädagogischen Aspekten (z. B. ihrer Mitarbeit).</del>	... präsentiert und diskutiert. Dabei sind die SuS überwiegend aktiv. <del>Von der Lehrkraft erhalten sie eine Rückmeldung zu ihrer Kompetenzentwicklung.</del>
<b>Kodierhinweise:</b>		<i>Kodieren, wenn in der Sicherungsphase nur ein Lernprodukt bzw. eine Lösung präsentiert wird, um ein richtiges Beispiel zu zeigen.</i>	<i>Kodieren, wenn in der Sicherungsphase nicht bloß die eine "richtige" Lösung präsentiert wird, sondern wenn Lernprodukte oder Lösungen von verschiedenen SuS präsentiert und die SuS aufgefordert werden, diese zu vergleichen bzw. zu diskutieren. Das Unterrichtsgespräch wird aktiv und vorwiegend von der Lehrkraft mithilfe kleinschrittiger Impulse gesteuert (s. Verlaufsplan). Falls vorhanden bezieht sich die Rückmeldung zur Arbeit der SuS auf ihre Mitarbeit, ihre Kreativität, ihren Fleiß o.ä., nicht auf den angestrebten Kompetenzzschwerpunkt.</i>	<i>Kodieren, wenn in der Sicherungsphase nicht bloß die eine "richtige" Lösung präsentiert wird, sondern wenn Lernprodukte oder Lösungen von verschiedenen SuS präsentiert und die SuS aufgefordert werden, diese zu vergleichen bzw. zu diskutieren. Die Lehrkraft überlässt den SuS einen hohen Gesprächsanteil im Unterrichtsgespräch. Nur kodieren, wenn das Unterrichtsgespräch erkennbar in Richtung des angestrebten Kompetenzzuwachses gelenkt wird (z. B. sollte in einer Erkenntnisgewinnungsstunde in der Sicherung nicht vorwiegend über das biologische Fachwissen nachgedacht werden).</i>

### Literaturverzeichnis

- Anderson, L.W. & Krathwohl, D.R. (2001): *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Harlow, London: Pearson.
- Gropengießer, H. & Kattmann, U. (2013): Didaktische Rekonstruktion. In: Harald Gropengießer, Ute Harms & Ulrich Kattmann (Hrsg.), *Fachdidaktik Biologie* (S. 16-23), 11. Auflage. Hallbergmoos: Aulis Verlag.
- Hamman, M. & Asshoff, R. (2014): *Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten*. Seelze: Klett Kallmeyer.
- Helmke, A. (2015): *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett Kallmeyer.
- Kattmann, U. (2015): *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht*. Hallbergmoos: Aulis Verlag.
- KMK (2005): *Bildungsstandards im Fach Biologie für den mittleren Schulabschluss*. Beschluss vom 16.12.2004. München, Neuwied: Luchterhand.
- KMK (2020): *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife*. Beschluss vom 18.06.2020. Bonn, Berlin: Wolters Kluwer.
- König, J., Buchholtz, C. & Dohmen, D. (2015): Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18(2), S. 375-404.
- Leisen, J. (2014): Wie soll ich meinen Unterricht planen? – Lehr-Lern-Prozesse planen am Beispiel Elektrizitätslehre in Physik. In: Uwe Maier (Hrsg.), *Lehr-Lernprozesse in der Schule: Referendariat Praxiswissen für den Vorbereitungsdienst* (S. 102-117). Stuttgart: Klinkhardt.
- Lersch, R. (2010): Didaktik und Praxis kompetenzfördernden Unterrichts. In: Katja Faulstich-Christ, Rainer Lersch & Katja Moegling (Hrsg.), *Kompetenzorientierung in Theorie, Forschung und Praxis* (S. 31-60). Immenhausen: Prolog.
- Meyer, H. (2020): *Unterrichtsmethoden I – Theorieband*, 19. Auflage. Berlin: Cornelsen.
- Weinert, F. E. (2001): Vergleichende Leistungsmessungen an Schulen. In: Franz E. Weinert (Hrsg.), *Leistungsmessungen in Schulen*. Weinheim, Basel: Beltz, S. 17-31.
- Weingarten, J. (2019): *Wie planen angehende Lehrkräfte ihren Unterricht? Empirische Analysen zur kompetenzorientierten Gestaltung von Lernangeboten*. Münster: Waxmann.

## Publikationen dieser Arbeit

Der Beitrag 1 wurde gemeinsam mit der Ko-Erstautorin konzipiert. Zur Durchführung des Reviews und zum Verfassen des Manuskripts haben beide gleichermaßen beigetragen. Daher wird für diesen Beitrag eine geteilte Erstautorenschaft angegeben. Die weiteren Ko-Autoren haben den Prozess begleitet und das Manuskript kritisch kommentiert, ehe es eingereicht worden ist. Der Drittautor hat insbesondere an der methodischen Ausrichtung des Beitrags mitgewirkt, die Fünftautorin hat insbesondere die Darstellung und Diskussion der Ergebnisse begleitet.

Die Beiträge 2, 3, 4, 5, 6, 7, und 8 wurden vom Erstautor konzipiert. Die Datenerhebung in Form der Beantragung von Staatsexamensentwürfen bei der Senatsverwaltung für Bildung, Jugend und Familie des Landes Berlin basiert auf einer Idee des Erstautors. Gemeinsam mit dem Betreuer dieser Dissertation wurde die Analyse formal beantragt. Sämtliche Analysen sind vom Erstautor durchgeführt und die Manuskripte von ihm verfasst worden. Der Zweitautor hat den Prozess begleitet und sämtliche Beiträge mehrfach kritisch kommentiert, ehe sie eingereicht worden sind. Diejenigen Beiträge, die ein Review-Verfahren durchlaufen haben (Beiträge 2, 3, 4, 5, 6 und 7) sind vom Zweitautor nach Einarbeitung der Reviewer-Kommentare durch den Erstautor erneut kritisch kommentiert worden, ehe die Beiträge erneut eingereicht worden sind.

Beitrag 9 basiert auf einer Idee des Drittautors. Der Erstautor hat den Beitrag konzipiert und das Manuskript verfasst. Die weiteren Ko-Autoren haben das Manuskript kritisch kommentiert, ehe es eingereicht worden ist.

Die Idee zur Herausgabe eines Mystery-Hefts wurde gemeinsam vom Erst- und Zweitherausgeber entwickelt. Die Beiträge 10 und 11 wurden vom Verfasser dieser Arbeit allein konzipiert und verfasst. Beitrag 12 basiert wie dort auf S. 45 angegeben auf einer Idee zweier Studierender, die dem Erst- und Zweitautor die Genehmigung erteilt haben, ihre Idee weiterzuentwickeln und zu publizieren. Der Erstautor hat die Materialien grundlegend überarbeitet, der Zweitautor hat kritisch kommentiert.

Beitrag 13 basiert auf einer Idee des Erstautors, der das Manuskript konzipiert und verfasst hat. Die Zweitautorin hat die Fischschwarmsimulation und die Idee für die darin zu berücksichtigenden Variablen entwickelt. Die Drittautorin hat die Programmierung in NetLogo® unterstützt und das Manuskript kritisch kommentiert.

Beitrag 14 basiert auf der Idee der Erstautorin. Der Verfasser dieser Dissertation ist als Zweitautor beteiligt. Er hat die Erstautorin dabei unterstützt, das Thema in einen unterrichtsrelevanten Kontext zu kleiden, und das Manuskript mehrfach kritisch gelesen und kommentiert.

## Beitrag 1

**Großmann, L.**, Koberstein-Schwarz, M., Scholl, D., Krüger, D. & Meisert, A. (eingereicht). Establishing common ground in empirical research on science teachers' lesson planning – A scoping review. [geteilte Erstautorenschaft]



1 **Establishing Common Ground in Empirical Research on Science Teachers'**

2 **Lesson Planning: A Scoping Review**

3 Leroy Großmann<sup>a\*</sup> (ORCID: 0000-0001-7635-1737),

4 Maren Koberstein-Schwarz<sup>b</sup> (ORCID: 0000-0003-0522-8189),

5 Daniel Scholl<sup>c</sup> (ORCID: 0000-0003-3403-9811),

6 Dirk Krüger<sup>a</sup> (ORCID: 0000-0003-0999-4382), and

7 Anke Meisert<sup>b</sup> (ORCID: 0000-0002-7671-5075)

8

9 <sup>a</sup>Department of Biology Education, Freie Universität Berlin, Berlin, Germany

10 <sup>b</sup>Department of Biology, Universität Hildesheim, Hildesheim, Germany

11 <sup>c</sup>Department of Educational Science, Universität Siegen, Siegen, Germany

12 *Leroy Großmann and Maren Koberstein-Schwarz contributed equally to this work and*  
13 *should be considered joint first authors.*

14 *Dirk Krüger and Anke Meisert should be considered joint senior authors.*

15 \*corresponding author: [leroy.grossmann@fu-berlin.de](mailto:leroy.grossmann@fu-berlin.de)

16

17 In case of acceptance, we kindly ask you to recommend the following citation of the paper, so that Leroy

18 Großmann and Maren Koberstein-Schwarz are both named: Großmann, Koberstein-Schwarz et al. (2023).

19

20

21

22

23

## 24 **Abstract**

25 Although lesson planning forms a core part of science teachers' professional expertise,  
26 research on science teachers' lesson planning is still relatively scarce. There is neither a  
27 heuristic model describing which knowledge and skills are necessary and illustrating their  
28 interplay in lesson planning processes nor an overview of this emerging field of research  
29 consolidating current findings and proposing trajectories for further investigation. To  
30 close this gap, we conducted a scoping review of  $N = 66$  empirical studies in science  
31 education research investigating science teachers' lesson planning between 1987 and  
32 2022. We find that most studies target the identification of conceptual knowledge rather  
33 than investigating science teachers' procedural knowledge in lesson planning processes.  
34 Moreover, most studies focus on the creation of mental or written lesson plans, and only  
35 a few additionally consider the justification of planning decisions. An analysis of the  
36 bibliographic coupling among the reviewed studies reveals that researchers in the field do  
37 not yet relate to a common theoretical ground. Hence, we summarize our findings in a  
38 modified version of the Refined Consensus Model of PCK specified for lesson planning,  
39 which might facilitate research on science teachers' lesson planning.

## 40 **Keywords**

41 lesson planning, science teacher education, Refined Consensus Model of PCK,  
42 scoping review

## 43 **Introduction**

44 Lesson planning forms a core part of science teachers' professional expertise (Carlson et  
45 al., 2019). It is assumed that lesson planning has an impact on the quality of teaching and,  
46 thus, on student outcomes (Darling-Hammond et al., 2013; König, Blömeke et al., 2021).

47 As there is a consensus that lesson planning is teachable and learnable (König et al.,  
48 2015), the development of knowledge and skills for lesson planning is an essential goal  
49 of teacher education (e.g., Australia: Australian Institute for Teaching and School  
50 Leadership [AITSL], 2018; Estonia: Kutsekoda, 2013; France: Ministère de l'Éducation  
51 Nationale [MEN], 2013; Germany: Ständige Konferenz der Kultusminister der Länder in  
52 der Bundesrepublik Deutschland [KMK], 2019; Ireland: Céim, 2020; Norway: National  
53 Council for Teacher Education [NRLU], 2016) realized in non-formal (e.g., internships)  
54 as well as formal (e.g., lectures) learning opportunities (Kleickmann et al., 2013).

55 Due to its importance, the first research on lesson planning was carried out as long as 50  
56 years ago. In the 1970s and 1980s, many researchers investigated teachers' lesson  
57 planning almost exclusively from a generic cognitive psychological perspective. This  
58 research focused on decision-making in the planning processes (e.g., Morine-Dersheimer,  
59 1979; Yinger, 1980; Zahorik, 1975). This focus waned in the following years and then  
60 resumed in the last 15 years from a subject-specific perspective (König & Rothland,  
61 2022). This subject-specific orientation results from the undisputed idea that a teacher's  
62 professional knowledge and skills are partly subject-specific, based on Shulman's (1987)  
63 notion of Pedagogical content knowledge (PCK). Since then, research has focused on the  
64 knowledge resources necessary for lesson planning. However, perspectives on what  
65 constitutes knowledge and skills for lesson planning and how it can be measured vary  
66 significantly. For instance, knowledge resources for planning are reconstructed from  
67 planning artifacts; planning skills are described in terms of strategies in planning  
68 processes; or performance tests are used to measure knowledge and skills. However,  
69 despite the consensus that lesson planning is one of teachers' core activities and the  
70 increasing amount of research in this area, there is no overview of this field of research  
71 that consolidates current findings and proposes trajectories for further research.

72 Consequently, there are no models of (science) teachers' lesson planning illustrating the  
73 interplay of knowledge and skills of (science) lesson planning supported by empirical  
74 evidence. Current models of general teacher profession do not yet represent these  
75 components of lesson planning in detail but address holistic ideas of teachers'  
76 professional competence (e.g., Kunter et al., 2013) or PCK (e.g., Carlson et al., 2019)

77 One fruitful approach might be the adaptation of the Refined Consensus Model of PCK  
78 (RCM; Carlson et al., 2019), which maps a teacher's subject-specific knowledge and skills  
79 for teacher action. However, a model of lesson planning primarily requires further  
80 specification of the strategies needed for this purpose since successful lesson planning is  
81 not realized by knowledge resources alone but, for example, by their integration  
82 (Westerman, 1991) within an iterative (Yinger, 1980) planning process of creation and  
83 justification (Vogelsang & Riese, 2017).

84 Accordingly, the purpose of this scoping review is threefold: First, we aim to examine  
85 the extent, range, and nature of research activities concerning science teachers'  
86 knowledge and skills for lesson planning. Second, we aim to map how knowledge and  
87 skills for lesson planning are measured in empirical research. Third, we aim to examine  
88 factors influencing science teachers' lesson planning. As a heuristic theoretical lens, we  
89 will use the RCM to deduce the categories we will apply in the coding process of the  
90 included articles. Finally, the results will be summarized in a modified version of the  
91 RCM specified for lesson planning. This will establish a tentative model of science  
92 teachers' lesson planning model. Since the field is still unsystematic, a scoping review is  
93 a suitable approach, including all studies that examine the dependence of lesson planning  
94 on another variable (e.g., knowledge, beliefs). In doing so, we include all studies  
95 matching this focus, regardless of whether the authors explicitly targeted lesson planning  
96 or aimed at capturing knowledge resources (e.g., PCK) in written lesson plans. Thus, we

97 can provide an extensive overview of all currently available empirical research on science  
98 teachers' lesson planning and suggest further research trajectories in the field.

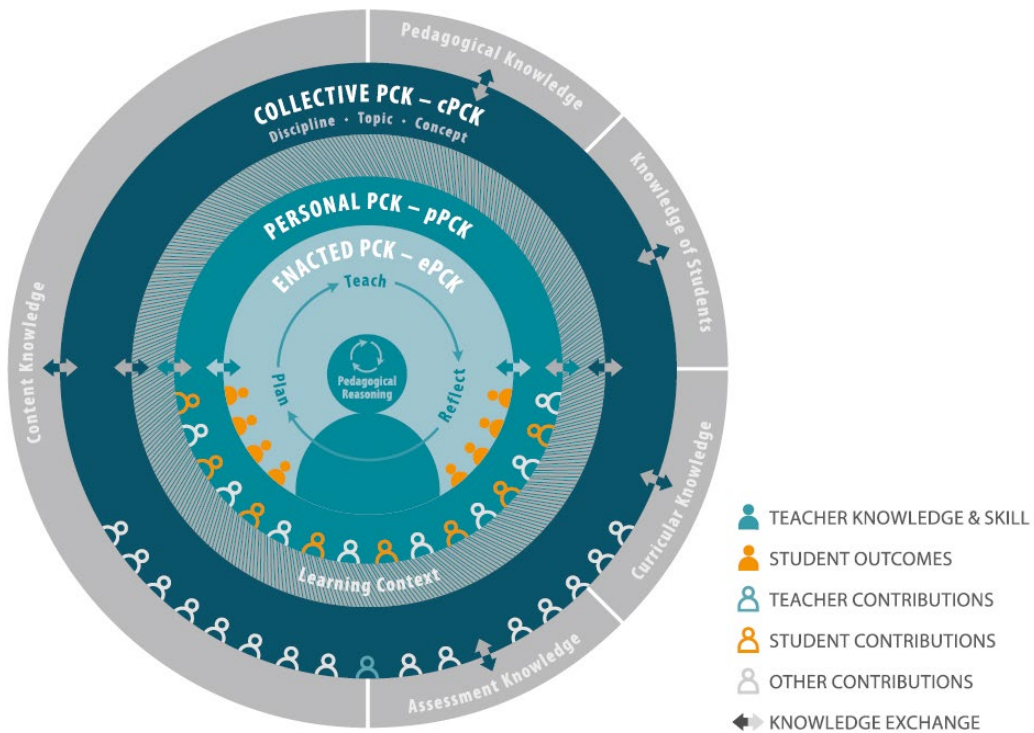
## 99 **Conceptual framework**

### 100 **Teachers' professional knowledge**

101 Shulman defines PCK as a “special amalgam of content and pedagogy that is uniquely  
102 the province of teachers” (Shulman, 1987, p. 8). In doing so, he distinguishes PCK from  
103 general *Pedagogical Knowledge* (PK, e.g., classroom management) on the one hand and  
104 from subject-specific *Content Knowledge* (CK, e.g., knowledge of photosynthesis,  
105 chemical bonds, gravity) on the other. Meanwhile, empirical research has repeatedly  
106 confirmed Shulman's assumption that PK, CK, and PCK are distinct knowledge bases  
107 (e.g., Großschedl et al., 2015), even if they influence each other (Chan & Hume, 2019;  
108 Kirschner et al., 2016). Accordingly, PCK is conceptualized as a unique knowledge base  
109 interacting with CK and PK in the RCM (Figure 1; Carlson et al., 2019). According to  
110 this model, PCK is further differentiated in discipline- (e.g., knowledge of inquiry-based  
111 learning in science), topic- (e.g., knowledge of students' ideas about photosynthesis), or  
112 concept-specific PCK (e.g., knowledge of students' understanding of energy flow;  
113 Carlson et al., 2019; Mavhunga, 2020).

### 114 **Figure 1**

115 *The Refined Consensus Model of PCK (Carlson et al., 2019, p. 83)*



116

117 The RCM distinguishes between three realms of PCK, with knowledge flowing between  
 118 them): collective PCK (cPCK), personal PCK (pPCK), and enacted PCK (ePCK). cPCK  
 119 is the shared knowledge of a community of researchers or educators in science education.  
 120 Accordingly, cPCK is the “knowledge [...] generated through science education research  
 121 or best practice examples developed or approved by a professional community“ (Sorge,  
 122 Stender, & Neumann, 2019, p. 151). This knowledge is shared throughout teacher  
 123 education and ”is, in contrast to a person’s pPCK, not necessarily influenced by their  
 124 personal experiences” (Irmer et al., 2023, p.3).

125 Based upon cPCK, teachers develop their unique pPCK. A teacher’s pPCK “serves as a  
 126 reservoir of knowledge and skills that a teacher can draw upon during the practice”  
 127 (Carlson et al., 2019, p. 85). Therefore, pPCK is an individual teacher’s personal  
 128 knowledge applicable in action. This action is shown in the innermost circle of the RCM.  
 129 In action, teachers generate ePCK through pedagogical reasoning (Carlson et al., 2019).  
 130 As there are three significant practices in the teaching cycle, Alonzo et al. (2019)

131 distinguish three forms of ePCK: ePCK<sub>planning</sub> (ePCK<sub>p</sub>), ePCK<sub>teaching</sub> (ePCK<sub>t</sub>), ePCK<sub>reflecting</sub>  
132 (ePCK<sub>r</sub>)).

133 To specify the difference between individual cPCK, pPCK, and ePCK, amplifiers and  
134 filters are assumed between these two realms of knowledge that mediate the  
135 transformation of cPCK to pPCK and pPCK to ePCK (Nilsson & Karlsson, 2019).  
136 Amplifiers and filters can be divided into *personal* and *extra-personal* factors (Henze &  
137 Barendsen, 2019). Personal factors are affective dispositions like orientations and beliefs  
138 (Carlson et al., 2019; Gess-Newsome, 2015; Henze & Barendsen, 2019) or personal  
139 experiences (Sorge et al., 2019). Extra-personal factors include elements of the setting,  
140 such as school context, lesson planning guides, or curriculum material (Barendsen &  
141 Henze, 2019; Behling et al., 2022; Irmer et al., 2023).

142 Thus, the RCM is a holistic model of teachers' professional knowledge and skills,  
143 mapping cognitive (e.g., CK, PK, cPCK) and affective dispositions (e.g., amplifiers and  
144 filters). Furthermore, the RCM considers the circumstances under which teachers' action  
145 is placed (learning context) and describes situation-specific skills (ePCK, pedagogical  
146 reasoning). However, as the RCM illustrates teachers' PCK in general, it does not  
147 concretize teachers' PCK and its interdependence with other variables during planning,  
148 teaching, and reflecting. Thus, the RCM provides a solid basis for a model of science  
149 teachers' knowledge and skills for lesson planning, which needs to be further specified.

150

### 151 **Teachers' lesson planning**

152 Lesson planning aims to design a sequence of learning activities to help students with  
153 different prerequisites to achieve a specific learning goal (Gunckel, 2011; Rusznyak &  
154 Walton, 2011). This sequence of learning activities, named lesson plan, is a written or  
155 mental product (Ball et al., 2007) of an iterative lesson planning process (John, 2006;

156 Karlström & Hamza, 2021; Yinger, 1980). In the planning process, teachers have various  
157 resources at their disposal. In most European countries, planning is left up to the teacher,  
158 and teachers use the internet, textbooks, or colleagues as inspiration for lesson plans or  
159 materials (Pepin et al., 2013). In this case, planning can be described as a material-based  
160 design process (Koberstein-Schwarz & Meisert, 2022) guided by superordinate curricular  
161 requirements. In the U.S., on the other hand, planning based on curriculum materials is  
162 common. Curriculum material often includes lesson plans and materials, which need to  
163 be analysed and adapted in the planning process (e.g., Davis, 2006). Nonetheless, in both  
164 cases of lesson planning, teachers make decisions by interconnecting different aspects  
165 relevant to lesson planning, such as learning objectives, activities, or tasks in an iterative  
166 process (König, Krepf et al., 2021; Rusznyak & Walton, 2011; Yinger, 1980). To do so,  
167 teachers integrate their professional knowledge in a process called pedagogical reasoning  
168 (Alonzo et al., 2019; Koberstein-Schwarz & Meisert, 2022; Shulman, 1987; Starkey,  
169 2010; Westerman, 1991). Pedagogical reasoning in lesson planning is divided into  
170 creation (e.g., selecting or designing a representation or a task) and justification (e.g.,  
171 using criteria such as the fit of the representation to the student's prerequisites to  
172 substantiate the planning decision), which form an iterative and circular process (Ball et  
173 al., 2007; Vogelsang & Riese, 2017; Westerman, 1991; Yinger, 1980). In pedagogical  
174 reasoning, teachers make use of knowledge that Shavelson et al. (2005) locate on different  
175 levels: Teachers need both conceptual knowledge (*knowing that*) as well as *procedural*  
176 *knowledge (knowing how)*. *Knowing that* may be specified by the PCK facets, according  
177 to Magnusson et al. (1999). According to that, teachers need *Orientation towards Science*  
178 *Teaching (OST)*, *Knowledge of Science Curriculum (KSC)*, *Knowledge of Students'*  
179 *Understanding in Science (KSU)*, *Knowledge of Instructional Strategies and*  
180 *Representations (KISR)*, and *Knowledge of Assessment of Science Learning (KAs)*. The



181 latter has hardly been specified in the PCK discourse, but essential examples of  
182 procedural knowledge for planning emerge from the discourse on lesson planning (e.g.,  
183 Großmann & Krüger, 2022; Koberstein-Schwarz & Meisert, 2022). For instance,  
184 knowing how to integrate the different PCK facets in the creation and justification of  
185 planning decisions (e.g., Westerman, 1991) or knowing how to interconnect planning  
186 decisions in other planning areas (e.g., Scholl et al., 2022) are essential skills in lesson  
187 planning. Therefore, reconstructing lesson planning in front of the background of the  
188 RCM, ePCK<sub>p</sub> is the knowledge generated by creation and justification in terms of  
189 pedagogical reasoning using available conceptual and procedural pPCK.

190 Although the literature describes the activities and demands of lesson planning (e.g.,  
191 König, Krepf et al., 2021), there is no consistent model of science teachers' lesson  
192 planning researchers in the field can refer to (Rothland, 2021). Different studies examine  
193 the use of conceptual knowledge in the creation of planning decisions (e.g., Forbes &  
194 Davis, 2010; Hanuscin et al., 2018) or strategies of their justification (e.g., Gunckel, 2011;  
195 Kang, 2017). However, these studies remain unconnected to each other. So far, no review  
196 relates these studies to each other and transfers results into a science-specific tentative  
197 lesson planning model.

198 Evidence about teachers' knowledge and skills for lesson planning is provided in studies  
199 with a designated research focus on lesson planning and in all studies investigating  
200 teachers' PCK in lesson planning processes or products. To gain an overview of the  
201 research field of science teachers' lesson planning and to specify the RCM as a model of  
202 science teachers' knowledge and skills for lesson planning from the literature, we need to  
203 include (1) studies that have a designated focus on lesson planning and (2) studies that  
204 analyse PCK using lesson planning products or processes as a data source.

205 **Research focus**

206 In light of the recently increasing research activity in the field of lesson planning and  
207 knowing that there is no review of this branch of research in science education yet, the  
208 time is ripe to consolidate the current state of knowledge and establish common ground  
209 in empirical research on science lesson planning. More specifically, gaining an overview  
210 of the extent, range, and nature of research activities concerning (pre-service) science  
211 teachers' lesson planning and summarizing current research findings seems crucial to  
212 develop a tentative lesson planning model for science teachers and identifying research  
213 gaps and, thus avoid repetitive or unfocussed research attempts in the future.

214 Two observations have guided our review process: First, researchers use various terms,  
215 methodological approaches, and conceptualizations of lesson planning. This  
216 heterogeneous nature of theoretical and empirical approaches challenges the  
217 comparability of research findings. Therefore, it is necessary to gain an overview of what  
218 approaches are prevalent in current research on lesson planning. For that purpose, the  
219 present review examines studies in which ePCK<sub>p</sub> or pPCK in lesson planning are  
220 examined as a variable depending on another influencing factor to develop a tentative  
221 lesson planning model. Based upon that, suggestions for establishing common ground  
222 concerning the conceptualization of lesson planning can be made. Second, as learning to  
223 plan lessons is a crucial purpose of teacher education that needs to be learned and  
224 enhanced in teachers' professional development, we were especially interested in what  
225 learning opportunities are employed to help science teachers develop knowledge and  
226 skills for lesson planning.

227

## 228 **Methodology of Scoping Reviews**

229 Examining the extent, range, and nature of research activity on science teachers'  
230 knowledge and skills for lesson planning and summarizing current research findings as a  
231 basis for a tentative lesson planning model requires mapping the research landscape and  
232 providing an overview of the studies investigating lesson planning. The field is still  
233 emerging. Hence we decided to conduct a scoping review (Peters et al., 2020), which is  
234 the appropriate methodology to target the abovementioned purposes (Arksey & O'Malley,  
235 2005). As “a scoping review of a body of literature can be of particular use when the topic  
236 has not yet been extensively reviewed or is of a complex or heterogeneous nature” (Pham  
237 et al., 2014, p. 371), this choice seems appropriate in the context of lesson planning  
238 research. To our knowledge, no literature review on lesson planning research in science  
239 education has been published yet. Moreover, the inconsistent use of terminology (e.g.,  
240 lesson planning competence, lesson planning skills, ability to plan lessons) and the  
241 absence of a general or science-specific consensus model of lesson planning make it  
242 necessary to summarize current research and to show research perspectives.

243 Among the different types of literature reviews, scoping reviews are a relatively recently  
244 developed methodology (Arksey & O'Malley, 2005; Colquhoun et al., 2014; Pham et al.,  
245 2014) that is still rare in science education. They are defined as “a form of knowledge  
246 synthesis that addresses an exploratory research question aimed at mapping key concepts,  
247 types of evidence, and gaps in research related to a defined area or field by systematically  
248 searching, selecting, and synthesizing existing knowledge” (Colquhoun et al., 2014, pp.  
249 1292-1294). Hence, scoping reviews are guided by the intention to cover a comprehensive  
250 breadth of research literature, yet they might differ in the degree of depth to extract  
251 information from the included studies (Arksey & O'Malley, 2005).

252 Even though they are conducted as rigorously and transparently as the more common  
253 methodology of systematic literature reviews, they differ in three significant  
254 characteristics (Munn et al., 2018; Pham et al., 2014): First, unlike systematic literature  
255 reviews, scoping reviews do not aim to summarize the *best* available evidence for a  
256 specific question but to comprehensively identify and map *all* available studies related to  
257 a particular question. Second, scoping reviews investigate broader research questions to  
258 determine the coverage of a specific body of literature regarding a particular topic. In  
259 contrast, systematic literature reviews are conducted to answer specific questions and  
260 confirm or refute an assumption. Third, scoping reviews do not critically appraise the  
261 quality of the identified studies and thus include all studies matching the inclusion criteria,  
262 notwithstanding potential biases or methodological deficiencies. Therefore, scoping  
263 reviews can be regarded as a precursor of systematic literature reviews. Beyond  
264 systematic literature reviews, traditional narrative literature reviews could be an option  
265 for our purpose. However, we opted against a narrative review since this methodology  
266 heavily relies on the authors' subjective and thus biased knowledge and experience,  
267 which subsequently limits the replicability and generalizability of the results. Since we  
268 aim to establish common ground in research on lesson planning and clarify concepts and  
269 terminology, we aimed to conduct a transparent and highly structured literature review.  
270 In a nutshell, agreeing with Deehan et al. (2022), the scoping review methodology  
271 provides an appropriate balance between highly structured reviews (i.e., systematic  
272 literature reviews) on the one hand and less structured open types of reviews (i.e.,  
273 narrative reviews) on the other. Consequently, the scoping review methodology matches  
274 the purposes of the present paper.

275 In the few existing publications that apply the scoping review methodology in science  
276 education (e.g., Deehan et al., 2022; Jack & Lin, 2017; Kafai et al., 2022), the authors

277 unanimously justify their decision in favor of a scoping review by emphasizing its  
278 potential to characterize research activity in a particular field on a relatively broad level  
279 and, thus, gain an overview about current trends and open questions. In particular, Jack  
280 and Lin (2017) refer back to the original proposition of a five-stage process for conducting  
281 a scoping review by Arksey and O'Malley (2005) and emphasize “its ability to provide a  
282 rigorous, transparent and stand-alone methodology for comprehensively identifying and  
283 analysing all relevant literature about a research question and determining whether a full  
284 systematic review of literature is feasible or necessary” (Jack & Lin, 2017, p. 144):

285       Stage 1: Identifying the research question;

286       Stage 2: Identifying relevant studies;

287       Stage 3: Study selection;

288       Stage 4: Charting the data;

289       Stage 5: Collating, summarising, and reporting the results.

290 As Colquhoun et al. (2014) proposed, we will rigorously adhere to this structure  
291 throughout the present paper instead of the established chapter structure (research  
292 questions, methods, results, and discussion). In doing so, we followed the “Preferred  
293 Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping  
294 Reviews” (PRISMA-ScR) checklist (Tricco et al., 2018), which provides guidelines for  
295 conducting and reporting scoping reviews.

## 296 **Stage 1: Identifying the research questions**

297 Despite the repeatedly expressed claim that research on lesson planning is relatively  
298 scarce (e.g., Campbell et al., 2022, p. 109; Koberstein-Schwarz & Meisert, 2022, p. 2;  
299 König, Krepf et al., 2021, p. 467), the authors of this paper took notice of studies that do  
300 explicitly or implicitly refer to science teachers’ lesson planning. This raises the question  
301 of to what extent researchers in science education consider lesson planning in their

302 research, what common themes emerge from their findings, and what research gaps might  
303 need to be filled. More specifically, if researchers in this field barely relate to each other  
304 and hardly take note of other researchers' findings regarding lesson planning, the time is  
305 ripe to consolidate the current state of evidence, map research trajectories, and establish  
306 common ground in research on lesson planning. Hence, this review aims to examine the  
307 extent, range, and nature of research activities concerning teachers' lesson planning in  
308 science education and summarize current research findings. Due to the absence of  
309 previous literature reviews on lesson planning, we decided to pursue a threefold target:  
310 First, we aim to examine the extent, range, and nature of research activities concerning  
311 science teachers' knowledge and skills for lesson planning (RQ 1):

312 **RQ 1:** What is the state of science education research on lesson planning  
313 regarding demographics, methodological approaches, and bibliographic coupling?

314 Second, we aim to map how knowledge and skills for lesson planning are  
315 measured in empirical research using RCM as a heuristic theoretical lens. More  
316 specifically, we wonder which aspects of this complex construct were measured in the  
317 related studies and which data sources were investigated, leading to another  
318 methodological research question (RQ 2):

319 **RQ 2:** Which aspects of science teachers' knowledge and skills for lesson  
320 planning regarding the RCM are investigated in science education research?

321 Third, we aim to examine factors influencing science teachers' lesson planning.  
322 Therefore, we focus on studies in which lesson planning forms the dependent variable.  
323 Thus, we aim to identify which independent variables are investigated in empirical  
324 research and what effects on knowledge and skills for lesson planning are described. In  
325 terms of the RCM, these variables relate to the amplifiers and filters, which are  
326 symbolized as double arrows in Figure 1 (RQ 3):

327           **RQ 3:** Which variables influence science teachers' knowledge and skills for  
328 lesson planning?

### 329 **Stage 2: Identifying relevant studies**

330 Due to the lack of a connecting theory or model, authors use various terms to refer to  
331 lesson planning and science teachers' knowledge and skills for lesson planning. As we  
332 aim to identify all studies related to lesson planning, we were required to reflect the  
333 heterogenous terminology during the data search. To meet this challenge, we build a  
334 search string comprising three significant topics: lesson planning, teacher education, and  
335 science. The operator AND connected the three following parts of the search string in the  
336 database search. The letters "ab" specified that we searched the abstracts, and the letters  
337 "ti" specified that we explored the titles. The asterisk \* indicates that all words will be  
338 searched that start with the same letters and vary in their endings (e.g., adapt\* comprises  
339 "adapting" and "adaptation"). We decided to search in the databases ERIC and SAGE.  
340 Using ERIC, we accessed the most significant database for education and teacher training  
341 publications. We intended to identify publications in life science or social science journals  
342 using SAGE. We used the search strings listed below in both databases.

#### 343 ***Lesson planning***

344 Depending on researchers' perspectives on lesson planning, they use specific terms to  
345 refer to it. Consequently, searching for "lesson planning" only would entail the risk of  
346 missing studies that apply different terms, such as "lesson preparation" (e.g., van der Valk  
347 & Broekman, 1999) or "adapt and critique science curriculum materials" (e.g., Beyer &  
348 Davis, 2012), evidently an important aspect of lesson planning without mentioning these  
349 exact words. Hence, the first part of the search string was developed:

350

351 ((ab(plan OR planning OR "lesson proposal" OR "lesson preparation" OR (adapt\* AND  
352 "curriculum materials")) OR (critique\* AND "curriculum materials")) OR ti(plan OR  
353 planning OR "lesson proposal" OR "lesson preparation" OR (adapt\* AND "curriculum  
354 materials")) OR (critique\* AND "curriculum materials"))  
355

### 356 ***Teacher Education***

357 In an initial trial search, we found many articles about planning in various contexts, such  
358 as urban planning. To focus the search on lesson planning, we were required to develop  
359 a second part of the search string that explicitly refers to lesson planning in the context of  
360 teacher education. Research on lesson planning has been conducted since the 1970s.  
361 Research paradigms develop, thus resulting in new perspectives researchers take on a  
362 topic. For instance, early research mainly focused on teachers' decision-making (e.g.,  
363 Shavelson et al., 1977; Westerman, 1991), an essential aspect of lesson planning.  
364 Furthermore, researchers investigating professional teachers' expertise often refer to  
365 pedagogical reasoning (e.g., Loughran et al., 2016; Penso & Shoham, 2003; Shulman,  
366 1987; Zaragoza et al., 2021), that is the ability to use professional knowledge to justify  
367 planning decisions, or to specific aspects of professional knowledge, such as pedagogical  
368 content knowledge or content knowledge (e.g., Koberstein-Schwarz & Meisert, 2022).  
369 Hence, the second part of the search string was developed:

370 (ab(lesson OR teach\* OR instruction\* OR "pedagogical reasoning" OR "decision making"  
371 OR "pedagogical content knowledge" OR "lesson study" OR "content knowledge" OR  
372 "curriculum")) OR ti(lesson OR teach\* OR instruction\* OR "pedagogical reasoning" OR  
373 "decision making" OR "pedagogical content knowledge" OR "lesson study" OR "content  
374 knowledge" OR "curriculum"))

### 375 ***Science***

376 To limit the scope of the search to science teacher education, we developed a third part  
377 of the search string that clarifies the context in which the planning occurs: science teacher  
378 education. As science teacher programs vary significantly among countries (Pedersen et  
379 al., 2017), we specified subjects. For instance, some countries separate biology,



380 chemistry, physics, and geography from each other (e.g., Chile: Cofré et al., 2015), and  
381 others have both depending on the state (USA: Olson et al., 2015). Still, others have  
382 integrated science or STEM in lower classes and biology, chemistry, and physics as  
383 specialist subjects in higher classes (e.g., England: Evagorou et al., 2015; Germany:  
384 Neumann et al., 2017). Hence, the following search string was developed:

385 (ab("science education" OR "elementary science" OR "science teach\*" OR "science student  
386 teacher\*" OR "science preservice teacher\*" OR "STEM" OR biolog\* OR physics OR  
387 chemistry OR geography\* OR "inquiry-based" OR "SSI" OR "socioscientific issues" OR  
388 "socio-scientific issues") OR ti("science education" OR "elementary science" OR "science  
389 teach\*" OR "science student teacher\*" OR "science preservice teacher\*" OR "STEM" OR  
390 biolog\* OR physics OR chemistry OR geography\* OR "inquiry-based" OR "SSI" OR  
391 "socioscientific issues" OR "socio-scientific issues"))  
392

393 We applied these three search strings in an initial database search and noticed that some  
394 papers the two joint first authors had come across before were not among the results.  
395 These papers were published in highly recognized science education journals. As we  
396 aimed to include as many research papers as possible investigating science teachers'  
397 lesson planning, we opted against searching exclusively for the most recognized science  
398 education journals. However, we were open to include papers published in any  
399 international peer-reviewed journal. We conducted an additional journal-specific  
400 database search in ERIC to ensure that we did not miss any papers that we happened to  
401 know before and were published in highly recognized science education journals. The  
402 first two search strings were used as described before, but in place of the context search  
403 string, we used the JN operator to search the following science education journals  
404 specifically:

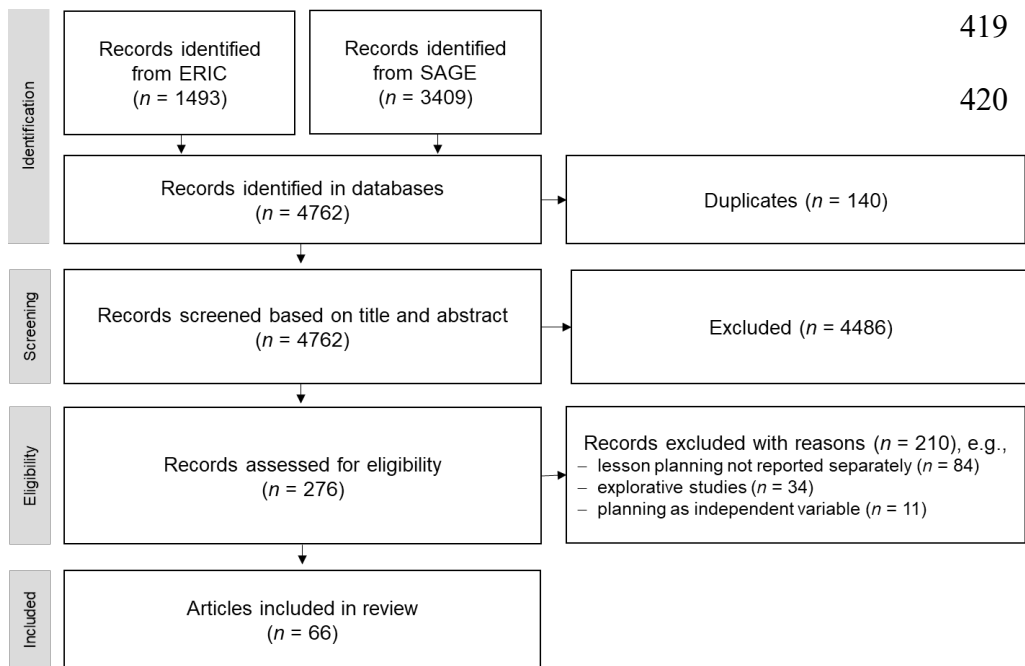
405 JN("Electronic Journal of Science Education" OR "Journal of Research in Science Teaching"  
406 OR "Science Education" OR "International Journal of Science Education" OR "Research in  
407 Science Education" OR "Journal of Science Teacher Education" OR "Science & Education"  
408 OR "International Journal of Science and Mathematics Education" OR "Chemistry  
409 Education Research and Practice" OR "Eurasia Journal of Mathematics, Science and  
410 Technology Education" OR "Research in Science and Technological Education"))

411 Using that, we identified all previously missing papers. We were confident enough that  
412 the relevant studies were identified, even though we cannot claim that the search is

413 entirely unbiased and exhaustive (Arksey & O'Malley, 2005). In total, 4902 records were  
 414 identified in databases (Figure 2). The duplicates ( $n = 140$ ) were excluded, and the  
 415 remaining records ( $n = 4762$ ) were screened based on title and abstract.

416 **Figure 2**

417 *Flowchart of the search procedure and data analysis used for the systematic literature*  
 418 *review, following PRISMA (Moher et al., 2009).*



421 **Stage 3: Study selection**

422 To ensure that the articles included in the review meet the abovementioned research  
 423 purposes and to make the process transparent, we defined inclusion and exclusion criteria  
 424 (Table 1) and screened the abstracts and titles. The criteria reflect the targets we pursued  
 425 by our research questions; that is, we only included studies that investigate elementary or  
 426 secondary (*educational stage*) science teachers' (*science education*) lesson planning as  
 427 the dependent variable (*focus*) in empirical articles (*design*) published in peer-reviewed  
 428 journals (*sources*) between 1987 and 2022 (*time*) in English (*language*). To ensure a  
 429 substantial range of literature relating to the topic, we extended the search back to 1987,

430 after Shulman introduced the notion of PCK in his seminal papers (Shulman, 1986;  
 431 Shulman, 1987). In doing so, we can trace the research on teachers' professional  
 432 development from the beginning to when we searched databases (June 17<sup>th</sup>, 2022). It has  
 433 not escaped our notice that lesson planning research has increased in some countries for  
 434 the last five to ten years. Even though some of the recently published empirical papers  
 435 (e.g., Großmann & Krüger, 2022; Schröder et al., 2020) might contribute to fruitfully  
 436 answering our research questions, we decided to focus solely on papers published in  
 437 English. It would counteract our intention to establish common ground in empirical  
 438 research on lesson planning if we included studies that most international science  
 439 educators would not be able to receive.

440 **Table 1**

441 *Inclusion and exclusion criteria.*

<b>Criteria</b>	<b>Studies included</b>	<b>Studies excluded</b>
Educational stage	elementary school, secondary school	preschool, kindergarten, higher education
Science Education	the sample is related to science education	the sample is not related to science education
Focus	lesson planning as the dependent variable	lesson planning as an independent variable, explorative studies
Design	empirical studies	theoretical studies
Sources	journal articles	dissertations, grey literature
Time	1987 – 17 <sup>th</sup> June 2022	before 1987 and after 17 <sup>th</sup> June 2022
Language	English	other language than English
Presentation	Lesson planning is separately reported in the results	Lesson planning is not separately reported in the results
Science-specificity	focus on science-specific facets of the RCM (i.e., PCK, CK)	focus on generic aspects of the RCM (i.e., PK, beliefs)

442

443 The two joint first authors independently applied the inclusion and exclusion criteria to a  
 444 randomly selected 100 papers to ensure intersubjective comprehensibility. A few

445 uncertainties and disagreements were discussed. Afterward, the remaining 4662 papers  
446 were equally assigned to the joint first authors and either included, excluded, or marked  
447 as uncertain. Uncertain cases were discussed once an agreement was reached. Finally,  
448 276 studies remained after the screening process (Figure 1). These studies were assessed  
449 for eligibility by reading the full texts and checking whether all inclusion criteria or at  
450 least one exclusion criterion were met. Again, both joint first authors divided the sample  
451 up. Uncertainties were discussed until an agreement was reached. We noticed that even  
452 though many studies referred to lesson planning (e.g., using lesson plans as a data source  
453 for analysing PCK), their focus was not on science teachers' lesson planning as the  
454 authors did not explicitly refer to lesson planning in the results section ( $n = 84$ ;  
455 presentation, Tab. 1). Moreover, we identified studies ( $n = 39$ ) that explored planning  
456 processes or products without linking lesson planning to another variable (*focus*, Tab. 1).  
457 Another set of studies ( $n = 11$ ) investigated the effects of lesson planning on other  
458 variables, such as student outcomes (*focus*, Tab. 1). This is not the focus of the present  
459 scoping review but might hint at the so far barely investigated role of lesson planning as  
460 an independent variable within the teaching cycle (Alonzo et al., 2019; Carlson et al.,  
461 2019). The remaining papers were excluded because at least one of the other exclusion  
462 criteria was met. One paper was not accessible and was therefore excluded from the  
463 further analysis. Finally,  $N = 66$  studies were included in the scoping review.

#### 464 **Stage 4: Charting the data**

465 Arksey and O'Malley (2005) define charting as “a technique for synthesizing and  
466 interpreting qualitative data by sifting, charting and sorting material according to key  
467 issues and themes” (p. 26). In this process, we used the Microsoft software Excel. We  
468 gathered general (e.g., authors, year of publication, journal, country) and specific

469 information (e.g., sample size, research design, variables, key findings) in a data sheet.

470 The following information were recorded:

- 471 • Authors, study title, journal, year of publication, study location (country)
- 472 • Sample size, participants' experience (pre-service, in-service), school type  
473 (elementary, secondary), subject (science, STEM, biology, chemistry, physics,  
474 geography)
- 475 • Number of data collections (cross-sectional, longitudinal), design (non-  
476 experimental, quasi-experimental, experimental)
- 477 • Data source (self-report, performance), lesson topic, authenticity (simulated, in  
478 situ), planning situation (individual planning, collaborative planning), focus  
479 (planning process, planning product), planning activity (creation, justification)
- 480 • Indicators to capture ePCK<sub>p</sub> (conceptual knowledge: PCK-facets, CK; procedural  
481 knowledge: planning strategies); only coded for ePCK<sub>p</sub> investigated in  
482 performance situations
- 483 • Independent variables (dispositions, learning opportunities, conditions)

484 We sought a uniform approach to all  $N = 66$  studies as proposed by Arksey and O'Malley  
485 (2005), who themselves admit that this is sometimes not possible for each study due to  
486 missing information. For instance, some authors did not explicitly state in which country  
487 the study was conducted. Nevertheless, we obtained a nearly complete data sheet after  
488 going through an iterative process. Before the analysis, the research team had collectively  
489 determined which variables to extract from the included studies, as Colquhoun et al.  
490 (2014) suggested. Based on that, the two joint first authors selected ten random papers,  
491 applied the category system, discussed occurring disagreements, and applied the version  
492 to another five articles and discussed disagreements again, leading to two updates of the  
493 category system. In many studies, research designs were rather complex. Lesson planning

494 was assessed by different means at different times so that ticking the corresponding  
495 categories per publication resulted in a scarcely meaningful table. Some research designs  
496 comprise several data materials or a step-wise procedure, such as analysing written lesson  
497 plans and semi-structured interviews about the lesson planning process afterward. In such  
498 cases, we counted the articles for both codes, i.e., the analysis of the lesson plan  
499 (performance, planning product) and the analysis of the interview (self-report, planning  
500 process). As a result, the total numbers in the reported tables in the results section will  
501 exceed  $N = 66$ .

502 Moreover, we had to find that analysing the research designs, that is, which variables  
503 influence which other variables in the papers, was often challenging because researchers  
504 usually did not explicitly state the variables in their research questions or the text.  
505 Consequently, we decided not to split the remaining articles and calculate the interrater  
506 agreement. Instead, both joint first authors coded all  $N = 66$  papers independently and  
507 discussed all disagreements until a consensus was reached to ensure intersubjective  
508 comprehensibility.

509 While reviewing the  $N = 66$  articles, we noticed that these papers mainly do not refer to  
510 the theoretical or empirical literature on lesson planning in their theoretical background  
511 chapters. Most studies investigated constructs like PCK and used planning processes or  
512 planning products as a data source. Thus, we gained the impression that the researchers  
513 in the field do barely refer to a common ground of theoretical or empirical literature on  
514 lesson planning but focus on literature related to their constructs of interest (e.g., PCK)  
515 or methodologies (e.g., lesson study). To investigate whether this impression is correct,  
516 we modified RQ 1 and added the bibliographic coupling analysis sample: Two documents  
517 A and B are bibliographically coupled if they both cite a third document C (Kessler,  
518 1963). The more documents A and B cite in common, the higher the bibliographic

519 coupling strength between them. Bibliographic coupling is a widely used methodology in  
520 bibliometric analyses (Ma et al., 2022) that enables researchers to gain insights into  
521 document-document similarity (Ahlgren & Jarneving, 2008). This methodology matches  
522 our purposes as we are interested in how far the  $N = 66$  articles relate to a common ground  
523 of empirical or theoretical research on lesson planning. Therefore, we used the software  
524 VOSviewer, version 1.6.9 (van Eck & Waltman, 2010; van Eck & Waltman, 2014). The  
525 DOI links of the reviewed articles were imported into the software. VOSviewer  
526 automatically applies normalization, mapping, and cluster techniques, resulting in a  
527 network map visualizing the bibliographic coupling among the reviewed studies. The  
528 network map visualizes three pieces of information: (1) There are *nodes*, i.e., the circles  
529 with the authors' names. Each node represents one of the publications in this study. The  
530 node size visualizes the total link strengths. This means a relatively big node is  
531 bibliographically coupled to many other publications in our sample. A small node  
532 indicates that this publication commonly cites different publications with only a few (or  
533 even no) other publications in our sample. (2) The nodes are connected by *edges*, i.e., the  
534 lines of varying strength between two nodes. An edge of strength 1 indicates that the two  
535 related nodes cite one publication in common. The more publications are cited in  
536 common, the stronger the bibliographic coupling between these two publications and,  
537 thus, the thicker the line between them in the network map. (3) Based on the clustering  
538 technique that VOSviewer applies, the publications are assigned to clusters, which are  
539 visualized by the color of and the distance between the nodes. Publications located close  
540 to each other tend to cite the same publications, while publications located at far distance  
541 cite different publications.

542 Briefly, the network map will visualize to what extent the reviewed  $N = 66$  studies refer  
543 to a common ground of literature. As the network map only visualizes the bibliographic

544 coupling in terms of quantitative measures and does not provide information on which  
545 publications are cited in common, we provide a tabular overview of the most frequently  
546 cited publications among the  $N = 66$  articles. This table will show which publications  
547 form the majority of the edges in the network map.

548

### 549 **Stage 5: Collating, summarising, and reporting the results**

550 This section describes Stage 5 of Arksey and O'Malley's (2005) methodological  
551 framework for conducting scoping reviews. In line with Colquhoun et al. (2014), we will  
552 undertake two steps in this section: First, we will analyse the  $N = 66$  studies in a  
553 descriptive quantitative and a qualitative way and report the results. Second, we will  
554 discuss the findings and provide implications for further research on lesson planning in  
555 science education. We will proceed this way for all three research questions separately.

556 ***RQ 1: What is the state of science education research on lesson planning regarding***  
557 ***demographics, methodological approaches, and bibliographic coupling?***

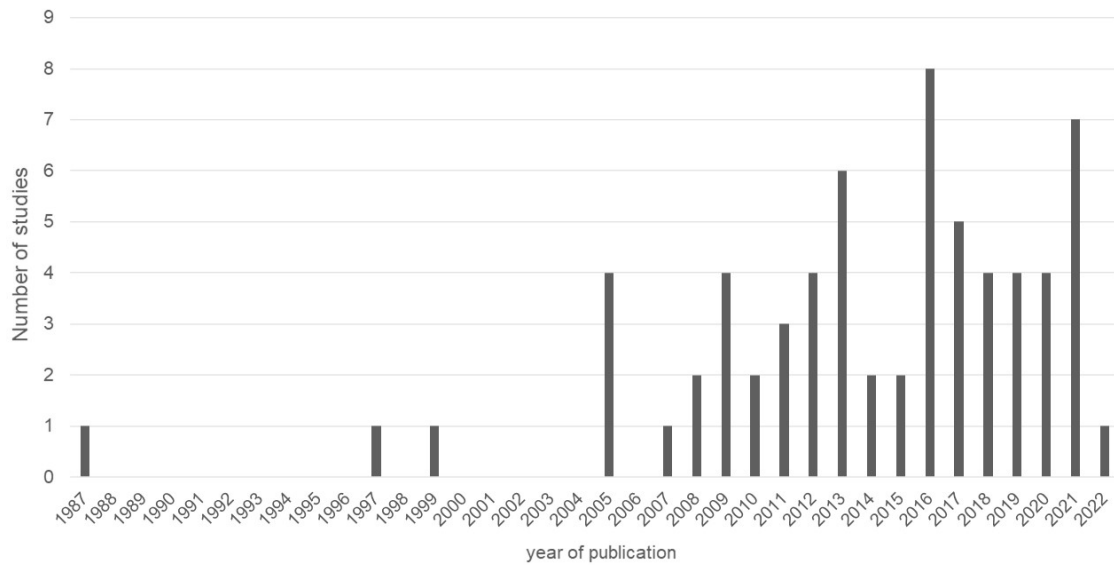
#### 558 *Demographics*

559 The review focused on peer-reviewed journal articles published in English between 1987  
560 and June 2022, since Shulman introduced the notion of PCK into research in science  
561 education. As Figure 3 shows, the most significant number of articles concerning lesson  
562 planning has been published since 2005. Only the studies of Hashweh (1987), W. M. So  
563 (1997), and Frederik et al. (1999) date from before 2005. Due to the timing of the  
564 literature search for this scoping review, publications were only considered up to June  
565 2022.

#### 566 **Figure 3**

567 *Breakdown of studies ( $N = 66$ ) referring to lesson planning by year of publication.*





568

569 Regarding the countries where the studies were conducted, 19 countries are represented  
 570 in this review (Table 2). Half of the considered studies are conducted in the United States,  
 571 constituting by far the largest group ( $n = 33$ ), followed by Turkey ( $n = 8$ ), China, and  
 572 Germany (both  $n = 4$ ). Only one study is included from ten other countries spread around  
 573 the world.

574 **Table 2**

575 *Publications by countries of origin.*

576 *Note.* The sum in the table equals 68 because two studies were conducted in two countries  
 577 (Frederik et al., 1999: Netherlands and Portugal; Weitzel & Blank, 2020: Germany and  
 578 Switzerland).

<i>n</i>	Countries
33	USA
8	Turkey
4	China, Germany
2	Israel, Netherlands, South Africa, Sweden, United Kingdom
1	Australia, Canada, Finland, Jordania, Lebanon, Norway, Palestine, Portugal, Spain, Switzerland

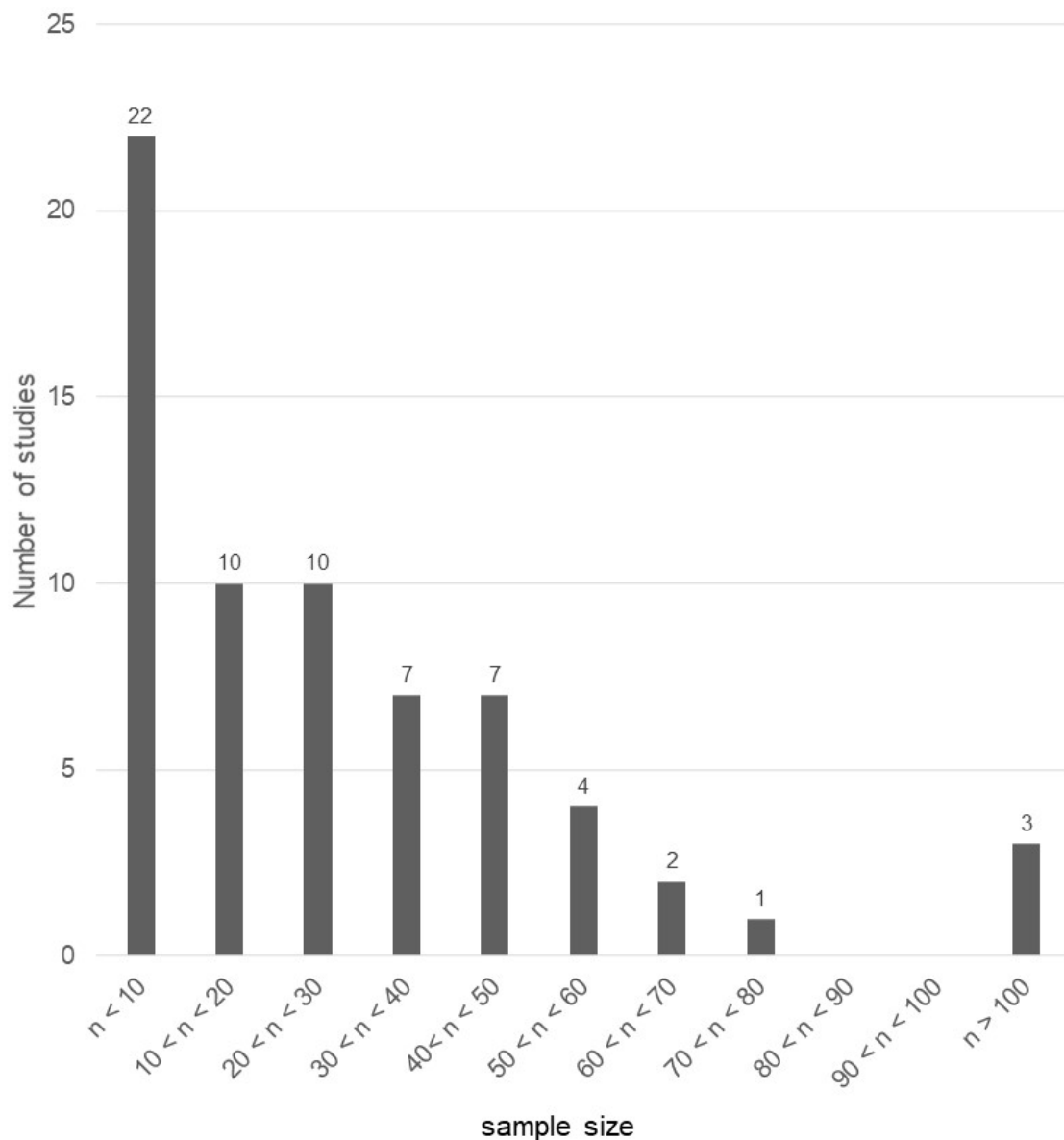
579

580 *Sample sizes*

581 Most studies employ research designs based on relatively small sample sizes (Figure 4).  
 582 One-third of the studies investigated sample sizes between 1 and 9 participants. For  
 583 instance,  $n = 4$  studies (Domingos-Grilo et al., 2012; Grove et al., 2009; Gunckel, 2011;  
 584 Schneider, 2013) investigated one single science teacher in an in-depth qualitative  
 585 approach. In contrast, only  $n = 3$  studies (Childs et al., 2011; Luera et al., 2005; Weitzel  
 586 & Blank, 2020) included sample sizes beyond 100 participants in their research designs.  
 587

588 **Figure 4**

589 *Breakdown of studies (N = 66) referring to lesson planning by sample size.*



590

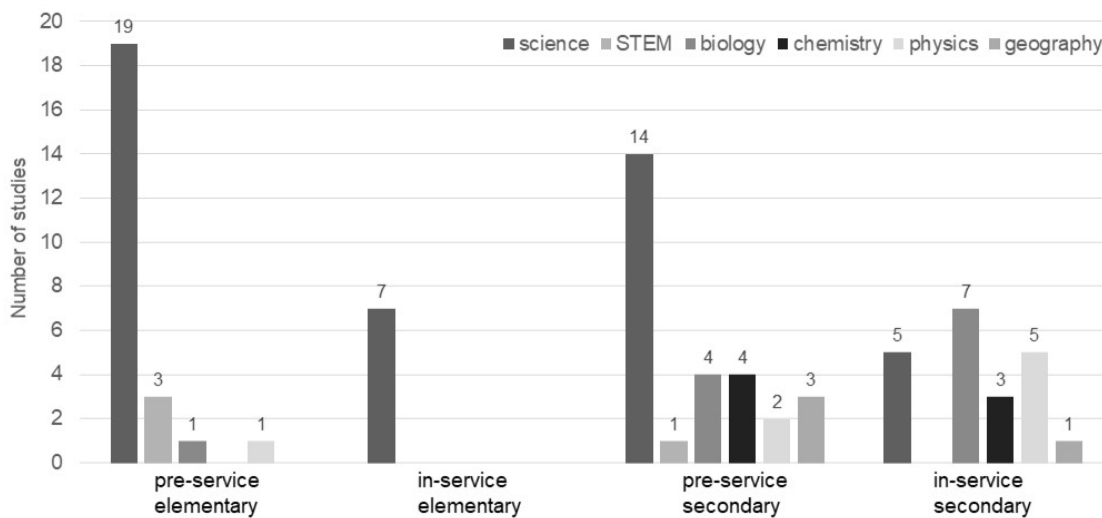
591 *Participants*

592 Taking a closer look at the participants included in the studies, we find that both pre-  
593 service science teachers (PSTs;  $n = 47$ ) and in-service science teachers (ISTs;  $n = 21$ ) are  
594 investigated, yet a focus lies on PSTs (Figure 5). As far as elementary teachers are  
595 concerned, only  $n = 7$  studies include ISTs. Most studies investigate science or STEM  
596 teachers ( $n = 43$ ).

597 **Figure 5**

598 *Breakdown of studies (N = 66) referring to lesson planning by grade level and*  
599 *participants' subject.*

600 *Note.* Studies that included more than one group of participants were counted several  
601 times.



602

603 *Study design*

604 A range of study designs is reported in the included articles (Figure 6). Knowledge and  
605 skills for lesson planning are measured in simulated contexts (e.g., university courses;  $n$   
606 = 34) and in situ (e.g., in authentic classroom contexts;  $n = 35$ ). In most studies, the  
607 participants' individual planning processes or products in simulated ( $n = 29$ ) or in situ  
608 settings ( $n = 28$ ) are analysed ( $n = 55$ ). However, many studies also employ collaborative

609 planning in simulated ( $n = 13$ ) and in situ settings ( $n = 13$ ). Overall, the majority of  
610 studies examine lesson planning in a discipline-specific way ( $n = 42$ , e.g., planning for  
611 socio-scientific issues, inquiry), but some also examine lesson planning in a topic-specific  
612 way ( $n = 24$ ).

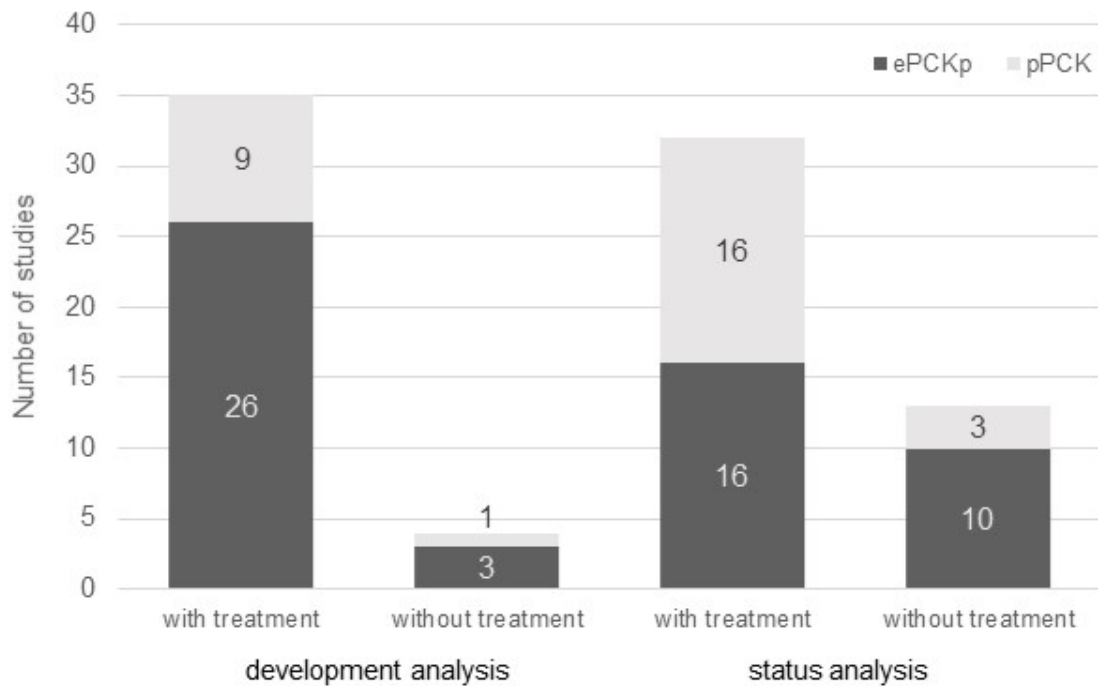
613  
614

615 Figure 6 displays the employed study designs among the reviewed studies. Most studies  
616 focus on performance ( $n = 55$ ), and fewer refer to self-reports ( $n = 29$ ). Among the  
617 performance studies, most studies analyse the development of knowledge and skills for  
618 lesson planning by application of a treatment ( $n = 50$ ). Most self-reports studies measure  
619 knowledge and skills for lesson planning once after a treatment ( $n = 16$ ). Almost all  
620 studies included in this review use a design without a control group ( $n = 60$ ). The  
621 remaining studies use an experimental-control design ( $n = 3$ ) or a quasi-experimental  
622 design ( $n = 3$ ).

623 **Figure 6**

624 *Breakdown of studies ( $N = 66$ ) referring to lesson planning by study design.*

625 *Note. Studies that include both performance and self-reports were counted several times.*



626

627 *Journal spectrum*

628 As we attempted to cover all peer-reviewed papers that contribute to answering our  
 629 research questions, we did not limit the search to the most influential science education  
 630 journals. Consequently, the  $N = 66$  articles included in this review were published in a  
 631 wide range of journals (Table 3). Two findings are particularly noteworthy: First,  
 632 approximately a fourth of the articles included in this review was published in the *Journal*  
 633 *of Science Teacher Education*. Second, among the 37 journals, only seven are ranked in  
 634 the top 100 in the Scimago Journal and Country Rank (SJR) within the field of education,  
 635 adding up to only  $n = 9$  articles published in these most impactful educational journals.

636 **Table 3**

637 *Journals that published the articles included in this review ( $N = 66$ ).*

638 *Note.* \* indicates that the referenced journal is ranked in the top 100 in the Scimago  
 639 Journal and Country Rank (SJR) within the field of education as of 2021.

---

Journals	$n$
----------	-----

---

---

Journal of Science Teacher Education	19
International Journal of Science Education	6
Journal of Research in Science Teaching*; Journal of Chemical Education; Research in Science Education*; School Science and Mathematics; Science education*; Teaching and Teacher Education*	2
Australian Journal of Educational Technology; Chemistry Education Research; Education Sciences; Educational Policy Analysis and Strategic Research; Educational Studies; Environmental Education Research; EURASIA Journal of Mathematics, Science and Technology Education; Eurasian Journal of Educational Research; European Journal of Teacher Education*; International Journal of Technology in Education; International Journal for the Scholarship of Teaching and Learning; International Journal of Curriculum and Instruction; International Journal of Environmental and Science Education; International Online Journal of Education and Teaching; Journal of Biological Education; Journal of Computer Assisted Learning*; Journal of Elementary Science Education; Journal of Geography; Journal of Learning Design; Journal of Science Education and Technology; Journal of Teacher Education*; Physics Education; Professional Development in Education; Science & education; South African Journal of Education; Teacher Development; Teacher Education and Practice; Technology, Knowledge and Learning; Technology, Pedagogy and Education; The Clearing House; The Teacher Educator	1

---

640

641 *Bibliographic coupling*

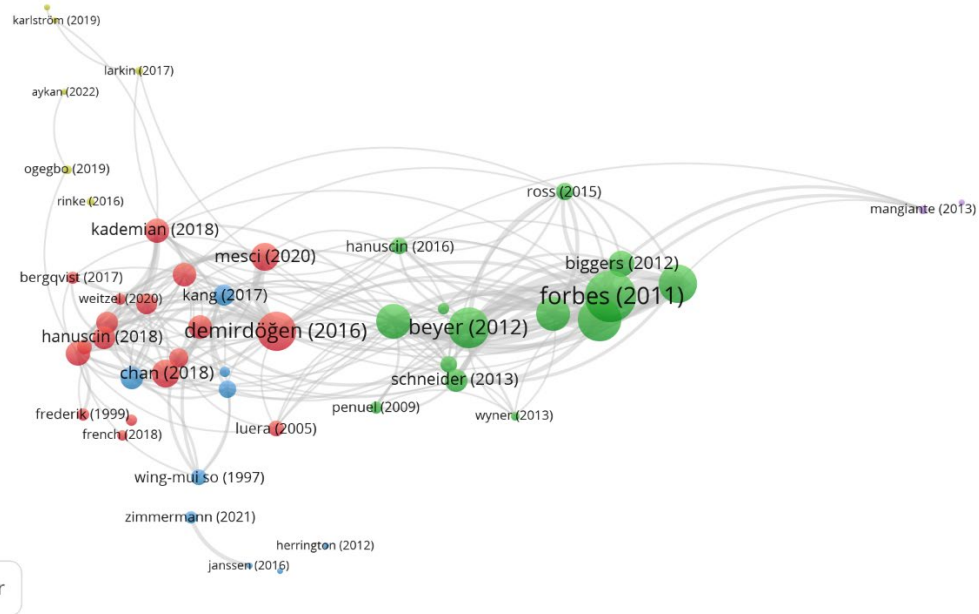
642 As a result of the bibliographic coupling analysis, the software VOSviewer produced a  
 643 network map (Figure 7).

644 **Figure 7**

645 *Bibliographic coupling network map.*

646 *Note.* Only edges  $> 2$  are shown. Since VOSviewer requires to use DOIs and four  
 647 publications in this sample have no DOI link, we only imported  $n = 62$  articles.

648



649

650 Two significant observations can be highlighted here: First, there are two major clusters  
 651 (red and green), among which we find the strongest bibliographic coupling. Demirdögen  
 652 and Uzuntiryaki-Kondakçı (2016) in the red cluster is bibliographically coupled to 24,  
 653 Forbes (2011) in the green cluster to 21 publications in this sample. Thus, they are the  
 654 two publications most strongly coupled to other publications in this sample. However,  
 655 they are especially coupled to publications within their clusters: For instance, the green  
 656 cluster comprises many publications focusing on the creation and adaption of curriculum  
 657 materials (e.g., Beyer & Davis, 2012; Biggers & Forbes, 2012; Ross & Cartier, 2015).  
 658 They refer to similar theoretical and methodological publications, resulting in a relatively  
 659 strong bibliographic coupling. Even though there are connections between these two  
 660 major clusters, they barely relate to the same publications. Second, many publications in  
 661 this sample are hardly bibliographically coupled (e.g., Janssen & Lazonder, 2016;  
 662 Herrington et al., 2012; Aykan & Yıldırım, 2022; Mangiante, 2013) and Childs et al.  
 663 (2011) even form their own cluster in remarkable distance to the remaining publications.

664 To gain further insight into which publications contribute most significantly to the  
 665 bibliographic coupling and to investigate whether they are directly related to lesson  
 666 planning research, Table 4 displays the most frequently cited publications in this review.

667 **Table 4**

668 *Publications most frequently cited by the N = 66 articles included in this review.*

669 *Note.* The list only contains empirical, theoretical, or methodological publications.

670 Standard documents (e.g., American Association for the Advancement of Science  
 671 [AAAS], 1993) are not included here, although they were cited frequently.

Number of Reference citations	
23	Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. <i>Educational Researcher</i> , 15(2), 4–14.
15	Magnusson, S., J. Krajcik, and H. Borko (1999). “Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching.” In <i>Examining Pedagogical Content Knowledge</i> , edited by J. Gess-Newsome and N. G. Lederman, 95–132. Dordrecht: Kluwer.
15	Miles M. B. and Huberman A. M., (1994). <i>Qualitative data analysis: an expanded sourcebook</i> , 2nd ed, Thousand Oaks: Sage Publications. [or different editions]
13	Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. <i>Educational Researcher</i> , 34(3), 3–14
12	Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. <i>Harvard Educational Review</i> , 57(1), 1-22.
9	Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. <i>Review of Educational Research</i> , 76(4), 607–651.
9	Remillard, J. T. (2005). Examining key concepts in research on teachers’ use of mathematics curricula. <i>Review of Educational Research</i> , 75(2), 211-246.
8	Davis, E.A. (2006). Preservice elementary teachers’ critique of instructional materials for science. <i>Science Education</i> , 90(2), 348–375.
8	Merriam, S. B. (2009). <i>Qualitative research: a guide to design and implementation</i> . Revised and expanded from qualitative research and case study applications in education (3rd edition). San Francisco: Jossey-Bass Publishers. [or different editions]
8	Patton, M.Q. <i>Qualitative Research &amp; Evaluation Methods</i> , 3rd ed.; Sage Publications, Inc.: Thousand Oaks, CA, USA, 2002. [or previous editions]
7	Bryan, L. A., & Abell, S. K. (1999). Development of professional knowledge in learning to teach elementary science. <i>Journal of Research in Science Teaching</i> , 36(2), 121–139.
7	Clark, C. M., & Peterson, P. L. (1986). Teacher’s thought processes. In M. C. Wittrock (Ed.), <i>Handbook of research on teaching</i> (3rd ed., pp 255–296). New York: Macmillan.
7	Davis, E. A., & Smithey, J. (2009). Beginning teachers moving toward effective elementary science teaching. <i>Science Education</i> , 93(4), 745–770.
7	Gess-Newsome J., (2015). A model of teacher professional knowledge and skill including PCK: results of the thinking from the PCK Summit, in Berry A., Friedrichsen P. and Loughran J. (ed.) <i>Re-examining pedagogical content knowledge in science education</i> , New York: Routledge, pp. 28–42.
7	Grossman, P. L. (1990). <i>The making of a teacher: Teacher knowledge and teacher education</i> . New York: Teachers College Press.



672

673 As far as lesson planning is concerned, we only find Remillard's (2005) seminal literature  
674 review and the literature on critiquing and adapting curriculum materials (i.e., Davis,  
675 2006; Davis & Krajcik, 2005; Davis et al., 2006; Davis & Smithey, 2009) among the most  
676 frequently cited publications in our sample. However, it should be noted that these  
677 publications are partly authored by the same researchers as some of the publications  
678 included in this review (e.g., Beyer & Davis, 2012). Accordingly, this results in the green  
679 cluster in Figure 8. The remaining articles cover Shulman's seminal PCK papers  
680 (Shulman, 1986; 1987) and publications in the field of teacher professionalization that  
681 build upon Shulman (i.e., Magnusson et al., 1999; Gess-Newsome, 2015; Grossmann,  
682 1990), and methodological literature (i.e., Miles & Huberman, 1994; Merriam, 2009;  
683 Patton, 2002). In summary, the most frequently cited publications among the reviewed  
684 studies are not related to lesson planning but to different theoretical or methodological  
685 topics.

686

### 687 *Discussion*

688 Mapping the research landscape brought some critical findings to light that might advance  
689 the field of empirical research on science teachers' lesson planning. Generally speaking,  
690 the lesson planning literature can be classified as an emergent field of research that still  
691 lacks a coherent theoretical foundation and that builds upon each other's empirical  
692 evidence (Table 4): Apparently, most researchers in the field refer to literature not specific  
693 for lesson planning, but that instead addresses generic aspects of teachers' profession (i.e.,  
694 Magnusson et al., 1999; Shulman, 1986).

695 Most strikingly, knowledge and skills for lesson planning are primarily analysed in  
696 qualitative research designs based on small sample sizes. As the field is still emerging  
697 and there is no consensual theoretically sound lesson planning model, researchers still  
698 gather qualitative insights into science teachers' lesson planning. Without such a model,  
699 researchers cannot deduce hypotheses and empirically test them in experimental control-  
700 group study designs. Consequently, our findings indicate that the field is still in its  
701 explorative beginnings, which makes qualitative approaches most suitable at the moment.  
702 Qualitative research attempts are promising to gain insights into science teachers' thought  
703 processes, differences between experts and novices (e.g., Westerman, 1991), or cognitive  
704 demands they need to meet during lesson planning processes (e.g., König, Krepf et al.,  
705 2021).

706 Undoubtedly, qualitative research is valuable in shedding light on individual cases and  
707 providing in-depth insights into teachers' thinking. In the best case, a purposeful sampling  
708 strategy (e.g., Patton, 1990) is chosen, and a rigorous data analysis procedure is employed  
709 to yield at least somehow generalizable results. However, research designs including  
710 larger sample sizes would be required to ensure that research findings are generalizable  
711 in a broader sense. Quantitative measures must be employed when sufficient qualitative  
712 findings are gathered and systematically analysed. However, barely any instruments are  
713 available so far intended to measure knowledge and skills for lesson planning  
714 quantitatively explicitly. Moreover, the vast majority of studies included in this review  
715 do not employ an experimental-control group design. On the one hand, this does not make  
716 these studies obsolete, as they might still contribute relevant findings to the community.  
717 However, as science teachers' lesson planning is still an underexplored construct and the  
718 variables influencing lesson planning are still barely investigated, it is necessary to  
719 carefully target certain variables (e.g., PCK) and control all other variables to establish

720 transparent cause-effect relationships. In line with the scoping review methodology, we  
721 did not evaluate the quality of the reviewed studies.

722 Two further aspects concerning the research designs are given insufficient attention so  
723 far: First, although a total of 19 nations contributed to the articles reviewed in this paper,  
724 the USA, Turkey, China, and Germany account for nearly three-quarters (74.2%) of the  
725 research output (Table 2). Since the development of knowledge and skills for lesson  
726 planning is an essential goal of teacher education in many countries, and teacher educators  
727 are expected to help PSTs develop their lesson planning by application of different ways  
728 of learning, it might be enriching for the field to cover a more diverse range of countries  
729 contributing to this research. Second, research activities focus mainly on PSTs (Figure  
730 5). However, developing knowledge and skills for lesson planning is a life-long process  
731 (Munthe & Conway, 2017). As PSTs at universities still lack practical experience in  
732 everyday practice at schools, it might be worth giving much more attention to experienced  
733 and effective ISTs' planning processes and products (Rothland, 2021). Experienced and  
734 effective ISTs can be characterized as teachers who integrate their profound knowledge  
735 bases and consider their students' cognitive processes to plan lessons (Westerman, 1991).  
736 Such teachers would score highly in teaching quality measures and, as a consequence,  
737 contribute significantly to students' learning gains. Such research might yield insights  
738 into effective teachers' thinking that might inform teacher educators and researchers how  
739 distinctive lesson planning can be characterized, which, in turn, might help to tailor  
740 teacher education to PSTs' needs.

741 Beyond these generic methodological shortfalls, our findings can be interpreted as  
742 indicators of particular challenges that are specific to lesson planning and that require  
743 great methodological efforts:

744 First, nearly half of the studies have participants plan in situ (i.e., for authentic classroom  
745 teaching), and the other half investigates lesson planning in simulated contexts (e.g., in a  
746 university course). Experienced ISTs plan lessons intuitively, usually without writing  
747 extensive lesson plans or taking notes (Ball et al., 2007). As Kang (2017) pointed out,  
748 “research on teachers’ planning was often conducted in a controlled setting to uncover  
749 cognitive decisions made by teachers. Researchers tend to rely on either teacher self-  
750 reported data or data generated from think-aloud methods [...]. Little empirical evidence  
751 focuses on teachers’ effective planning in the natural teaching environments” (Kang,  
752 2017, p.56). Second, researchers in the field investigate individual teachers’ lesson  
753 planning and planning in pairs or even groups of peers. Both approaches have pros and  
754 cons: Analysing individual teachers’ lesson planning allows researchers to capture  
755 ePCKp and gain insights into an individual’s current lesson planning knowledge and  
756 skills or the development over time. However, as PCK is tacit (Alonzo et al., 2019; Chan  
757 & Hume, 2019), and as lesson planning processes are usually not verbally expressed (Ball  
758 et al., 2007; Clark & Yinger, 1980), it is difficult to capture the full range of thoughts  
759 during the planning process. In contrast, analysing pairs or groups of teachers in  
760 collaborative planning talks provides more insights into the dynamics of teachers’  
761 thinking during lesson planning as they have to communicate their ideas and, thus, make  
762 them accessible for research (Roskos, 1996). Beyond that, as we also aimed to draw a  
763 more detailed picture of the theoretical and empirical background the reviewed studies  
764 refer to, we analysed the bibliographic coupling in the  $N = 66$  articles. As expected, the  
765 field is rather diverse in the foci they place in their research (Figure 8). For instance, many  
766 papers are located in PCK research, whereas others employ lesson studies. Remarkably,  
767 these studies barely relate to each other in the sense of empirical or theoretical literature  
768 on lesson planning. Thus, the literature can be described as somewhat fragmented. To

769 strive forward, researchers in the field might consider leveraging synergies by agreeing  
770 on a common theoretical foundation. In this paper, we suggested referring to the RCM  
771 (Carlson et al., 2019) and thus locating lesson planning research within PCK research.  
772 More specifically, researchers might take notice of recent research activities, particularly  
773 in Germany. For instance, König, Krepf et al. (2021) developed a heuristic model of  
774 cognitive demands that teachers must master in lesson planning. Such models or theories  
775 that explicitly refer to lesson planning might be worth considering in empirical research  
776 to make research findings more comparable and build on one another. In that sense, the  
777 present scoping review might also help establish common ground in lesson planning  
778 research. We will summarize the findings of this review at the end of this paper and  
779 suggest a tentative model of science teachers' lesson planning and research trajectories  
780 for further empirical investigations.

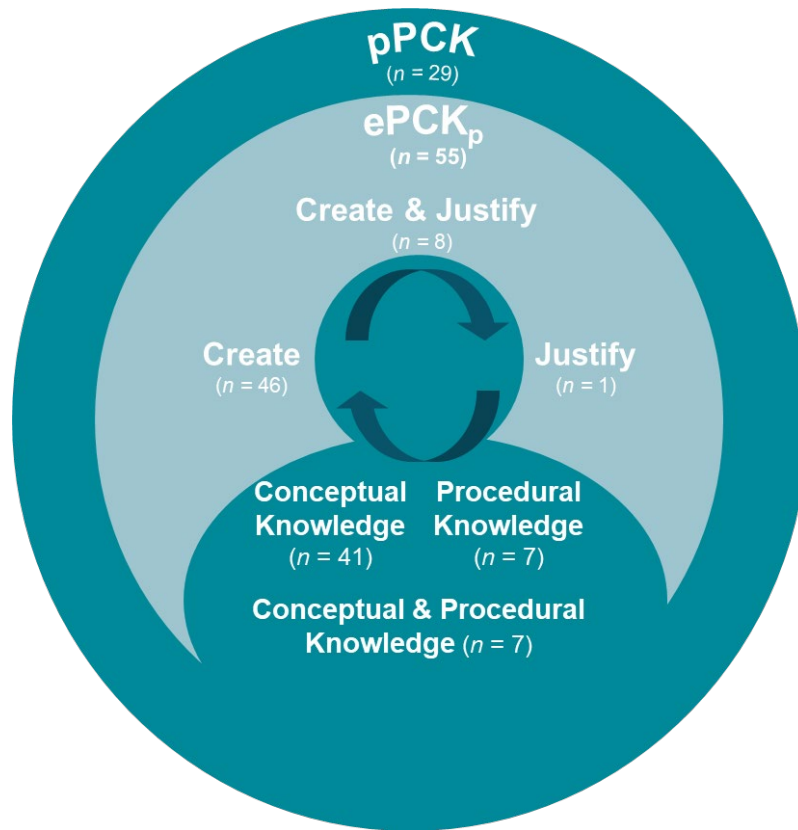
781

782 ***RQ 2: Which aspects of knowledge and skills for lesson planning regarding the RCM***  
783 ***are investigated in science education research?***

784 To connect to the RCM as a conceptual framework for this review, we adopted the  
785 original model and specified it for research designs employed in the  $N = 66$  reviewed  
786 studies. Figure 8 provides an overview of how many studies focused on the different  
787 dispositions, situation-specific skills, and realms of PCK.

788 **Figure 8**

789 *Overview of studies on lesson planning based on the RCM (Carlson et al., 2019).*



790

791 *pPCK and ePCK<sub>p</sub>*

792 Overall,  $n = 11$  focused on  $ePCK_p$  investigated by self-reports,  $n = 37$  focused on  $ePCK_p$   
 793 investigated in performance situations, and  $n = 18$  considered self-report and  
 794 performance (Figure 9). Regarding self-report studies, the participants were mostly  
 795 requested to report their attitudes and thoughts on lesson planning. For instance, Aykan  
 796 and Yıldırım (2022) employed a questionnaire and asked, among others, what challenges  
 797 the participants perceive in lesson planning. Similarly, Bartels et al. (2019) gathered  
 798 information on how PSTs proceed in their lesson-planning process by asking them  
 799 directly how they teach STEM in a post-instruction survey. Penuel (2009) had  
 800 participants report their instructional planning in an online questionnaire. A particular  
 801 way to capture teachers' self-reports on their lesson planning is provided by So and  
 802 Watkins (2005), who had their participants draw concept maps of the components and  
 803 the process of planning lessons. Schneider (2013) relied on different data sources, such

804 as a weekly written teacher journal and periodic informal interviews, to gather  
 805 information on her participant's thoughts on her lesson planning.  
 806 Regarding performance, most studies ( $n = 41$ ) investigate conceptual knowledge related  
 807 to lesson planning (i.e., PCK and CK), and only a few ( $n = 7$ ) take procedural knowledge  
 808 (i.e., planning strategies) or both ( $n = 7$ ) into account (Table 5).

809 **Table 5**

810 *Performance studies referring to conceptual knowledge, procedural knowledge, or both.*

<i>n</i>	Studies	Conceptual knowledge		Procedural knowledge
		PCK	CK	Strategies
32	Bahçivan (2016); Bartels et al. (2019); Beyer and Davis (2012); Biggers and Forbes (2012); Bozkurt Altan and Ucuncuoglu (2019); Deng et al. (2017); Domingos-Grilo et al. (2012); Firth and Winter (2007); Forbes and Davis (2010); Forbes (2011); Golightly (2010); Grove et al. (2009); Grueber and Özgün-Koca (2013); Güngören et al. (2020); Hanuscin and Zangori (2016); Hanuscin et al. (2018); Hashweh (1987); Herrington et al. (2012); Hubbard and Abell (2005); Juhler (2016); Jung and Brown (2016); Kademian and Davis (2018); Kang (2017); Kapon and Merzel (2019); Käpylä et al. (2009); Luera et al. (2005); Mangiante (2013); Qablan (2016); Wyner (2013a); Wyner (2013b); Yildiz (2021); Zimmermann et al. (2021)	x		
9	Aykan and Yıldırım (2022); Bergqvist and Chang Rundgren (2017); Dawkins et al. (2008); Frederik et al. (1999); Mesci et al. (2020); Minken et al. (2021); Rinke et al. (2016); Wahbeh and Abd-El-Khalick (2014); Weitzel and Blank (2020)	x	x	
7	Beyer and Davis (2009); Chan and Yung (2018); Forbes (2013); Gunckel (2011); Karlström and Hamza (2021) Larkin (2017); Ross and Cartier (2015)			x
4	Demirdöğen & Uzuntiryaki-Kondaççı (2016); French and Burrows (2018); Janssen and Lazonder (2016); Stender et al. (2017)	x		x
2	Dubek & Dyle-Jones (2021); Duruk (2020)	x	x	x
1	Eshchar-Netz and Vedder-Weiss (2021)		x	x

811  
 812 Two significant observations can be highlighted: First, the vast majority of studies focus  
 813 solely on the identification of single PCK components. While some studies refer to PCK  
 814 rather loosely and identify single PCK components in lesson plans (e.g., Bahçivan, 2016),  
 815 many studies do not explicitly refer to PCK as a conceptual framework, even though they  
 816 capture information that can be assigned to specific PCK components (e.g., Grueber &

817 Özgün-Koca, 2013; Kapon & Merzel, 2019; Qablan, 2016). Only a few studies explicitly  
 818 locate themselves within PCK research and capture PCK components qualitatively using  
 819 content representations (e.g., Hanuscin et al., 2018) or specific coding schemes (e.g.,  
 820 Demirdöğen & Uzuntiryaki-Kondakçı, 2016).

821 Second, relatively few performance studies investigate science teachers' procedural  
 822 knowledge (Table 6), namely, their planning strategies and ability to integrate the single  
 823 PCK components to plan a lesson ( $n = 14$ ). For instance, Stender et al. (2017) used a  
 824 performance test. They asked specific questions concerning their participants' ideas on  
 825 how to proceed in the planning process based on the vignettes without enacting the lesson  
 826 itself. Dubek and Doyle-Jones (2021) analysed transcribed co-planning sessions and  
 827 inductively described emerging patterns concerning the strategies employed in the  
 828 planning process.

829 **Table 6**

830 *Planning strategies reported in studies reporting procedural knowledge ( $n = 14$ ).*

Phase	Strategies	References
How to proceed in the lesson planning process?	searching information (e.g., about activities, content)	Dubek and Doyle-Jones (2021)
	critiquing and adapting curriculum materials	Beyer and Davis (2009); Forbes (2013); Ross and Cartier (2015)
	using planning guides or models	Gunckel (2011); Karlström and Hamza (2021)
	talking to peers/mentors (i.e., co-planning)	Dubek and Doyle-Jones (2021); Eshchar-Netz and Vedder-Weiss (2021)
How to make planning decisions?	planning spontaneously/roughly or thoughtfully/deliberately	Chan and Yung (2018)
	aligning instructional strategies to students' understanding in science	Demirdöğen and Uzuntiryaki-Kondakçı (2016)
	aligning instructional strategies and assessments to the intended learning outcome	Duruk (2020)
	making planning decisions in a particular order (e.g., starting with the intended learning outcome, with the topic, with students' thinking)	Eshchar-Netz and Vedder-Weiss (2021); Karlström and Hamza (2021)



eliciting students' ideas as a starting point for instructional planning	Beyer and Davis (2009); Larkin (2017)
discussing possible alternatives and weighing the pros and cons	Eshchar-Netz and Vedder-Weiss (2021)
making use of teaching scripts	Stender et al. (2017)
justifying planning decisions	Janssen and Lazonder (2016)

831

832 Among the performance studies referring to conceptual knowledge, the focus lies on  
833 capturing PCK facets instead of CK (Table 5). We find that *Knowledge of Instructional*  
834 *Strategies* (KISR;  $n = 42$ ) and *Knowledge of Students' Understanding in Science* (KSU;  
835  $n = 29$ ) are the most frequently investigated PCK facets analysed in the context of lesson  
836 planning (Table 10). *Orientation to Teaching Science* (OTS;  $n = 16$ ), *Knowledge of*  
837 *Science Curriculum* (KSC;  $n = 17$ ), and *Knowledge of Assessment of Science Learning*  
838 (KAs;  $n = 14$ ) are less frequently addressed. Most study designs do not focus on one  
839 single PCK component but capture at least two PCK components. (Table 7). Only  $n = 9$   
840 studies operationalize ePCK<sub>p</sub> using one single PCK facet.

841 **Table 7**

842 *Overview of ePCK<sub>p</sub> studies investigating PCK components (n = 55)*

<i>n</i>	ePCK <sub>p</sub> studies	OTS	KC	KSU	KISR	KAs
9	Beyer and Davis (2012); Bozkurt Altan and Ucuncuoglu (2019); Forbes (2011); French and Burrows (2018); Hanuscin et al. (2018); Juhler (2016); Jung and Brown (2016); Weitzel and Blank (2020); Yildiz (2021)		x	x	x	x
7	Bergqvist and Chang Rundgren (2017); Dawkins et al. (2008); Janssen and Lazonder (2016); Kademian and Davis (2018); Kang (2017); Minken et al. (2021); Stender et al. (2017)			x	x	
7	Hashweh (1987); Herrington et al. (2012); Kapon and Merzel (2019); Luera et al. (2005); Mangiante (2013); Wyner (2013b); Zimmermann et al. (2021)				x	
4	Bahçivan (2016); Grueber and Özgün-Koca (2013); Hanuscin and Zangori (2016); Käpylä et al. (2009)	x	x	x	x	
4	Forbes and Davis (2010); Mesci et al. (2020); Qablan (2016); Wyner (2013a)	x			x	
3	Biggers and Forbes (2012); Domingos-Grilo et al. (2012); Hubbard and Abell (2005)	x		x	x	
2	Deng et al. (2017); Rinke et al. (2016)		x		x	
2	Bartels et al. (2019); Güngören et al. (2020)	x		x	x	x
1	Golightly (2010)	x			x	x

1	Firth and Winter (2007)		x	x	x	
1	Aykan and Yıldırım (2022)			x	x	x
1	Frederik et al. (1999)			x		
1	Grove et al. (2009)		x			
1	Wahbeh and Abd-El-Khalick (2014)		x	x	x	x

843

844 A more specific look at the KSU and KISR reveals the most frequently investigated  
845 aspects of PSTs’ and ISTs’ lesson planning.

846 Concerning KSU, three significant aspects were addressed in the reviewed  
847 studies. First, many researchers refer to students’ misconceptions (e.g., Demirdöğen &  
848 Uzuntiryaki-Kondakçı, 2016; Gunckel, 2011; Hubbard & Abell, 2005), alternatively  
849 named as “students’ preconceptions” (e.g., Penuel et al., 2009; Weitzel & Blank, 2020)  
850 or “alternative ideas” (e.g., Domingos-Grilo et al., 2012; Kademian & Davis, 2018).  
851 Beyond the terminology, all these studies refer to the ideas about specific scientific topics  
852 that students bring to the classroom and that teachers might (or might not) consider in  
853 their instructional planning. Second, connecting to students’ misconceptions, some  
854 researchers explicitly refer to students’ conceptual difficulties in learning and  
855 understanding a particular topic (e.g., Bahçivan, 2016; Käpylä et al., 2009). Third, in a  
856 broader sense, some researchers refer to students’ prior knowledge (e.g., Bartels et al.,  
857 2019; Wahbeh & Abd-El-Khalick, 2014; Weitzel & Blank, 2020). In contrast to students’  
858 misconceptions and conceptual difficulties, investigations of students’ prior knowledge  
859 refer more generally to knowledge acquired in the lesson before or in everyday life. In  
860 addition to prior knowledge, few studies take a broader view and explicitly refer to  
861 students’ skills for learning a particular scientific concept and variations in students’  
862 approaches to acquiring new science knowledge (e.g., Hanuscin et al., 2018).

863 Concerning KISR, two significant conceptualizations of instructional strategies  
864 were used in the reviewed studies (Table 8): First, some studies emphasize the topic- and  
865 subject-specificity of instructional strategies (e.g., Bergqvist et al., 2016; Demirdöğen &

866 Uzuntiryaki-Kondakçı, 2016; Henke & Höttecke, 2015). They focus on instructional  
 867 strategies that are specific to teaching science. For instance, some studies focus on  
 868 scientific inquiry (e.g., Biggers & Forbes, 2012; Herrington et al., 2012; Mesci et al.,  
 869 2020; Qablan, 2016). Thus, they conceptualize KISR as strategies that foster students'  
 870 inquiry skills, such as generating hypotheses, explaining data, or evaluating alternative  
 871 explanations. For the particular case of teaching the history and philosophy of science,  
 872 Henke and Höttecke (2015) suggest creative writing, role plays, and using replicas of  
 873 historical instruments.

874 Second, many studies also refer to generic instructional strategies that would also  
 875 need to be considered beyond science teaching, such as ensuring that the teaching phases  
 876 build on each other (Weitzel & Blank, 2020), activities like watching a video (Käpylä et  
 877 al., 2009), or creating cognitively demanding tasks (Kang, 2017).

878 **Table 8**

879 *Overview of instructional strategies investigated in the reviewed studies (N = 66).*

Conceptualization of KISR	instructional strategy	Exemplary references
science-specific	inquiry-orientation	Biggers and Forbes (2012); Qablan (2016)
	Representations	Bergqvist and Chang Rundgren (2017); Henke and Höttecke (2015)
	Models	Demirdöğen and Uzuntiryaki-Kondakçı (2016)
	analogies	Demirdöğen and Uzuntiryaki-Kondakçı (2016)
	creative writing	Henke and Höttecke (2015)
generic	role play	Henke and Höttecke (2015)
	language	Jung and Brown (2016)
	phases/timing	Weitzel and Blank (2020)
	worksheets	Weitzel and Blank (2020)
	methods/activities	Käpylä et al. (2009)
	Tasks	Kang (2017)
questions	Sarieddine and BouJaoude (2014)	
Media	Dubek and Doyle-Jones (2021)	
use of technology	Herrington et al. (2012)	

880

881 *Creation and Justification*

882 Among the studies aiming at ePCK<sub>p</sub> investigated in performance situations ( $n = 55$ ), most  
883 studies focused on the creation of mental or written lesson plans ( $n = 46$ ), only  $n = 1$  study  
884 focused on the justification of planning decisions, and  $n = 8$  studies comprise the creation  
885 and the justification (Figure 8). Overall, science teachers' justifications were barely  
886 investigated. Among these  $n = 9$  studies, researchers employed different techniques to  
887 capture science teachers' justifications. For instance, some researchers conducted semi-  
888 structured interviews to gain further insights into their participants' reasoning (e.g.,  
889 Bahçivan, 2016; Chan & Yung, 2018), whereas others used additional data sources, such  
890 as a documentary journal in which the participants were asked to record thoughts and  
891 observations during the lesson planning and enacting (e.g., Dubek & Doyle-Jones, 2021).  
892 In contrast, Janssen and Lazonder (2016) analysed their participants' justifications as  
893 noted in their written lesson plans.

894

#### 895 *Planning processes and products*

896 Overall, most ePCK<sub>p</sub> studies investigated in performance situations ( $n = 55$ ) exclusively  
897 analysed planning products (e.g., written lesson plans). Among these studies, most ( $n =$   
898 30) focused on the analysis of lesson plans without including any other data sources in  
899 the analysis. For instance, Janssen and Lazonder (2016) investigated the effectiveness of  
900 two approaches to support PSTs in designing technology-infused lesson plans. In an  
901 experimental design, they compared two sub-samples' lesson plans concerning the  
902 learning activities the PSTs chose, their justifications, and the overall lesson plan quality  
903 in terms of PCK and TPACK. Other studies focused on the transition between the  
904 planning stage and the enactment (i.e., classroom teaching). For instance, Güngören et al.  
905 (2020) had PSTs plan lessons based on the 5E instructional model for teaching scientific  
906 inquiry. The lesson plans were analysed with an established rubric published by Goldston

907 et al. (2013), and the PSTs received feedback on their lesson plans. Afterward, they  
908 enacted the planned lessons in a microteaching setting in their university course. Such  
909 research designs focus on the planning product and its significance for classroom  
910 teaching.

911 In contrast, few studies ( $n = 14$ ) focussed on planning processes exclusively. Two major  
912 approaches exist in the reviewed literature. First, in some studies, researchers captured  
913 the collaborative planning processes of novice and expert teachers. For instance, Eshchar-  
914 Netz and Vedder-Weiss (2021) analysed a planning conversation between a novice and  
915 two experienced teachers. Similarly, Juhler (2016) analysed planning conversations  
916 between PSTs and their mentoring teachers at school. Second, in contrast, other studies  
917 focused on planning conversations among peers, that is, without more experienced ISTs  
918 who support them or question their ideas. For instance, Karlström and Hamza (2021) had  
919 PSTs plan a microteaching lesson in groups of five or six. They analysed in which order  
920 the PSTs elaborated on aspects relevant to the lesson planning process (e.g., students'  
921 prior knowledge, assessment).

922 Many studies analyse the planning process and the products ( $n = 22$ ). For instance, Kang  
923 (2017) had PSTs plan instructional tasks, a crucial element of a lesson plan. Afterward,  
924 the participants enacted the plan with students at their local school. For her analysis, Kang  
925 (2017) used the lesson plans, the PSTs' reports on how the plans were enacted at school,  
926 and teaching artifacts, such as worksheets and presentation slides. This part focuses on  
927 the planning product. In addition, the participants were interviewed. They were asked  
928 how they got their ideas for the planned tasks, what curriculum resources were used, and  
929 how they were adapted. Thus, Kang (2017) provides a holistic inside into the planning  
930 process and the products of the example of planning cognitively demanding instructional  
931 tasks. Similarly, Weitzel and Blank (2020) developed a research design that explicitly

932 interconnects planning processes and products: They investigated whether content-  
933 focused peer coaching supports PSTs in lesson planning. For that purpose, they employed  
934 an experimental design and compared PSTs who learned peer coaching in a workshop to  
935 those who did not. First, they analysed all PSTs' written lesson plans (planning products).  
936 Afterward, the participants discussed their lesson plans in pairs and were requested to  
937 improve them. These planning dialogs were videotaped and analysed concerning the PCK  
938 components the PSTs referred to (planning processes) and the number of modifications  
939 the pairs made to their initial lesson plans.

#### 940 **Discussion**

941 As lesson planning is a multi-faceted construct and, thus, research on it is a challenging  
942 endeavour. The reviewed studies investigate many different aspects of science teachers'  
943 lesson planning, e.g., the application of conceptual and procedural knowledge in creating  
944 and justifying planning decisions. In doing so, they target various aspects relevant to  
945 lesson-planning processes and products. Looking at it from the other side, this diversity  
946 can also be interpreted as a still lacking common understanding of lesson planning as a  
947 construct. The analysis revealed three findings of significant relevance:

948 First, researchers focus on conceptual knowledge necessary in lesson planning rather than  
949 procedural knowledge. In other words, which PCK components are considered in  
950 planning processes or products is often investigated, but only a few studies target the  
951 integration of these components. This is surprising as lesson planning comprises many  
952 different aspects that need to be interconnected. For instance, König, Krepf et al. (2021)  
953 six cognitive demands of lesson planning. These demands must be met in the lesson  
954 planning process to plan lessons appropriately. More specifically, the tasks, for example,  
955 should be aligned to the intended learning outcome and adapt to the students' current

956 needs. Thus, teachers are expected to relate these more or less distinct aspects to each  
957 other and build a lesson. Similarly, Karlström and Hamza (2021) illustrate the desired  
958 interconnectedness in their model of didactic relations. According to that, teachers need  
959 to answer five interrelated questions (i.e., *What* to teach? To *whom* to teach? *How* to teach  
960 it? *Why* to teach it (this way)? *How* to *assess* science learning?) in their planning  
961 processes. Assuming that the quality of a lesson plan lies in the degree of  
962 interconnectedness between relevant aspects is in line with PCK research: It is assumed  
963 and was empirically shown that the interconnectedness of PCK components is positively  
964 correlated with teaching quality (e.g., Reynolds & Park, 2021), even though contradicting  
965 findings were made for the quality of written lesson plans and the interconnectedness of  
966 ePCKp identified in them (Großmann & Krüger, 2022). Given these assumptions, it is  
967 surprising that researchers focus mainly on science teachers' conceptual knowledge  
968 instead of shedding light on how they integrate the different funds of knowledge (i.e.,  
969 how they apply their procedural knowledge). Our findings support recent remarks on the  
970 empirical underrepresentation of teachers' situation-specific skills in lesson planning  
971 (e.g., Campbell et al., 2022; König, Krepf et al., 2021). Hence, it might be fruitful to shed  
972 further light on the constructs captured in empirical research: As far as PCK is concerned,  
973 research in the field tends to focus on identifying and describing conceptual knowledge.  
974 More specifically, based on PCK models that describe distinct components of PCK (e.g.,  
975 Magnusson et al., 1999), researchers undertook significant efforts to analyse which PCK  
976 components are considered by science teachers (for an overview, see Chan & Hume,  
977 2019). Consequently, the idea of conceptual knowledge is emphasized in the RCM  
978 (Figure 1) in terms of the outermost circle rather than procedural knowledge. For lesson  
979 planning as one of teachers' core practices, the RCM in its current form might blur the  
980 integration of conceptual knowledge by the term "pedagogical reasoning" in the inner

981 circle. Specifically, PCK is described as knowledge and skill (Carlson et al., 2019; Gess-  
982 Newsome, 2015). However, it seems that researchers in the field focus on identifying the  
983 knowledge bases rather than on enacting this knowledge and the ability to plan lessons as  
984 one of teachers' most essential skills. As we identified several studies focussing on  
985 science teachers' planning strategies (i.e., their procedural knowledge; Table 6), it might  
986 be worth considering increasing research activities that try to analyse in detail how  
987 science teachers use their conceptual knowledge and integrate it to make planning  
988 decisions.

989         Second, most studies investigate how science teachers create a lesson plan, but  
990 only a few capture the justifications for their decisions. This is astonishing because  
991 justifying planning decisions would encourage science teachers to verbalize their  
992 thoughts and, thus, make them accessible for empirical analysis. Furthermore, justifying  
993 a planning decision or a prediction about the effects of planned activities on student  
994 learning requires the deliberate application of professional knowledge (Zaragoza et al.,  
995 2023). As many researchers in the field locate their studies on lesson planning within  
996 PCK research, it might be worth investigating the interplay of creation and justification  
997 in lesson planning processes by identifying which components of PCK (i.e., conceptual  
998 knowledge) are referred to in the creation process, and how they are interconnected in the  
999 justifications (i.e., procedural knowledge).

1000 Third, we identified an imbalance between studies investigating planning products and  
1001 those investigating planning processes. Collecting written lesson plans might be time-  
1002 consuming to gather information about the final decisions in the lesson planning process.  
1003 In contrast, capturing the lesson planning process is challenging because it is an iterative  
1004 and recursive process that can last for days (John, 2006; Yinger, 1980).

1005



1006

1007 ***RQ 3: Which variables influence science teachers' knowledge and skills for***  
1008 ***lesson planning?***

1009 In the following, we use the conceptual background to report on the variables that  
1010 influence science teachers' knowledge and skills for lesson planning. Accordingly, we  
1011 distinguish studies according to independent variables (1) dispositions, (2) amplifiers and  
1012 filters, and (3) learning opportunities that influence the dependent variable ePCK<sub>p</sub> (Table  
1013 9). The dependent variable in the sense of facets of ePCK<sub>p</sub> is shown in Tables 5, 6, and  
1014 7. There are many combinations of independent and dependent variables in the reviewed  
1015 studies. In the following, we will give a systematic overview of independent variables  
1016 and explain their influence on ePCK<sub>p</sub> in planning by selected examples.

1017

1018 *Dispositions*

1019 One-fifth of the studies ( $n = 12$ ) focus on the influence of dispositions on lesson planning  
1020 (Table 9). Most of them analyse the influence of teachers' knowledge bases (pPCK, CK,  
1021 PK) on ePCK<sub>p</sub>. Some studies ( $n = 5$ , e.g., Aykan & Yıldırım, 2022; Deng et al., 2017;  
1022 Stender et al., 2017) show that PCK influences lesson planning. For instance, Käpylä et  
1023 al. (2009) report that secondary PSTs ( $N = 10$ ) indicate *Knowledge of Students'*  
1024 *Understanding in Science* (KSU) as a prerequisite for lesson planning. Stender et al.  
1025 (2017) confirm the influence of pPCK on ePCK<sub>p</sub> measured with  $N = 49$  ISTs in a paper-  
1026 and-pencil test.

1027 Furthermore,  $n = 7$  studies show that CK is crucial for lesson planning (e.g., Chan &  
1028 Yung, 2018; Frederik et al., 1999; Luera et al., 2005). For instance, Luera et al. (2005)  
1029 report a positive correlation between CK (measured in a paper and pencil test with  $N =$   
1030 234 PSTs) and the ability to plan inquiry-based lessons. Similarly, Sarieddine and

1031 BouJaoude (2014) show in a study with  $N = 7$  ISTs that teachers' conceptions of NOS  
1032 influence the integration of NOS in lesson plans. Käpylä et al. (2009) report that nine out  
1033 of ten elementary PSTs indicate CK as crucial for planning and that a lack of CK causes  
1034 difficulties in lesson planning. There is no study investigating the influence of PK on  
1035 science teachers' ePCK<sub>p</sub>. A more non-specific disposition is reported by So (1997), who  
1036 compares the planning of ISTs with  $N = 5$  non-science teachers and finds that science  
1037 teachers show more interconnected reasoning than other teachers when planning science  
1038 lessons.

1039

#### 1040 *Amplifiers and filters*

1041 Most studies report on the influence of extra-personal factors ( $n = 23$ ), while only  $n = 10$   
1042 report on the influence of personal factors (Table 7). Studies on extra-personal factors  
1043 refer primarily to materials for planning support (e.g., 2009; Janssen & Lazonder, 2016)  
1044 and school environment (e.g., Forbes, 2013; Grove et al., 2009). These studies describe a  
1045 wide range of materials to support the planning process. Some studies investigate the  
1046 influence of discipline-specific planning tools. For instance, Gunckel (2011) shows in her  
1047 analysis of one female PST lesson planning that the *Inquiry-Application Instructional*  
1048 *Model* (I-AIM) combined with curriculum materials supports this PST in planning  
1049 conceptual change. However, she still struggles with planning inquiry instructions. Jung  
1050 and Brown (2016) investigate the lesson plans of seven elementary PSTs and report on  
1051 the usefulness of the *Academic Language Planning Organizer* (ALPO) to integrate  
1052 language support in lesson plans. Additionally, there are few studies investigating the  
1053 influence of generic planning tools, e.g., Karlström and Hamza (2021) show that generic  
1054 planning models “did not serve as an algorithm for planning” (p. 679) analysing

1055 procedural knowledge in planning processes of PST ( $N = 22$ ) planning lessons for  
1056 microteaching sequences in groups of five to six students.

1057 Few studies analyse the potential of educative support for dealing with curriculum  
1058 materials (e.g., Beyer & Davis, 2012; Forbes, 2011). When planning with curriculum  
1059 materials, planning means adapting the materials. Educative supports, such as  
1060 explanatory texts related to PCK or PK, might facilitate this process. Beyer and Davis  
1061 (2009) show in their study on  $N = 24$  elementary PSTs that educative support addressing  
1062 students' thinking helps PST to identify strengths and weaknesses in curriculum  
1063 materials. However, they barely continued doing so after no longer receiving such support  
1064 (Beyer & Davis, 2009). Similarly, Schneider (2013) reports that information about  
1065 students' ideas, tasks, and idea-specific support is helpful for teachers' lesson planning  
1066 based on curriculum materials. She based the results on an analysis of teacher journals,  
1067 informal interviews, and daily classroom videotapes of one elementary IST. Janssen and  
1068 Lazonder (2016) investigate in an experimental study ( $N = 54$  secondary PST) the  
1069 difference between educative support that integrates or separates PCK- and PK-related  
1070 information and show that integrated educative support fosters the use of PCK in the  
1071 reasoning of planning decisions.

1072 Beyond the planning tool and educative support, few studies show that curriculum and  
1073 textbooks influence teachers' lesson planning (e.g., Bergqvist & Chang Rundgren, 2017;  
1074 Hanuscin & Zangori, 2016; Kang, 2017; Sarriddine & BouJaoude, 2014; Verma, 2008).

1075 For example, it is reported that PSTs have problems integrating curriculum standards in  
1076 planning (e.g., National Science Education Standards [NSES],  $N = 48$  elementary PSTs;  
1077 Next Generation Science Standards [NGSS], Hanuscin & Zangori, 2016;  $N = 18$   
1078 elementary PSTs). Bergqvist and Chang Rundgren (2017) investigate lesson plans of  $N =$   
1079 10 secondary ISTs and report the influence of textbooks on teachers' planning and that

1080 teachers have problems analysing and adapting representations presented in textbooks.  
1081 Further extra-personal factors analysed are the school environment, such as colleagues  
1082 (e.g., Forbes, 2013; Henke & Höttecke, 2015; Kang, 2017), time (e.g., Käpylä et al., 2009)  
1083 or school equipment (e.g., Saredidine & BouJaoude, 2014).

1084 In planning, personal factors influencing teachers' ePCK<sub>p</sub> are orientations and beliefs ( $n$   
1085 = 6) and experience ( $n = 8$ ). For example, Käpylä et al. (2009) reported that orientations  
1086 influence science teachers' planning by comparing orientations of elementary PST ( $N =$   
1087 10) with secondary PSTs ( $N = 10$ ). Zimmermann et al. (2021) reported that TPACK self-  
1088 efficacy and attitudes influence the integration of digital tools in lesson plans; based on  
1089 an analysis of lesson plans from secondary PSTs ( $N = 41$ ). Additionally, Stender et al.  
1090 (2017) reported that orientations and beliefs moderate the influence of pPCK (i.e., topic-  
1091 specific pedagogical knowledge) on teaching scripts in planning. In contrast, Bahçivan  
1092 (2016) found a discrepancy between PSTs' epistemological beliefs and their lesson  
1093 planning in his analysis of lesson plans written by secondary PSTs ( $N = 3$ ): Their personal  
1094 beliefs regarding the nature of knowledge and scientific practices to gain new knowledge  
1095 do not have an influence on the instructional approaches they are planning to teach.

1096 Five studies show the influence of experience as a personal factor. For example, So and  
1097 Watkins (2005) show that the interconnectedness of planning decisions increases with  
1098 experience. The analysis is based on concept maps that describe planning drawn by  $N =$   
1099 25 elementary PSTs. Hanuscin et al. (2018) specify that it is not the general experience  
1100 that is decisive but the experience in the class level for which teachers plan lessons,  
1101 referring to their study of lesson plans from  $N = 18$  elementary PSTs.

1102

1103 *Learning opportunities.* The influence of learning opportunities on lesson planning was  
1104 investigated mainly by focusing on the professional development of PSTs ( $n = 24$ ) and

1105 ISTs ( $n = 8$ ). The learning opportunities have several foci:  $n = 7$  studies focus solely on  
1106 knowledge bases, and  $n = 8$  studies on transforming pPCK to ePCK (Table 7). Half of  
1107 the studies ( $n = 16$ ) investigate learning opportunities that promote knowledge bases and  
1108 their transformation to ePCK<sub>p</sub>. Some studies focus on the promotion of one knowledge  
1109 resource (e.g., Hubbard & Abell, 2005). They report that the use of conceptual knowledge  
1110 in lesson planning increases when ( $N = 6$ ) elementary PSTs participate in an inquiry  
1111 course focussing on CK. However, in several studies, different knowledge bases are  
1112 promoted simultaneously. For example, studies with a method course combining CK and  
1113 PCK as a learning opportunity (e.g., Duruk, 2020; Demirdöğen & Uzuntiryaki-Kondakçı,  
1114 2016; French & Burrows, 2018; Hudson, 2005; Wahbeh & Abd-El-Khalick, 2014).  
1115 Likewise, studies combine phases of planning, teaching, and reflecting to promote the  
1116 transformation of pPCK to ePCK<sub>p</sub>, for example, through lesson study (Aykan & Yıldırım,  
1117 2022; Domingos-Grilo et al., 2012; Grueber & Özgün-Koca, 2013; Larkin, 2017; Ogegbo  
1118 et al., 2019), mentoring programs (Dubek & Doyle-Jones, 2021; Güngören et al., 2020;  
1119 Minken et al., 2021) or microteaching (e.g., Golightly, 2010; Yildiz, 2021). For example,  
1120 Mesci et al. (2020) analysed lesson plans of  $N = 34$  secondary PSTs in a pre-post design.  
1121 They reported that a program including science experiments (CK), best practice lesson  
1122 plans (pPCK), collaborative lesson planning, and microteaching (ePCK<sub>p</sub>, ePCK<sub>t</sub>, ePCK<sub>r</sub>)  
1123 facilitate the integration of nature of science aspects in lesson plans.  
1124 Overall, several studies report increased knowledge and skills for lesson planning through  
1125 different learning opportunities. However, the learning opportunities are manifold and are  
1126 rarely tested in (quasi) experimental-control designs ( $n = 5$ ). For example, Herrington et  
1127 al. (2012) compare two programs, one focussing on CK ('Research experience for  
1128 teachers') and one focussing on CK as well as the preparation of instructional materials  
1129 ('Target inquiry'). They report that teachers who participated in the first program

1130 integrated less inquiry into their lesson plans than teachers who participated in the second  
 1131 program.

1132

1133 **Table 9**

1134 *Factors that influence science teachers' ePCK or pPCK in planning.*

1135

		Studies	<i>n</i>
<b>Independent variable</b>			
cognitive dispositions	teacher knowledge bases (CK, pPCK, PK)	Aykan and Yıldırım (2022); Chan and Yung (2018); Deng et al. (2017); Frederik et al. (1999); Gunckel (2011); Hashweh (1987); Käpylä et al. (2009); Luera et al. (2005); Mesci et al. (2020); Saredidine and BouJaoude (2014); Stender et al. (2017)	12
	subject	So (1997)	1
amplifiers and filters	conditions/context (e.g., planning guides, curriculum material, colleagues, topic)	Bergqvist and Chang Rundgren (2017); Beyer and Davis (2009); Beyer and Davis (2012); Forbes and Davis (2010); Forbes (2011); Forbes (2013); Grove et al. (2009); Gunckel (2011); Hanuscin and Zangori (2016); Henke and Höttecke (2015); Janssen and Lazonder (2016); Jung and Brown (2016); Kademian and Davis (2018); Kang (2017); Käpylä et al. (2009); Karlström and Hamza (2021); Mangiante (2013); Nixon and Campbell (2021); Ross and Cartier (2015); Saredidine and BouJaoude (2014); Schneider (2013); Verma (2008); Wyner (2013a); Wyner (2013b)	24
	orientations and beliefs	Bahçivan (2016); Deng et al. (2017); Käpylä et al. (2009); Mesci et al. (2020); Saredidine and BouJaoude (2014); Stender et al. (2017); Zimmermann et al. (2021)	7
	experience	Chan and Yung (2018); Dawkins et al. (2008); Eshchar-Netz and Vedder-Weiss (2021); Hanuscin et al. (2018); Henke and Höttecke (2015); Saredidine and BouJaoude (2014); So and Watkins (2005); Zimmermann et al. (2021)	8
learning opportunities	Teacher knowledge bases (CK, pPCK, PK)	Bozkurt Altan and Ucuncuoglu (2019); Demirdögen and Uzuntiryaki-Kondakçı (2016); French and Burrows (2018); Grove et al. (2009); Hubbard and Abell (2005); Rinke et al. (2016); Zimmermann et al. (2021)	7

Transformation pPCK -> ePCK <sub>p</sub>	Aykan and Yıldırım (2022); Domingos-Grilo et al. (2012); Dubek and Doyle-Jones (2021); Grueber and Özgün-Koca (2013); Güngören et al. (2020); Kapon and Merzel (2019); Larkin (2017); Minken et al. (2021); Ogegbo et al. (2019)	9
Teacher knowledge bases & Transformation pPCK -> ePCK <sub>p</sub>	Bartels et al. (2019); C. Beyer and Davis (2009); C. J. Beyer and Davis (2012); Biggers and Forbes (2012); Childs et al. (2011); Duruk (2020); Firth and Winter (2007); Forbes and Davis (2010); Forbes (2011); Golightly (2010); Hanuscin and Zangori (2016); Henke and Höttecke (2015); Herrington et al. (2012); Hudson (2005); Juhler (2016); Mesci et al. (2020); Penuel et al. (2009); Qablan (2016); Wahbeh and Abd-El-Khalick (2014); Weitzel and Blank (2020); Yildiz (2021)	21

1136

1137 **Discussion**

1138 In addressing the third research question, we identified four factors influencing science  
1139 teachers' lesson planning: cognitive dispositions, affective dispositions, context, and  
1140 learning opportunities. In the following, we discuss the influence of these four factors on  
1141 lesson planning. Reflecting on the RCM, firstly, some studies show the influence of  
1142 cognitive dispositions. These are already anchored in the RCM: CK, PK, and pPCK. The  
1143 prerequisite of sufficient CK for teacher action is often propagated (e.g., Carlson et al.,  
1144 2019; Gess-Newsome, 2015; Kind, 2009; Krepf et al., 2018). An essential contribution  
1145 to this is the study of Luera et al. (2005) which has a huge sample size ( $N = 234$ ) focussing  
1146 on the planning of inquiry-based science lessons, an important concept in science  
1147 education that is particularly challenging for both PSTs and ISTs (Krämer et al., 2015;  
1148 Lotter et al., 2013; Newman et al., 2004; Yoon et al., 2012). In line with this, many  
1149 researchers discuss a lack of CK as a reason for difficulties in inquiry-based teaching  
1150 (e.g., Lotter et al., 2013; Yoon et al., 2012). These difficulties are already apparent in  
1151 lesson planning (Beyer & Davis, 2012). Other studies support the importance of sufficient

1152 CK for planning specific topics (e.g., photosynthesis, Käpylä et al., 2009; Hashweh, 1987;  
1153 polymerase chain reaction, Chan & Yung, 2018; temperature/heat, Frederik et al., 1999)  
1154 but these results are based on small samples ( $N < 10$  for three of the seven articles). In  
1155 contrast to studies reviewed here, Schröder et al. (2020) cannot find a correlation between  
1156 physics PSTs' CK and their lesson planning skills. Additionally, Vogelsang et al. (2022)  
1157 find a negative correlation between physics PSTs' CK and their lesson planning skills.  
1158 They state that “it seems that just PSTs with high PCK benefit from their CK” (p. 16).  
1159 This aligns with “the idea that PCK is inconceivable without sufficient CK, but that CK  
1160 cannot substitute for PCK” (Voss et al., 2011, p. 964). Therefore, one of the core  
1161 challenges of lesson planning research to be further explored is whether CK is a  
1162 prerequisite for planning and whether PCK mediates the influence.

1163 Regarding the influence of PK on science-specific lesson planning skills, not a single  
1164 study considers PK as a cognitive disposition. However, there is a consensus that “higher  
1165 PPK [pedagogical/psychological knowledge] matters for teaching” (Voss et al., 2011, p.  
1166 965). Likewise, PK is an essential prerequisite for planning (e.g., Zaragoza et al., 2021;  
1167 König, Bremerich-Vos, Buchholtz, & Glutsch, 2020; König, Bremerich-Vos, Buchholtz,  
1168 Fladung, & Glutsch, 2020b). For German language PSTs, König, Bremerich-Vos,  
1169 Buchholtz, and Glutsch (2020) confirm this regarding pedagogical adaptivity, which is  
1170 the ability to tailor instructional tasks to the students' current needs in terms of their  
1171 cognitive level. Declarative PK of adaptivity “is a significant predictor for the situation-  
1172 specific skill of pedagogical adaptivity in written lesson plans” (p. 800). Likewise,  
1173 Schröder et al. (2020) report the influence of PK on lesson planning, including subject-  
1174 specific and generic decisions, especially at the beginning of teacher training. This aligns  
1175 with the findings that PK correlates with PCK in the early phase of teacher training (e.g.,  
1176 Sorge, Kröger et al., 2019; Sorge, Stender, & Neumann, 2019). Accordingly, as with CK,



1177 it remains questionable whether PK as a knowledge resource directly influences lesson  
1178 planning or whether PCK mediates this.

1179 The influence of individual pPCK on lesson planning is undisputed (e.g., Kunter et al.,  
1180 2013; Vogelsang et al., 2022). Again, few studies in this review show this connection.  
1181 Above all, it remains unclear which procedural or conceptual knowledge components are  
1182 decisive as cognitive dispositions for lesson planning (Scholl et al., 2019), with  
1183 substantial consequences for teacher training mainly targeting conceptual knowledge.

1184 In RCM, the transformation of pPCK to ePCK is amplified and filtered by personal  
1185 factors (affective dispositions) and extra-personal factors (context) (Behling et al., 2022;  
1186 Henze & Barendsen, 2019). These are not yet well described, neither in general nor for  
1187 lesson planning (Behling et al., 2022). Among the reviewed publications, we found at  
1188 least some studies examining the influence of amplifiers and filters on lesson planning,  
1189 which might be a promising starting point for further exploring these variables.

1190 The results provide a good insight into a wide range of extra-personal factors. The studies  
1191 show that both educative support and planning guides can provide support in lesson  
1192 planning, whereby the topic- or subject-specificity may represent an advantage over  
1193 generic models. The use of generic planning models or guides has been widespread in  
1194 teacher education since the 1950s (e.g., Tyler, 1950), even though there is little evidence  
1195 about their usefulness (Ball et al., 2007; John, 2006; Karlström & Hamza, 2021) and much  
1196 criticism of their linearity (e.g., John, 2006). Therefore, it would be beneficial for teacher  
1197 education to evaluate topic-specific educational support and subject-specific planning  
1198 guides. Here an experimental comparison of different supports is worthwhile. This has  
1199 been lacking up to now.

1200 The personal factors divide into orientations and beliefs, and teaching experience. Both  
1201 are affective dispositions (Behling et al., 2022). Some authors explore the influence of

1202 experience by comparing experienced and less experienced teachers (e.g., Hanuscin et  
1203 al., 2018) or by studying teachers' lesson planning over time (e.g., Dawkins et al., 2008).  
1204 However, the critical question is not how much time has passed but what has happened  
1205 during this time. Since experience is shaped by context, more precisely by the students,  
1206 the colleagues, or the school environment (Beauchamp & Thomas, 2009), experience  
1207 from time as a student can also be decisive (Zeichner & Tabachnick, 1981).  
1208 Furthermore, experiences can shape orientations and beliefs (Cady et al., 2006),  
1209 impacting lesson planning. The number of studies examining this influence is small,  
1210 although the range of orientations and beliefs, such as motivational orientations,  
1211 constructivist beliefs, and self-efficacy (e.g., Bandura, 1978; Baumert & Kunter, 2006;  
1212 Richardson, 1996), is relatively large. The assumption that orientations and beliefs  
1213 influence planning is congruent with the assumption that orientations and beliefs  
1214 influence teaching (e.g., Dubberke et al., 2008). One study in this review (Bahçivan,  
1215 2016) found no causal relationship between personal beliefs about the nature of  
1216 knowledge and scientific practice in instructional planning. Other studies not reviewed  
1217 here show similar results. For instance, Kotzebue (2022) reports that beliefs about  
1218 learning with digital technologies do not predict the instructional quality of lesson plans.  
1219 Kunter et al. (2013) report that constructivist beliefs do not positively affect the  
1220 instructional quality assessed by students. Contrary to this, studies show an influence of  
1221 constructivist beliefs on instructional planning, e.g., for maths PSTs and ISTs (Lui &  
1222 Bonner, 2016). Therefore, future research on lesson planning should further examine  
1223 which various orientations and beliefs influence science teachers' lesson planning. Lastly,  
1224 almost half of the studies report on the impact of learning opportunities. Some learning  
1225 opportunities aim to gain conceptual knowledge, i.e., CK, PK, and PCK (Figure X). This  
1226 raises the dilemma that has long been discussed in teacher education: the theory-practice

1227 divide (Hammerness et al., 2002; Stender et al., 2017; Zaragoza et al., 2021). Teachers  
1228 struggle to apply conceptual knowledge in practice. This is accompanied by the demand  
1229 to support PSTs to overcome this divide (Zaragoza et al., 2021). Fortunately, our results  
1230 show that most learning opportunities address this issue and support the transformation  
1231 from pPCK to ePCK. This again highlights the importance of strategies for applying and  
1232 integrating conceptual knowledge, which might be the heart of transformation and has  
1233 been under-represented in previous research. In order to examine the impact of such  
1234 learning opportunities aimed at transforming pPCK into ePCK, it would therefore be  
1235 elementary to focus on these strategies. Furthermore, applying knowledge in planning  
1236 might be a crucial step for applying knowledge in teaching practice (Stender et al., 2017;  
1237 Zaragoza et al., 2023). However, in order to be able to give a conclusion about the  
1238 effectiveness of learning opportunities, an experimental design is the way forward. This  
1239 has hardly been practiced so far among the reviewed studies.

## 1240 **Limitations**

1241 Even though we conducted this scoping review rigorously and transparently throughout  
1242 the entire process, our scoping review has some limitations necessary to acknowledge.  
1243 First, in line with the methodology of scoping reviews (e.g., Arksey & O'Malley, 2005;  
1244 Munn et al., 2018; Tricco et al., 2018), we did not critically appraise the quality of the  
1245 reviewed studies. As analysed in RQ1, the field is still emerging and explores lesson  
1246 planning mainly with qualitative approaches. As qualitative research is subject to  
1247 numerous quality criteria and measures of quality control (e.g., Göhner & Krell, 2020;  
1248 Morse, 2022), it might be necessary to analyse to what extent researchers in our sample  
1249 employed and justified an appropriate sampling strategy (e.g., Patton, 1990), tested for  
1250 intra- and interrater reliability (e.g., Landis & Koch, 1977), or considered different

1251 sources of validity evidence (e.g., AERA et al., 2014). Consequently, our attempt to map  
1252 the current state of research in the field of lesson planning might include qualitative  
1253 studies demonstrating rigor and transparency on the one hand and studies of dubious  
1254 quality on the other. However, as previously described by Pham et al. (2014), we also  
1255 encountered the challenge of weighing between an appropriate breadth (i.e., *How many*  
1256 *studies need to be covered?*) and depth (i.e., *How critically do we analyse and appraise*  
1257 *the included studies?*). In line with Arksey and O'Malley (2005), we gave more weight to  
1258 covering all available research activities matching our inclusion criteria. We opted against  
1259 a detailed quality appraisal of the included studies.

1260 Second, in contrast to our intention to cover as much evidence as possible, we only  
1261 included studies published in English, as this allows the international research community  
1262 to receive our findings and interpretations. However, particularly in Germany, empirical  
1263 research on lesson planning forms an emerging field in the last decade that has brought  
1264 forward a couple of considerable and high-quality studies, in parts published in English  
1265 but not focusing on science teachers (e.g., König, Bremerich-Vos, Buchholtz, Fladung,  
1266 & Glutsch, 2020b; König, Bremerich-Vos, Buchholtz, Fladung, & Glutsch, 2020a;  
1267 König, Krepf et al., 2021; Krepf & König, 2022a; Krepf & König, 2022b). Nevertheless,  
1268 as lesson planning comprises not only science-specific but also generic components (i.e.,  
1269 PK, planning strategies), such research might also be noteworthy for researchers in  
1270 science education. Moreover, there are high-quality attempts to analyse interdependent  
1271 planning decisions by application of a vignette test (Scholl et al., 2022) or to measure  
1272 physics teachers' lesson planning skills using a standardized performance assessment  
1273 (Schröder et al., 2020) which are published in German. Moreover, possibly, as we found  
1274 an intensive research activity in Turkey (Table 2), for example, we assume there are more  
1275 publications in the Turkish language that we could not include in this sample. Such

1276 studies would need to be considered in the field as they contribute to our understanding  
1277 of lesson planning processes and developing knowledge and skills for lesson planning  
1278 during teacher training and beyond. However, we did not include them for  
1279 comprehensibility in the international research community. Researchers in the field might  
1280 reconsider whether their research might be better placed in international journals and,  
1281 thus, participate in the emerging international discourse on lesson planning research.

1282 Third, we might not only have missed essential studies published in languages other than  
1283 English but also grey literature (i.e., conference papers, books, dissertations) that  
1284 undoubtedly might also shed light on lesson planning. In addition to the limitations  
1285 mentioned above, we had to ensure the review was still feasible for us to conduct. The  
1286 application of two databases, the detailed search strings, and the inclusion of all studies  
1287 published in peer-reviewed journals beyond the most impactful journals resulted in a  
1288 considerable number of papers we had to screen and check for eligibility (Figure 2). As  
1289 the review builds on  $N = 66$  publications, we argue that this is sufficient to claim that this  
1290 review provides a convincing and reliable overview of the current research landscape.

### 1291 **Establishing common ground in lesson planning research**

1292 This scoping review overviews the current research landscape and characterizes research  
1293 designs and methods for investigating science teachers' lesson planning. Research on  
1294 lesson planning is an emerging field that already covers many aspects of lesson planning.  
1295 Based on the findings of the current scoping review, we propose to specify the RCM  
1296 (Carlson et al., 2019) to establish a heuristic model of science teachers' knowledge and  
1297 skills for lesson planning as a summary of the current state of research on science  
1298 teachers' lesson planning (Figure 9).

1299 As an important reduction of the RCM, we omit the underlying knowledge bases (e.g.  
1300 Assessment knowledge) and cPCK from the model of science teachers' knowledge and  
1301 skills for lesson planning. We assume that there is a knowledge of the community related  
1302 to each of a teacher's professional knowledge bases (cPCK, CK, PK etc.) and that each  
1303 teacher holds only a part of this knowledge. However, since we want to illustrate the  
1304 knowledge and skills for planning of an individual teacher, this collective knowledge is  
1305 out of place in the model. Accordingly, we define CK, PK and pPCK as a teachers'  
1306 personal knowledge, which is influenced by the context in which a teacher finds  
1307 him/herself e.g., colleagues or planning guides from literature. The context, i.e., the extra-  
1308 personal amplifiers and filters, is therefore the outer ring in our model of science teachers'  
1309 knowledge and skills for lesson planning, which influences the teachers' personal  
1310 knowledge and skills (CK, PK, pPCK) presented in the next two inner rings (Figure 9).  
1311 This personal knowledge is the basis for lesson planning and, as pPCK in the RCM, serves  
1312 as a reservoir of knowledge and skills for lesson planning activity (Carlson et al., 2019).  
1313 As a specification of the RCM, we divided pPCK into conceptual and procedural  
1314 knowledge. While conceptual knowledge as PCK facets (according to Magnusson et al.,  
1315 1999) is deeply anchored in RCM, the focus on procedural knowledge is a particularly  
1316 worthwhile extension regarding the use of planning strategies in lesson planning (Table  
1317 6). Although Carlson et al. (2019) also refer to knowledge and skills, a more precise  
1318 description of what these "skills" are is missing. This scoping review, which describes  
1319 procedural knowledge for lesson planning in more detail, provides an essential  
1320 specification of pPCK and ePCK for lesson planning. At the center of the model, we  
1321 locate the application procedural and conceptual knowledge in the lesson-planning  
1322 process, more precisely in the creation and justification of planning decisions. These

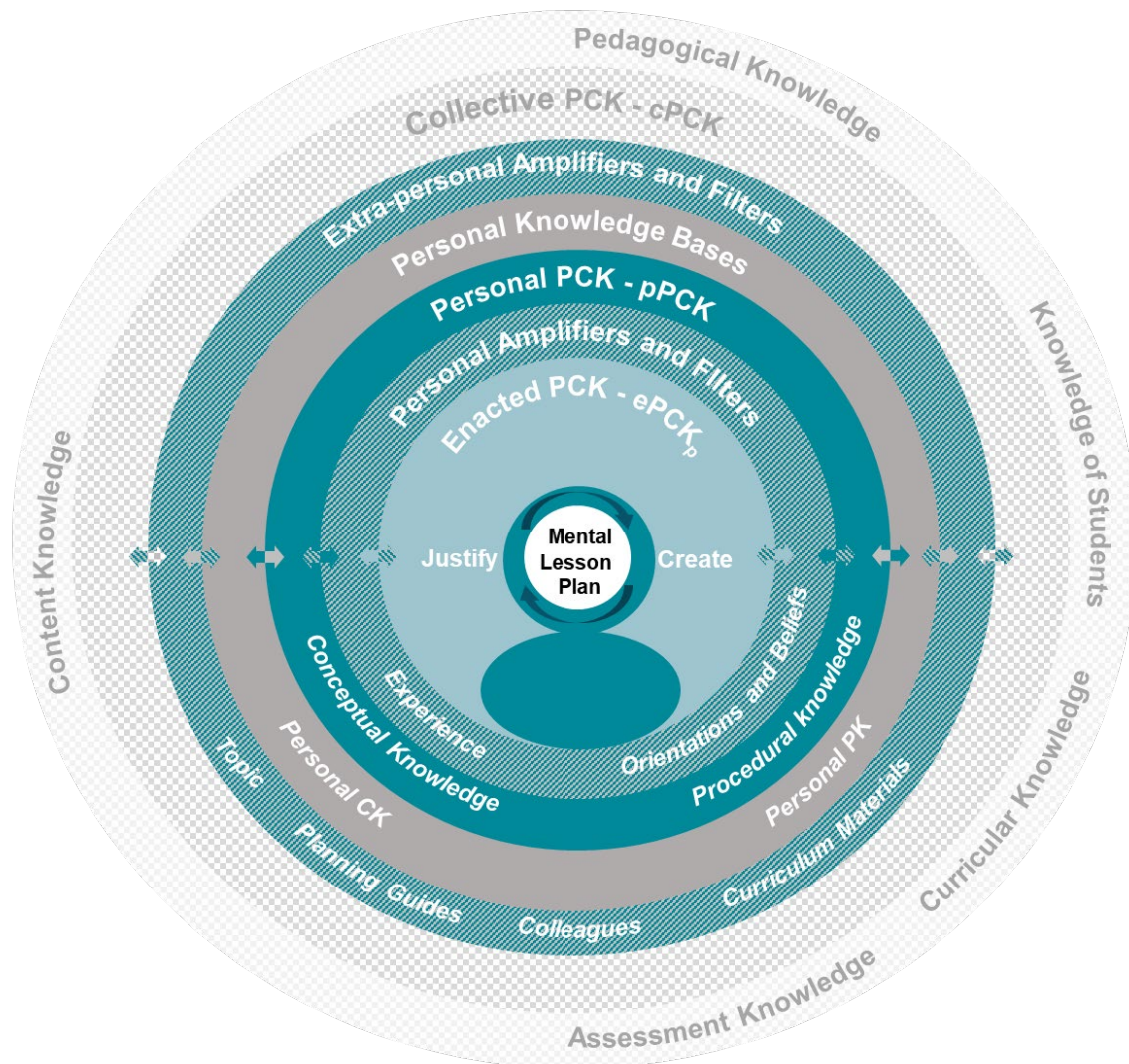
1323 enacted knowledge and skills are defined as ePCK<sub>p</sub> in analogy to the RCM. The interplay  
1324 creation and justification are another lesson planning-specific extension of the RCM.  
1325 Teachers never apply all their pPCK in ePCK. For instance, just because a teacher has  
1326 knowledge about students' ideas, these will not necessarily be taken into account in  
1327 planning (Weitzel & Blank, 2019). Likewise, a teacher who knows a certain planning  
1328 strategy (e.g., from a planning model) will not necessarily apply it in lesson planning due  
1329 to personal orientations and beliefs (Ball, Knobloch and Hoop, 2007). Therefore, which  
1330 part of the knowledge is integrated in lesson planning is influenced by personal amplifiers  
1331 and filters, represented as a ring between pPCK and ePCK. . Behling et al. (2022) assume  
1332 the abilities of noticing and knowledge-based reasoning as amplifiers and filters between  
1333 pPCK and ePCK. However, this is not plausible in the model presented here since both  
1334 are part of procedural knowledge, which is already included in the notions of pPCK and  
1335 ePCK.

1336 The planning process results in a mental lesson plan located in the middle of the model.  
1337 Writing down this plan requires additional knowledge and skills not mentioned in this  
1338 model (e.g., “creative and robust writing ability to communicate an idea”, Kaya & Nafiz  
1339 Kaya, 2023, p.18). It might be amplified and filtered by further personal and extra-  
1340 personal factors (e.g., formal requirements for written lesson plans). Accordingly, written  
1341 lesson plans as products of lesson planning processes are not part of the model. However,  
1342 written lesson plans are essential for reconstructing science teachers' knowledge and skills  
1343 for lesson planning.

1344 **Figure 9**

1345 *Heuristic model of science teachers' knowledge and skills for lesson planning (based on*  
1346 *Carlson et al., 2019)*

1347 *Note.* Double arrows indicate the flow of knowledge and skills between the concentric  
 1348 circles. The two outer circles are grayed out as we only focus on an individual teacher’s  
 1349 knowledge and skills.



1350  
 1351 As the field, in its current state, is still diverse in theoretical conceptualizations, research  
 1352 questions, and methodological approaches, we believe the heuristic model of science  
 1353 teachers’ knowledge and skills for lesson planning (Figure 9) might help the emerging  
 1354 field to strive in the same direction. More specifically, we consider the following two  
 1355 suggestions fruitful for future research.  
 1356 First, in the reviewed studies, we do not find approaches that attempt to cover the full  
 1357 range of lesson planning competence as a continuum (Blömeke et al., 2015). For instance,



1358 the flows of knowledge and skills between pPCK and ePCK are barely investigated so  
1359 far. Further insights into the transformation of cPCK – an underrepresented construct in  
1360 PCK research (Chan & Hume, 2019) – to pPCK might help teacher educators tailor  
1361 instruction in teacher training. Suppose researchers figure out which aspects of cPCK, as  
1362 the shared knowledge of researchers and teacher educators, are accessible and challenging  
1363 for PSTs to understand and apply. In that case, these findings might inform teacher  
1364 training and help PSTs acquire pPCK and transform it into their ePCK.

1365 In that sense, König et al. (2021) note that situation-specific skills are barely considered  
1366 in empirical research. This is astonishing because these skills refer to integrate conceptual  
1367 and procedural knowledge acquired during teacher training, a crucial goal of teacher  
1368 education. Hence, science teachers need to be able to interconnect their professional  
1369 knowledge and apply it when planning lessons in varying contexts for varying students  
1370 (Zaragoza et al., 2021; Zaragoza et al., 2023). Consequently, solely identifying PSTs'  
1371 conceptual knowledge in lesson plans skips the process when PSTs integrate their  
1372 conceptual knowledge and, thus, transform it into procedural knowledge. As only a few  
1373 studies in this sample consider procedural knowledge (Figure 8), we suggest shedding  
1374 further light on the complexity of lesson-planning processes and science teachers'  
1375 strategies. In particular, we noticed that researchers mainly analyse PSTs' planning  
1376 strategies and barely experienced and high-quality ISTs' planning strategies. It is  
1377 reasonable to assume that the most sophisticated manifestations of knowledge and skills  
1378 for lesson planning might not be observed in PSTs who still learn lesson planning but in  
1379 teachers who provide high-quality instruction resulting in students' performance gains.  
1380 Such research designs might explicitly target successful teachers' strategies and, thus, add  
1381 to older research on the differences between novices and experts in lesson planning  
1382 (Borko & Livingston, 1989; Westerman, 1991). In addition, few attempts target the

1383 performance in terms of criteria to evaluate the quality of mental or written lesson plans.  
1384 Although there are some rubrics available to assess lesson plan quality (e.g., Jacobs et al.,  
1385 2008; Ndiokubwayo et al., 2022), these instruments do not refer to a consensual set of  
1386 quality criteria and differ significantly in the extent to which they were tested for the  
1387 validity of test score interpretations. More strikingly, such instruments were not used in  
1388 our sample to quantify the quality of the analysed lesson plans. This might facilitate  
1389 further correlational analyses and help gain insights into the relationship between the  
1390 performance quality and other quantifiable variables (e.g., cPCK measured with  
1391 questionnaires).

1392 These observations raise the following research questions:

- 1393 – How exactly do amplifiers and filters (e.g., orientations and beliefs) affect the quality  
1394 of science teachers' lesson plans?
- 1395 – What situation-specific skills do science teachers apply to master the cognitive  
1396 demands they encounter in lesson planning processes?
- 1397 – How do experienced and successful expert teachers proceed in their lesson-planning  
1398 processes?
- 1399 – How can lesson plan quality be defined and assessed validly?

1400

1401 Second, in addition to the research gaps mentioned above, we suggest reflecting on the  
1402 research methods employed to investigate knowledge and skills for lesson planning. Most  
1403 researchers in the field analyse written lesson plans or conduct interviews and ask PSTs  
1404 to reflect on their planning processes or products. These approaches might be  
1405 undoubtedly appropriate to answer specific research questions. However, as far as lesson  
1406 planning processes are concerned, the nature of lesson planning raises methodological  
1407 difficulties: Unlike the act of teaching, which can be observed in class to its full extent,

1408 planning processes can be tedious, fragmented, messy, and mainly take place in science  
1409 teachers' thoughts. Consequently, they are not directly observable. When asking science  
1410 teachers about their planning processes in interviews, it should be considered that such  
1411 self-reports may barely cover the chronology of the thoughts to its full extent.  
1412 Accordingly, it is questionable whether self-reports are at all suitable for surveying ePCK  
1413 or whether self-reports can actually only be used to survey pPCK, since the report on the  
1414 planning process may deviate from the real process. It can also be questioned whether  
1415 planning products are suitable for investigating ePCK.

1416 Some authors (e.g., Stender and Brückmann, 2020) state that planning products are only  
1417 artefacts that are not suitable to capture the ePCK of a teacher, but the pPCK. Stender and  
1418 Brückmann (2020) even claim that the only way to investigate ePCK is to observe  
1419 classroom teaching. Alonzo et al. (2019), on the other hand, say that ePCK cannot be  
1420 captured at all because it is the tacit knowledge generated in the moment, which, once  
1421 verbalised, becomes pPCK. As we assume that there is a specific ePCK for lesson  
1422 planning, we also assume that it differs from the ePCK for teaching and could be collected  
1423 separately. However, which data sources (e.g., written planning products, think-aloud-  
1424 protocols of planning processes, self-reports about planning processes) are particularly  
1425 suitable for this is debatable. We assume that both self-reports on planning processes and  
1426 products and the planning processes and products themselves are suitable sources for  
1427 capturing ePCK. However, the suitability of data sources to capture ePCK still needs to  
1428 be verified, for instance, by comparing planning processes collected via self-reports and  
1429 those collected via think-aloud in real planning situations.

1430 Beyond the methods covered in the review, new paths of data collection are also  
1431 conceivable. Researchers might capture PSTs' or ISTs' planning processes by asking  
1432 them to take note of any thoughts on planning a specific lesson over time as precisely as

1433 possible. Such a methodological approach might be demanding for the participants.  
1434 However, it might be worth exploring the utility of such approaches to capture lesson-  
1435 planning processes as an essential part of teachers' knowledge and skills.  
1436 Moreover, we encourage researchers in the field to plan and employ rigorous  
1437 methodological research designs, particularly experimental-control-group designs. We  
1438 understand that as long as there is no consensual theoretical basis, such as a competence  
1439 model or an overarching theory, researchers need to explore the construct in question to  
1440 gain insights into the nature of lesson planning. Picking up the idea of relating to lesson  
1441 planning as a competence (e.g., König et al., 2021), we assume that it is possible to  
1442 purposefully select variables from the heuristic model and analyse their role for the  
1443 situation-specific skills (e.g., the planning strategies a science teacher applies) or for the  
1444 performance (i.e., the mental or written lesson plan). However, claiming that such  
1445 relationships are causal requires comparing an intervention and a control group.  
1446 These considerations raise the following research questions:  
1447 – How to capture ePCK in lesson planning processes reliably and validly?  
1448 – To what extent does science teachers' CK influence the quality of their lesson plans?  
1449 – To what extent does science teachers' PK influence the quality of their lesson plans?  
1450 – To what extent does science teachers' pPCK influence the quality of their lesson  
1451 plans?  
1452 – To what extent does experience influence the planning strategies science teachers  
1453 employ?  
1454 – To what extent do science teachers' affective-motivational dispositions influence the  
1455 quality of their lesson plans?  
1456 – What variables influence the sophistication of a science teacher's justifications for  
1457 planning decisions?

1458 **Funding details**

1459 The project K2Teach (*Know how to teach*) is part of the “Qualitätsoffensive  
1460 Lehrerbildung” in Germany, a joint initiative of the Federal Government and the Länder  
1461 which aims to improve the quality of teacher training. The program is funded by the  
1462 German Federal Ministry of Education and Research (grant number 01JA1802). The  
1463 authors are responsible for the content of this publication.

1464 **Disclosure Statement**

1465 The authors report there are no competing interests to declare.

1466 **Notes on Contributors**

1467 Leroy Großmann is a Ph.D student in Biology education at Freie Universität Berlin,  
1468 Germany.

1469 Maren-Koberstein-Schwarz is a Ph.D student in Biology education at Universität  
1470 Hildesheim, Germany.

1471 Daniel Scholl is a Professor of Pedagogy at Universität Siegen, Germany.

1472 Dirk Krüger is a Professor of Biology Education at Freie Universität Berlin, Germany.

1473 Anke Meisert is a Professor of Biology Education at Universität Hildesheim, Germany.

1474 **References**

1475 *References marked with an asterisk indicate the N = 66 studies included in the analysis.*

1476 AERA, APA, & NCME. (2014). Standards for educational and psychological testing.  
1477 American Psychological Association.

1478 Ahlgren, P., & Jarneving, B. (2008). Bibliographic coupling, common abstract stems and  
1479 clustering: A comparison of two document-document similarity approaches in the

1480 context of science mapping. *Scientometrics*, 76(2), 273–290.  
1481 <https://doi.org/10.1007/s11192-007-1935-1>

1482 Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science  
1483 teachers' PCK in action: enacted and personal PCK. In A. Hume, R. Cooper, & A.  
1484 Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers'*  
1485 *knowledge for teaching science* (pp. 271–286). Springer.  
1486 [https://doi.org/10.1007/978-981-13-5898-2\\_12](https://doi.org/10.1007/978-981-13-5898-2_12)

1487 American Association for the Advancement of Science. (1993). *Benchmarks for Science*  
1488 *Literacy*. Oxford.

1489 Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological  
1490 framework. *International Journal of Social Research Methodology*, 8(1), 19–32.  
1491 <https://doi.org/10.1080/1364557032000119616>

1492 Australian Institute for Teaching and School Leadership. (2018). *Australian Professional*  
1493 *Standards for Teachers*. [https://www.aitsl.edu.au/docs/default-source/national-](https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standards-for-teachers.pdf)  
1494 [policy-framework/australian-professional-standards-for-teachers.pdf](https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standards-for-teachers.pdf)

1495 \*Aykan, A., & Yıldırım, B. (2022). The Integration of a Lesson Study Model into  
1496 Distance STEM Education during the COVID-19 Pandemic: Teachers' Views and  
1497 Practice. *Technology, Knowledge and Learning*, 27(2), 609–637.  
1498 <https://doi.org/10.1007/s10758-021-09564-9>

1499 \*Bahçivan, E. (2016). Investigating the relationships among PSTs' teaching beliefs: Are  
1500 epistemological beliefs central? *Educational Studies*, 42(2), 221–238.  
1501 <https://doi.org/10.1080/03055698.2016.1160823>

- 1502 Ball, A. L., Knobloch, N. A., & Hoop, S. (2007). The Instructional Planning Experiences  
1503 of Beginning Teachers. *Journal of Agricultural Education*, 48(2), 56-65.  
1504 <http://pubs.aged.tamu.edu/jae>
- 1505 Bandura, A. (1978). Self-efficacy: Toward a unifying theory of behavioral change.  
1506 *Advances in Behaviour Research and Therapy*, 1(4), 139–161.  
1507 [https://doi.org/10.1016/0146-6402\(78\)90002-4](https://doi.org/10.1016/0146-6402(78)90002-4)
- 1508 Barendsen, E., & Henze, I. (2019). Relating Teacher PCK and Teacher Practice Using  
1509 Classroom Observation. *Research in Science Education*, 49(5), 1141–1175.  
1510 <https://doi.org/10.1007/s11165-017-9637-z>
- 1511 \*Bartels, S. L., Rupe, K. M., & Lederman, J. S. (2019). Shaping Preservice Teachers’  
1512 Understandings of STEM: A Collaborative Math and Science Methods Approach.  
1513 *Journal of Science Teacher Education*, 30(6), 666–680.  
1514 <https://doi.org/10.1080/1046560X.2019.1602803>
- 1515 Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften.  
1516 *Zeitschrift Für Erziehungswissenschaft*, 9(4), 469–520.  
1517 <https://doi.org/10.1007/s11618-006-0165-2>
- 1518 Beauchamp, C., & Thomas, L. (2009). Understanding teacher identity: an overview of  
1519 issues in the literature and implications for teacher education. *Cambridge Journal*  
1520 *of Education*, 39(2), 175–189. <https://doi.org/10.1080/03057640902902252>
- 1521 Behling, F., Förtsch, C., & Neuhaus, B. J. (2022). The Refined Consensus Model of  
1522 Pedagogical Content Knowledge (PCK): Detecting Filters between the Realms of  
1523 PCK. *Education Sciences*, 12(9), 592. <https://doi.org/10.3390/educsci12090592>
- 1524 \*Bergqvist, A., & Chang Rundgren, S.-N. (2017). The influence of textbooks on teachers’  
1525 knowledge of chemical bonding representations relative to students’ difficulties

- 1526 understanding. *Research in Science & Technological Education*, 35(2), 215–237.  
1527 <https://doi.org/10.1080/02635143.2017.1295934>
- 1528 Bergqvist, A., Drechsler, M., & Chang Rundgren, S.-N. (2016). Upper Secondary  
1529 Teachers' Knowledge for Teaching Chemical Bonding Models. *International*  
1530 *Journal of Science Education*, 38(2), 298–318.  
1531 <https://doi.org/10.1080/09500693.2015.1125034>
- 1532 \*Beyer, C., & Davis, E. A. (2009). Supporting Preservice Elementary Teachers' Critique  
1533 and Adaptation of Science Lesson Plans Using Educative Curriculum Materials.  
1534 *Journal of Science Teacher Education*, 20(6), 517. [https://doi.org/10.1007/s10972-](https://doi.org/10.1007/s10972-009-9148-5)  
1535 [009-9148-5](https://doi.org/10.1007/s10972-009-9148-5)
- 1536 \*Beyer, C. J., & Davis, E. A. (2012). Learning to critique and adapt science curriculum  
1537 materials: Examining the development of preservice elementary teachers'  
1538 pedagogical content knowledge. *Science Education*, 96(1), 130–157.  
1539 <https://doi.org/10.1002/sce.20466>
- 1540 \*Biggers, M., & Forbes, C. T. (2012). Balancing Teacher and Student Roles in  
1541 Elementary Classrooms: Preservice elementary teachers' learning about the inquiry  
1542 continuum. *International Journal of Science Education*, 34(14), 2205–2229.  
1543 <https://doi.org/10.1080/09500693.2012.694146>
- 1544 Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies:  
1545 competence viewed as a continuum. *Zeitschrift Für Psychologie*, 223(1), 3–13.  
1546 <https://doi.org/10.1027/2151-2604/a000194>
- 1547 Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in  
1548 mathematics instruction by expert and novice teachers. *American Educational*  
1549 *Research Journal*, 26(4), 473–498.



- 1550 \*Bozkurt Altan, E., & Ucuncuoglu, I. (2019). Examining the Development of Pre-Service  
 1551 Science Teachers' STEM-Focused Lesson Planning Skills. *Eurasian Journal of*  
 1552 *Educational Research*, 19(83), 103–124.  
 1553 <https://dergipark.org.tr/en/pub/ejer/issue/50254/648539>
- 1554 Cady, J., Meier, S. L., & Lubinski, C. A. (2006). The mathematical tale of two teachers:  
 1555 A longitudinal study relating mathematics instructional practices to level of  
 1556 intellectual development. *Mathematics Education Research Journal*, 18(1), 3–26.  
 1557 <https://doi.org/10.1007/BF03217427>
- 1558 Campbell, T., Gray, R., Fazio, X., & van Driel, J. H. (2022). Research on Secondary  
 1559 Science Teacher Preparation. In J. A. Luft & M. G. Jones (Eds.), *Handbook of*  
 1560 *Research on Science Teacher Education* (pp. 97–118). Routledge.  
 1561 <https://doi.org/10.4324/9781003098478-10>
- 1562 Carlson, J., Daehler, K. R., Alonzo, A., Barendsen, E., Berry, A., Borowski, A.,  
 1563 Carpendale, J., Chan, K. K. H., Cooper, R., Friedrichsen, P., Gess-Newsome, J.,  
 1564 Henze\_Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga,  
 1565 E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The refined consensus  
 1566 model of pedagogical content knowledge in science education. In A. Hume, R.  
 1567 Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in*  
 1568 *teachers' knowledge for teaching science* (pp. 77–92). Springer.  
 1569 [https://doi.org/10.1007/978-981-13-5898-2\\_2](https://doi.org/10.1007/978-981-13-5898-2_2)
- 1570 Céim. (2020). Standards for Initial Teacher Education.  
 1571 [https://www.teachingcouncil.ie/en/news-events/latest-news/ceim-standards-for-](https://www.teachingcouncil.ie/en/news-events/latest-news/ceim-standards-for-initial-teacher-education.pdf)  
 1572 [initial-teacher-education.pdf](https://www.teachingcouncil.ie/en/news-events/latest-news/ceim-standards-for-initial-teacher-education.pdf)

- 1573 Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: literature review of  
1574 how science teachers' pedagogical content knowledge is investigated in empirical  
1575 studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical  
1576 content knowledge in teachers' knowledge for teaching science* (pp. 3–76).  
1577 Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_1](https://doi.org/10.1007/978-981-13-5898-2_1)
- 1578 \*Chan, K. K. H., & Yung, B. H. W. (2018). Developing Pedagogical Content Knowledge  
1579 for teaching a new topic: more than teaching experience and subject matter  
1580 knowledge. *Research in Science Education*, 48(2), 233–265.  
1581 <https://doi.org/10.1007/s11165-016-9567-1>
- 1582 \*Childs, A., Sorensen, P., & Twidle, J. (2011). Using the Internet in science teaching?  
1583 Issues and challenges for initial teacher education. *Technology, Pedagogy and  
1584 Education*, 20(2), 143–160. <https://doi.org/10.1080/1475939X.2011.588413>
- 1585 Clark, C. M., & Yinger, R. J. (1980). *The Hidden World of Teaching: Implications of  
1586 Research on Teacher Planning*. Institute for Research on Teaching, College of  
1587 Education, Michigan State University.
- 1588 Cofré, H., González-Weil, C., Vergara, C., Santibáñez, D., Ahumada, G., Furman, M.,  
1589 Podesta, M. E., Camacho, J., Gallego, R., & Pérez, R. (2015). Science Teacher  
1590 Education in South America: The Case of Argentina, Colombia and Chile. *Journal  
1591 of Science Teacher Education : The Official Journal of the Association for the  
1592 Education of Teachers in Science*, 26(1), 45–63. [https://doi.org/10.1007/s10972-  
1593 015-9420-9](https://doi.org/10.1007/s10972-015-9420-9)
- 1594 Colquhoun, H. L., Levac, D., O'Brien, K. K., Straus, S. E., Tricco, A. C., Perrier, L.,  
1595 Kastner, M., & Moher, D. (2014). Scoping reviews: Time for clarity in definition,

1596 methods, and reporting. *Journal of Clinical Epidemiology*, 67(12), 1291–1294.  
1597 <https://doi.org/10.1016/j.jclinepi.2014.03.013>

1598 Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing  
1599 beginning teacher effectiveness: the potential of performance assessments.  
1600 *Educational Assessment, Evaluation and Accountability*, 25(3), 179–204.  
1601 <https://doi.org/10.1007/s11092-013-9163-0>

1602 Davis, E. A. (2006). Preservice elementary teachers' critique of instructional materials for  
1603 science. *Science Education*, 90(2), 348–375. <https://doi.org/10.1002/sce.20110>

1604 Davis, E. A., & Krajcik, J. S. (2005). Designing Educative Curriculum Materials to  
1605 Promote Teacher Learning. *Educational Researcher*, 34(3), 3–14.  
1606 <https://doi.org/10.3102/0013189X034003003>

1607 Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges New Science Teachers Face.  
1608 *Review of Educational Research*, 76(4), 607–651.  
1609 <https://doi.org/10.3102/00346543076004607>

1610 Davis, E. A., & Smithey, J. (2009). Beginning teachers moving toward effective  
1611 elementary science teaching. *Science Education*, 93(4), 745–770.  
1612 <https://doi.org/10.1002/sce.20311>

1613 \*Dawkins, K. R., Dickerson, D. L., McKinney, S. E., & Butler, S. (2008). Teaching  
1614 Density to Middle School Students: Preservice Science Teachers' Content  
1615 Knowledge and Pedagogical Practices. *The Clearing House: A Journal of*  
1616 *Educational Strategies, Issues and Ideas*, 82(1), 21–26.  
1617 <https://doi.org/10.3200/TCHS.82.1.21-26>

- 1618 Deehan, J., MacDonald, A., & Morris, C. (2022). A scoping review of interventions in  
1619 primary science education. *Studies in Science Education*, 1–43.  
1620 <https://doi.org/10.1080/03057267.2022.2154997>
- 1621 \*Demirdöğen, B., & Uzuntiryaki-Kondakçı, E. (2016). Closing the gap between beliefs  
1622 and practice: Change of pre-service chemistry teachers' orientations during a PCK-  
1623 based NOS course. *Chemistry Education Research and Practice*, 17(4), 818–841.  
1624 <https://doi.org/10.1039/c6rp00062b>
- 1625 \*Deng, F., Chai, C. S., So, H.-J., Qian, Y., & Chen, L. (2017). Examining the validity of  
1626 the technological pedagogical content knowledge (TPACK) framework for  
1627 preservice chemistry teachers. *Australasian Journal of Educational Technology*,  
1628 33(3). <https://doi.org/10.14742/ajet.3508>
- 1629 \*Domingos-Grilo, P., Reis-Grilo, C., Ruiz, C., & Mellado, V. (2012). An action-research  
1630 programme with secondary education teachers on teaching and learning  
1631 photosynthesis. *Journal of Biological Education*, 46(2), 72–80.  
1632 <https://doi.org/10.1080/00219266.2011.587522>
- 1633 Dubberke, T., Kunter, M., McElvany, N., Brunner, M., & Baumert, J. (2008).  
1634 Lerntheoretische Überzeugungen von Mathematiklehrkräften. *Zeitschrift Für*  
1635 *Pädagogische Psychologie*, 22(34), 193–206. [https://doi.org/10.1024/1010-](https://doi.org/10.1024/1010-0652.22.34.193)  
1636 [0652.22.34.193](https://doi.org/10.1024/1010-0652.22.34.193)
- 1637 \*Dubek, M., & Doyle-Jones, C. (2021). Faculty Co-Teaching with Their Teacher  
1638 Candidates in the Field: Co-Planning, Co-Instructing, and Co-Reflecting for STEM  
1639 Education Teacher Preparation. *The Teacher Educator*, 1–21.
- 1640 \*Duruk, Ü. (2020). Influence of a Socially-Mediated Contextual Professional  
1641 Development Program on Prospective Science Teachers' Understandings of Nature

- 1642 of Science, and Integrating It into Their Instructional Planning. *International Online*  
1643 *Journal of Education and Teaching*, 7(3), 912–943.  
1644 <https://eric.ed.gov/?id=ej1258700>
- 1645 \*Eshchar-Netz, L., & Vedder-Weiss, D. (2021). Teacher learning in communities of  
1646 practice: The affordances of co-planning for novice and veteran teachers' learning.  
1647 *Journal of Research in Science Teaching*, 58(3), 366–391.  
1648 <https://doi.org/10.1002/tea.21663>
- 1649 Evagorou, M., Dillon, J., Viiri, J., & Albe, V. (2015). Pre-service Science Teacher  
1650 Preparation in Europe: Comparing Pre-service Teacher Preparation Programs in  
1651 England, France, Finland and Cyprus. *Journal of Science Teacher Education : The*  
1652 *Official Journal of the Association for the Education of Teachers in Science*, 26(1),  
1653 99–115. <https://doi.org/10.1007/s10972-015-9421-8>
- 1654 \*Firth, R., & Winter, C. (2007). Constructing education for sustainable development: The  
1655 secondary school geography curriculum and initial teacher training. *Environmental*  
1656 *Education Research*, 13(5), 599–619. <https://doi.org/10.1080/13504620701659079>
- 1657 \*Forbes, C. T. (2011). Preservice elementary teachers' adaptation of science curriculum  
1658 materials for inquiry-based elementary science. *Science Education*, 95(5), 927–955.  
1659 <https://doi.org/10.1002/sce.20444>
- 1660 \*Forbes, C. T. (2013). Curriculum-Dependent and Curriculum-Independent Factors in  
1661 Preservice Elementary Teachers' Adaptation of Science Curriculum Materials for  
1662 Inquiry-Based Science. *Journal of Science Teacher Education*, 24(1), 179–197.  
1663 <https://doi.org/10.1007/s10972-011-9245-0>
- 1664 \*Forbes, C. T., & Davis, E. A. (2010). Curriculum design for inquiry: Preservice  
1665 elementary teachers' mobilization and adaptation of science curriculum materials.

1666 Journal of Research in Science Teaching, 47(7), 820–839.  
1667 <https://doi.org/10.1002/tea.20379>

1668 \*Frederik, I., van Der Valk, T., Leite, L., & Thorén, I. (1999). Pre-service Physics  
1669 Teachers and Conceptual Difficulties on Temperature and Heat. European Journal  
1670 of Teacher Education, 22(1), 61–74. <https://doi.org/10.1080/0261976990220105>

1671 \*French, D. A., & Burrows, A. C. (2018). Evidence of Science and Engineering Practices  
1672 in Preservice Secondary Science Teachers' Instructional Planning. Journal of  
1673 Science Education and Technology, 27(6), 536–549.  
1674 <https://doi.org/10.1007/s10956-018-9742-4>

1675 Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including  
1676 PCK: Results of the thinking from the PCK Summit. In A. Berry, P. M.  
1677 Friedrichsen, & J. Loughran (Eds.), Teaching and learning in science series. Re-  
1678 examining pedagogical content knowledge in science education (pp. 28–42).  
1679 Routledge.

1680 Göhner, M., & Krell, M. (2020). Qualitative Inhaltsanalyse in  
1681 naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von  
1682 Gütekriterien: Ein Review. Zeitschrift für Didaktik der Naturwissenschaften, 26,  
1683 207–225. <https://doi.org/10.1007/s40573-020-00111-0>

1684 Goldston, M. J., Dantzler, J., Day, J., & Webb, B. (2013). A Psychometric Approach to  
1685 the Development of a 5E Lesson Plan Scoring Instrument for Inquiry-Based  
1686 Teaching. Journal of Science Teacher Education, 24(3), 527–551.  
1687 <https://doi.org/10.1007/s10972-012-9327-7>

- 1688 \*Golightly, A. (2010). Microteaching to Assist Geography Teacher-Trainees in  
1689 Facilitating Learner-Centered Instruction. *Journal of Geography*, 109(6), 233–242.  
1690 <https://doi.org/10.1080/00221341.2010.509512>
- 1691 Großmann, L., & Krüger, D. (2022). Welche Rolle spielt das fachdidaktische Wissen von  
1692 Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift  
1693 Für Didaktik Der Naturwissenschaften*, 28(1). [https://doi.org/10.1007/s40573-022-  
1694 00141-w](https://doi.org/10.1007/s40573-022-00141-w)
- 1695 Grossmann, P. (1990). *The making of a teacher: Teacher knowledge & teacher education.*  
1696 Teachers College Press.
- 1697 Großschedl, J., Harms, U., Kleickmann, T., & Glowinski, I. (2015). Preservice biology  
1698 teachers' professional knowledge: structure and learning opportunities. *Journal of  
1699 Science Teacher Education*, 26(3), 291–318. [https://doi.org/10.1007/s10972-015-  
1700 9423-6](https://doi.org/10.1007/s10972-015-9423-6)
- 1701 \*Grove, C. M., Dixon, P. J., & Pop, M. M. (2009). Research experiences for teachers:  
1702 Influences related to expectancy and value of changes to practice in the American  
1703 classroom. *Professional Development in Education*, 35(2), 247–260.  
1704 <https://doi.org/10.1080/13674580802532712>
- 1705 \*Grueber, D., & Özgün-Koca, S. A. (2013). Developing Prospective Teachers'  
1706 Knowledge to Foster and Inspire Reasoning in STEM. *Teacher Education and  
1707 Practice*, 26(2).  
1708 [https://link.gale.com/apps/doc/A514683030/AONE?u=anon~5be80fff&sid=googl  
1709 eScholar&xid=71fc20d4](https://link.gale.com/apps/doc/A514683030/AONE?u=anon~5be80fff&sid=googleScholar&xid=71fc20d4)

- 1710 \*Gunckel, K. L. (2011). Mediators of a Preservice Teacher's Use of the Inquiry-  
1711 Application Instructional Model. *Journal of Science Teacher Education*, 22(1), 79–  
1712 100. <https://doi.org/10.1007/s10972-010-9223-y>
- 1713 \*Güngören, S. Ç., Hasançebi, F. Y., & Mesci, G. (2020). The Effect of Teacher  
1714 Evaluation and Self-Evaluation on Pre-Service Teachers' Inquiry-Based 5E Lesson  
1715 Plan Design and Teaching Practice. *Educational Policy Analysis and Strategic  
1716 Research*, 15(4), 367–388. <https://eric.ed.gov/?id=ej1280114>
- 1717 Hammerness, K., Darling-Hammond, L., & Shulman, L. (2002). Toward Expert  
1718 Thinking: How curriculum case writing prompts the development of theory-based  
1719 professional knowledge in student teachers. *Teaching Education*, 13(2), 219–243.  
1720 <https://doi.org/10.1080/1047621022000007594>
- 1721 \*Hanuscin, D. L., Cisterna, D., & Lipsitz, K. (2018). Elementary teachers' Pedagogical  
1722 Content Knowledge for teaching structure and properties of matter. *Journal of  
1723 Science Teacher Education : The Official Journal of the Association for the  
1724 Education of Teachers in Science*, 29(8), 665–692.  
1725 <https://doi.org/10.1080/1046560X.2018.1488486>
- 1726 \*Hanuscin, D. L., & Zangori, L. (2016). Developing Practical Knowledge of the Next  
1727 Generation Science Standards in Elementary Science Teacher Education. *Journal  
1728 of Science Teacher Education*, 27(8), 799–818. [https://doi.org/10.1007/s10972-  
1729 016-9489-9](https://doi.org/10.1007/s10972-016-9489-9)
- 1730 \*Hashweh, M. Z. (1987). Effects of subject-matter knowledge in the teaching of biology  
1731 and physics. *Teaching and Teacher Education*, 3(2), 109–120.  
1732 [https://doi.org/10.1016/0742-051X\(87\)90012-6](https://doi.org/10.1016/0742-051X(87)90012-6)



- 1733 \*Henke, A., & Höttecke, D. (2015). Physics Teachers' Challenges in Using History and  
1734 Philosophy of Science in Teaching. *Science & Education*, 24(4), 349–385.  
1735 <https://doi.org/10.1007/s11191-014-9737-3>
- 1736 Henze, I., & Barendsen, E. (2019). Unravelling Student Science Teachers' pPCK  
1737 Development and the Influence of Personal Factors Using Authentic Data Sources.  
1738 In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content*  
1739 *knowledge in teachers' knowledge for teaching science* (pp. 203–223). Springer.  
1740 [https://doi.org/10.1007/978-981-13-5898-2\\_9](https://doi.org/10.1007/978-981-13-5898-2_9)
- 1741 \*Herrington, D. G., Luxford, K., & Yeziarski, E. J. (2012). Target Inquiry: Helping  
1742 Teachers Use a Research Experience To Transform Their Teaching Practices.  
1743 *Journal of Chemical Education*, 89(4), 442–448.  
1744 <https://doi.org/10.1021/ed1006458>
- 1745 \*Hubbard, P., & Abell, S. (2005). Setting Sail or Missing the Boat: Comparing the Beliefs  
1746 of Preservice Elementary Teachers With and Without an Inquiry-Based Physics  
1747 Course. *Journal of Science Teacher Education*, 16(1), 5–25.  
1748 <https://doi.org/10.1007/s10972-005-6026-7>
- 1749 \*Hudson, P. (2005). Informing Future Learning Designs in Preservice Teacher Education  
1750 through Quantitative Research: A Primary Science Example. *Journal of Learning*  
1751 *Design*, 1(1), 33–44. <https://eric.ed.gov/?id=ej1066466>
- 1752 Irmer, M., Traub, D., Böhm, M., Förtsch, C., & Neuhaus, B. J. (2023). Using Video-  
1753 Based Simulations to Foster pPCK/ePCK—New Thoughts on the Refined  
1754 Consensus Model of PCK. *Education Sciences*, 13(3), 261.  
1755 <https://doi.org/10.3390/educsci13030261>

- 1756 Jack, B. M., & Lin, H. (2017). Making learning interesting and its application to the  
1757 science classroom. *Studies in Science Education*, 53(2), 137–164.  
1758 <https://doi.org/10.1080/03057267.2017.1305543>
- 1759 Jacobs, C. L., Martin, S. N., & Otieno, T. C. (2008). A science lesson plan analysis  
1760 instrument for formative and summative program evaluation of a teacher education  
1761 program. *Science Education*, 92(6), 1096–1126. <https://doi.org/10.1002/sce.20277>
- 1762 \*Janssen, N., & Lazonder, A. W. (2016). Supporting pre-service teachers in designing  
1763 technology-infused lesson plans. *Journal of Computer Assisted Learning*, 32(5),  
1764 456–467. <https://doi.org/10.1111/jcal.12146>
- 1765 John, P. D. (2006). Lesson planning and the student teacher: re-thinking the dominant  
1766 model. *Journal of Curriculum Studies*, 38(4), 483–498.  
1767 <https://doi.org/10.1080/00220270500363620>
- 1768 \*Juhler, M. V. (2016). The Use of Lesson Study Combined with Content Representation  
1769 in the Planning of Physics Lessons During Field Practice to Develop Pedagogical  
1770 Content Knowledge. *Journal of Science Teacher Education*, 27(5), 533–553.  
1771 <https://doi.org/10.1007/s10972-016-9473-4>
- 1772 \*Jung, K. G., & Brown, J. C. (2016). Examining the Effectiveness of an Academic  
1773 Language Planning Organizer as a Tool for Planning Science Academic Language  
1774 Instruction and Supports. *Journal of Science Teacher Education*, 27(8), 847–872.  
1775 <https://doi.org/10.1007/s10972-016-9491-2>
- 1776 \*Kademian, S. M., & Davis, E. A. (2018). Supporting beginning teacher planning of  
1777 investigation-based science discussions. *Journal of Science Teacher Education*,  
1778 29(8), 712–740. <https://doi.org/10.1080/1046560X.2018.1504266>

- 1779 Kafai, Y. B., Xin, Y., Fields, D., & Tofel-Grehl, C. (2022). Teaching and learning about  
1780 respiratory infectious diseases: A scoping review of interventions in K-12  
1781 education. *Journal of Research in Science Teaching*, 59(7), 1274–1300.  
1782 <https://doi.org/10.1002/tea.21797>
- 1783 \*Kang, H. (2017). Preservice teachers' learning to plan intellectually challenging tasks.  
1784 *Journal of Teacher Education*, 68(1), 55–68.  
1785 <https://doi.org/10.1177/0022487116676313>
- 1786 \*Kapon, S., & Merzel, A. (2019). Pre-service physics teachers' learning from guided  
1787 cycles of pedagogical design and teaching of lessons. *Physics Education*, 54(1),  
1788 15026. <https://doi.org/10.1088/1361-6552/aaebbe>
- 1789 \*Käpylä, M., Heikkinen, J.-P., & Asunta, T. (2009). Influence of Content Knowledge on  
1790 Pedagogical Content Knowledge: The case of teaching photosynthesis and plant  
1791 growth. *International Journal of Science Education*, 31(10), 1395–1415.  
1792 <https://doi.org/10.1080/09500690802082168>
- 1793 \*Karlström, M., & Hamza, K. (2021). How do we teach planning to pre-service teachers  
1794 – A tentative model. *Journal of Science Teacher Education*, 1–22.  
1795 <https://doi.org/10.1080/1046560X.2021.1875163>
- 1796 Kaya, Z., & Nafiz Kaya, O. (2023). Gathering Rich Data on Preservice Science Teachers'  
1797 Pedagogical Content Knowledge Through Their Lesson Plans. *Journal of Teacher*  
1798 *Education*, 74(1), 10–22. <https://doi.org/10.1177/00224871221105801>
- 1799 Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American*  
1800 *Documentation*, 14(1), 10–25. <https://doi.org/10.1002/asi.5090140103>

- 1801 Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and  
1802 potential for progress. *Studies in Science Education*, 45(2), 169–204.  
1803 <https://doi.org/10.1080/03057260903142285>
- 1804 Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert,  
1805 J. (2013). Teachers' content knowledge and pedagogical content knowledge: The  
1806 role of structural differences in teacher education. *Journal of Teacher Education*,  
1807 64(1), 90–106. <https://doi.org/10.1177/0022487112460398>
- 1808 Koberstein-Schwarz, M., & Meisert, A. (2022). Pedagogical content knowledge in  
1809 material-based lesson planning of preservice biology teachers. *Teaching and*  
1810 *Teacher Education*, 116, 103745. <https://doi.org/10.1016/j.tate.2022.103745>
- 1811 König, J., Blömeke, S., Jentsch, A., Schlesinger, L., née Nehls, C. F., Musekamp, F., &  
1812 Kaiser, G. (2021). The links between pedagogical competence, instructional  
1813 quality, and mathematics achievement in the lower secondary classroom.  
1814 *Educational Studies in Mathematics*, 107(1), 189–212.  
1815 <https://doi.org/10.1007/s10649-020-10021-0>
- 1816 König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I., & Glutsch, N. (2020a).  
1817 Planning Competence of Pre-Service German Language Teachers (PlanvoLL-D):  
1818 Conceptualization, Measurement, and Validation. In O. Zlatkin-Troitschanskaia, H.  
1819 A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student Learning in German Higher*  
1820 *Education: Innovative Measurement Approaches and Research Results* (pp. 53–74).  
1821 Springer. [https://doi.org/10.1007/978-3-658-27886-1\\_4](https://doi.org/10.1007/978-3-658-27886-1_4)
- 1822 König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I., & Glutsch, N. (2020b). Pre-  
1823 service teachers' generic and subject-specific lesson-planning skills: On learning

- 1824 adaptive teaching during initial teacher education. *European Journal of Teacher*  
1825 *Education*, 43(2), 131–150. <https://doi.org/10.1080/02619768.2019.1679115>
- 1826 König, J., Bremerich-Vos, A., Buchholtz, C., & Glutsch, N. (2020). General pedagogical  
1827 knowledge, pedagogical adaptivity in written lesson plans, and instructional  
1828 practice among preservice teachers. *Journal of Curriculum Studies*, 52(6), 800–822.  
1829 <https://doi.org/10.1080/00220272.2020.1752804>
- 1830 König, J., Buchholtz, C., & Dohmen, D. (2015). Analyse von schriftlichen  
1831 Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt  
1832 der Planungskompetenz angehender Lehrkräfte. *Zeitschrift Für*  
1833 *Erziehungswissenschaft*, 18(2), 375–404. [https://doi.org/10.1007/s11618-015-](https://doi.org/10.1007/s11618-015-0625-7)  
1834 [0625-7](https://doi.org/10.1007/s11618-015-0625-7)
- 1835 König, J., Krepf, M., Bremerich-Vos, A., & Buchholtz, C. (2021). Meeting cognitive  
1836 demands of lesson planning: introducing the CODE-PLAN Model to describe and  
1837 analyse teachers' planning competence. *The Teacher Educator*, 56(4), 466–487.  
1838 <https://doi.org/10.1080/08878730.2021.1938324>
- 1839 König, J., & Rothland, M. (2022). Stichwort: Unterrichtsplanungskompetenz. *Zeitschrift*  
1840 *Für Erziehungswissenschaft*, 25, 771–813. [https://doi.org/10.1007/s11618-022-](https://doi.org/10.1007/s11618-022-01107-x)  
1841 [01107-x](https://doi.org/10.1007/s11618-022-01107-x)
- 1842 Kotzebue, L. von (2022). Beliefs, self-reported or performance-assessed TPACK: What  
1843 can predict the quality of technology-enhanced biology lesson plans? *Journal of*  
1844 *Science Education and Technology*, 31(5), 570–582.  
1845 <https://doi.org/10.1007/s10956-022-09974-z>

- 1846 Krämer, P., Nessler, S. H., & Schlüter, K. (2015). Teacher students' dilemmas when  
1847 teaching science through inquiry. *Research in Science & Technological Education*,  
1848 33(3), 325–343. <https://doi.org/10.1080/02635143.2015.1047446>
- 1849 Krepf, M., & König, J. (2022a). Structuring lessons as an aspect of pre-service teachers'  
1850 planning competence: A scaling-up analysis. *Zeitschrift Für*  
1851 *Erziehungswissenschaft*, 25(4), 917–946. [https://doi.org/10.1007/s11618-022-](https://doi.org/10.1007/s11618-022-01125-9)  
1852 01125-9
- 1853 Krepf, M., & König, J. (2022b). Structuring the lesson: an empirical investigation of pre-  
1854 service teacher decision-making during the planning of a demonstration lesson.  
1855 *Journal of Education for Teaching*, 1–16.  
1856 <https://doi.org/10.1080/02607476.2022.2151877>
- 1857 Krepf, M., Plöger, W., Scholl, D., & Seifert, A. (2018). Pedagogical content knowledge  
1858 of experts and novices-what knowledge do they activate when analysing science  
1859 lessons? *Journal of Research in Science Teaching*, 55(1), 44–67.  
1860 <https://doi.org/10.1002/tea.21410>
- 1861 Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013).  
1862 Professional competence of teachers: Effects on instructional quality and student  
1863 development. *The Journal of Educational Psychology*, 105(3), 805–820.  
1864 <https://doi.org/10.1037/a0032583>
- 1865 Kutsekoda. (2013). Professional Standard - Senior Teacher Level 7.  
1866 [https://www.kutsekoda.ee/wp-content/uploads/2019/ENG\\_veeb/Teacher-level-](https://www.kutsekoda.ee/wp-content/uploads/2019/ENG_veeb/Teacher-level-7_eng-13.pdf)  
1867 7\_eng-13.pdf
- 1868 Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for  
1869 Categorical Data. *Biometrics*, 33(1), 159–174.

- 1870 \*Larkin, D. (2017). Planning for the Elicitation of Students' Ideas: A Lesson Study  
1871 Approach with Preservice Science Teachers. *Journal of Science Teacher Education*,  
1872 28(5), 425-443 (19 Seiten). <https://doi.org/10.1080/1046560X.2017.1352410>
- 1873 Lotter, C., Rushton, G. T., & Singer, J. (2013). Teacher Enactment Patterns: How Can  
1874 We Help Move All Teachers to Reform-Based Inquiry Practice Through  
1875 Professional Development? *Journal of Science Teacher Education*, 24(8), 1263–  
1876 1291. <https://doi.org/10.1007/s10972-013-9361-0>
- 1877 Loughran, J., Keast, S., & Cooper, R. (2016). Pedagogical Reasoning in Teacher  
1878 Education. In J. Loughran & M. L. Hamilton (Eds.), *International Handbook of*  
1879 *Teacher Education* (pp. 387–421). Springer Singapore.  
1880 [https://doi.org/10.1007/978-981-10-0366-0\\_10](https://doi.org/10.1007/978-981-10-0366-0_10)
- 1881 \*Luera, G. R., Moyer, R. H., & Everett, S. A. (2005). What type and level of science  
1882 content knowledge of elementary education students affect their ability to construct  
1883 an inquiry-based science lesson? *Journal of Elementary Science Education*, 17(1),  
1884 12–25. <https://doi.org/10.1007/BF03174670>
- 1885 Lui, A. M., & Bonner, S. M. (2016). Preservice and inservice teachers' knowledge,  
1886 beliefs, and instructional planning in primary school mathematics. *Teaching and*  
1887 *Teacher Education*, 56, 1–13. <https://doi.org/10.1016/j.tate.2016.01.015>
- 1888 Ma, T.-J., Lee, G.-G., Liu, J., & Lae, R. (2022). Bibliographic coupling: a main path  
1889 analysis from 1963 to 2020. *Information Research: An International Electronic*  
1890 *Journal*, 27(1). <https://doi.org/10.47989/irpaper918>
- 1891 Magnusson, S. J., Borko, H., & Krajcik, J. (1999). Nature, sources, and development of  
1892 pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G.

- 1893 Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Kluwer  
1894 Academic Publishers.
- 1895 \*Mangiante, E. S. (2013). Planning Science Instruction for Critical Thinking: Two Urban  
1896 Elementary Teachers' Responses to a State Science Assessment. *Education  
1897 Sciences*, 3(3), 222–258.
- 1898 Mavhunga, E. (2020). Revealing the structural complexity of component interactions of  
1899 topic-specific PCK when planning to teach. *Research in Science Education*, 50(3),  
1900 965–986. <https://doi.org/10.1007/s11165-018-9719-6>
- 1901 Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation* (2.  
1902 ed.). Jossey-Bass.
- 1903 \*Mesci, G., Çavus-Güngören, S., & Yesildag-Hasancebi, F. (2020). Investigating the  
1904 Development of Pre-Service Science Teachers' NOSI Views and Related Teaching  
1905 Practices. *International Journal of Science Education*, 42(1), 50-69.  
1906 <https://doi.org/10.1080/09500693.2019.1700316>
- 1907 Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded  
1908 sourcebook* (2. ed [Nachdr.]). Sage.
- 1909 Ministère de l'Éducation Nationale. (2013). Arrêté du 1er juillet 2013 relatif au référentiel  
1910 des compétences professionnelles des métiers du professorat et de l'éducation.  
1911 <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027721614/>
- 1912 \*Minken, Z., Macalalag, J. A., Clarke, A., Marco-Bujosa, L., & Rulli, C. (2021).  
1913 Development of Teachers' Pedagogical Content Knowledge during Lesson  
1914 Planning of Socioscientific Issues. *International Journal of Technology in  
1915 Education*, 4(2), 113–165. <https://doi.org/10.46328/ijte.50>



- 1916 Morine-Dershimer, G. (1979). Teacher plan and classroom reality: The South Bay Study,  
 1917 part IV. Research series No. 60. Institute for Research on Teaching College of  
 1918 Education Michigan State University.
- 1919 Morse, J. M. (2022). The Fallacy of Rigor: Examining Checklist Criteria as an Indicator  
 1920 of Quality. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Research Design*  
 1921 (pp. 1–37). Sage Publications. <https://doi.org/10.4135/9781529770278>
- 1922 Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E.  
 1923 (2018). Systematic review or scoping review? Guidance for authors when choosing  
 1924 between a systematic or scoping review approach. *BMC Medical Research*  
 1925 *Methodology*, 18(1), 143. <https://doi.org/10.1186/s12874-018-0611-x>
- 1926 Munthe, E., & Conway, P. F. (2017). Evolution of research on teachers' planning:  
 1927 implications for teacher education. In J. Hunsu & D. J. Clandinin (Eds.), *Evolution*  
 1928 *of research on teachers' planning: Implications for teacher education* (pp. 836–849).  
 1929 SAGE Publications Ltd.
- 1930 Ndiokubwayo, K., Byukusenge, C., Byusa, E., Habiyaremye, H. T., Mbonyirivuze, A.,  
 1931 & Mukagihana, J. (2022). Lesson plan analysis protocol (LPAP): A useful tool for  
 1932 researchers and educational evaluators. *Heliyon*, 8(1), 1-7.  
 1933 <https://doi.org/10.1016/j.heliyon.2022.e08730>
- 1934 Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation  
 1935 in Germany. In J. E. Pedersen, T. Isozaki, & T. Hirano (Eds.), *Model science*  
 1936 *teacher preparation programs: An international comparison of what works* (pp. 29–  
 1937 52). Information Age Publishing, Inc.
- 1938 Newman, W. J., Abell, S. K., Hubbard, P. D., McDonald, J., Otaala, J., & Martini, M.  
 1939 (2004). Dilemmas of Teaching Inquiry in Elementary Science Methods. *Journal of*

- 1940 Science Teacher Education, 15(4), 257–279.
- 1941 <https://doi.org/10.1023/B:JSTE.0000048330.07586.d6>
- 1942 Nilsson, P., & Karlsson, G. (2019). Capturing student teachers' pedagogical content  
 1943 knowledge (PCK) using CoRes and digital technology. *International Journal of*  
 1944 *Science Education*, 41(4), 419–447.  
 1945 <https://doi.org/10.1080/09500693.2018.1551642>
- 1946 \*Nixon, R. S., & Campbell, B. K. (2021). Attending to science concepts during planning:  
 1947 Exploring and supporting the process. *School Science and Mathematics*, 121(6),  
 1948 333–344. <https://doi.org/10.1111/ssm.12486>
- 1949 \*Ogegbo, A. A., Gaigher, E., & Salagaram, T. (2019). Benefits and Challenges of Lesson  
 1950 Study: A Case of Teaching Physical Sciences in South Africa. *South African*  
 1951 *Journal of Education*, 39(1), iten).
- 1952 Olson, J. K., Tippet, C. D., Milford, T. M., Ohana, C., & Clough, M. P. (2015). Science  
 1953 Teacher Preparation in a North American Context. *Journal of Science Teacher*  
 1954 *Education : The Official Journal of the Association for the Education of Teachers*  
 1955 *in Science*, 26(1), 7–28. <https://doi.org/10.1007/s10972-014-9417-9>
- 1956 Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2. ed. [Nachdr.]).  
 1957 Sage.
- 1958 Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3. ed.). Sage.  
 1959 <http://www.loc.gov/catdir/enhancements/fy0734/2001005181-b.html>
- 1960 Pedersen, J. E., Isozaki, T., & Hirano, T. (Eds.). (2017). *Model science teacher*  
 1961 *preparation programs: An international comparison of what works*. Information  
 1962 Age Publishing, Inc.

- 1963 Penso, S., & Shoham, E. (2003). Student teachers' reasoning while making pedagogical  
1964 decisions. *European Journal of Teacher Education*, 26(3), 313–328.  
1965 <https://doi.org/10.1080/0261976032000128166>
- 1966 \*Penuel, W. R., McWilliams, H., McAuliffe, C., Benbow, A. E., Mably, C., & Hayden,  
1967 M. M. (2009). Teaching for Understanding in Earth Science: Comparing Impacts  
1968 on Planning and Instruction in Three Professional Development Designs for Middle  
1969 School Science Teachers. *Journal of Science Teacher Education*, 20(5), 415–436.  
1970 <https://doi.org/10.1007/s10972-008-9120-9>
- 1971 Pepin, B., Gueudet, G., & Trouche, L. (2013). Re-sourcing teachers' work and  
1972 interactions: a collective perspective on resources, their use and transformation.  
1973 *ZDM*, 45(7), 929–943. <https://doi.org/10.1007/s11858-013-0534-2>
- 1974 Peters, M. D. J., Godfrey, C. M., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H.  
1975 (2020). Scoping reviews. In E. Aromataris & Z. Munn (Eds.), *JBIM Manual for*  
1976 *Evidence Synthesis* (pp. 406–450). JBI. <https://doi.org/10.46658/JBIMES-20-12>
- 1977 Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S.  
1978 A. (2014). A scoping review of scoping reviews: Advancing the approach and  
1979 enhancing the consistency. *Research Synthesis Methods*, 5(4), 371–385.  
1980 <https://doi.org/10.1002/jrsm.1123>
- 1981 \*Qablan, A. (2016). Teaching and Learning about Science Practices: Insights and  
1982 Challenges in Professional Development. *Teacher Development*, 20(1), 76-91 (16  
1983 Seiten). <https://doi.org/10.1080/13664530.2015.1111929>
- 1984 Reynolds, W. M., & Park, S. (2021). Examining the relationship between the educative  
1985 teacher performance assessment and preservice teachers' pedagogical content

- 1986 knowledge. *Journal of Research in Science Teaching*, 58, 721–748.
- 1987 <https://doi.org/10.1002/tea.1676>
- 1988 Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula
- 1989 (Ed.), *Handbook of research on teacher education: A project of the Association of*
- 1990 *Teacher Educators* (2nd ed., pp. 102–119). Macmillan.
- 1991 \*Rinke, C. R., Gladstone-Brown, W., Kinlaw, C. R., & Cappiello, J. (2016).
- 1992 *Characterizing STEM Teacher Education: Affordances and Constraints of Explicit*
- 1993 *STEM Preparation for Elementary Teachers. School Science and Mathematics*,
- 1994 116(6), 300-309 (10 Seiten). <https://doi.org/10.1111/ssm.12185>
- 1995 Roskos, K. (1996). When Two Heads are Better than One: Beginning Teachers' Planning
- 1996 Processes in an Integrated Instruction Planning Task. *Journal of Teacher Education*,
- 1997 47(2), 120–129. <https://doi.org/10.1177/0022487196047002005>
- 1998 \*Ross, D. K., & Cartier, J. L. (2015). Developing Pre-service Elementary Teachers'
- 1999 Pedagogical Practices While Planning Using the Learning Cycle. *Journal of*
- 2000 *Science Teacher Education*, 26(6), 573–591. [https://doi.org/10.1007/s10972-015-](https://doi.org/10.1007/s10972-015-9439-y)
- 2001 9439-y
- 2002 Rothland, M. (2021). Anmerkungen zur Modellierung und Operationalisierung
- 2003 (allgemeindidaktischer) Unterrichtsplanungskompetenz. *Unterrichtswissenschaft*,
- 2004 50, 347-372. <https://doi.org/10.1007/s42010-021-00111-0>
- 2005 Rusznyak, L., & Walton, E. (2011). Lesson planning guidelines for student teachers: A
- 2006 scaffold for the development of pedagogical content knowledge. *Education as*
- 2007 *Change*, 15(2), 271–285. <https://doi.org/10.1080/16823206.2011.619141>
- 2008 \*Sariieddine, D., & BouJaoude, S. (2014). Influence of Teachers' Conceptions of the
- 2009 Nature of Science on Classroom Practice. *EURASIA Journal of Mathematics*,

- 2010 Science and Technology Education, 10(2).
- 2011 <https://doi.org/10.12973/eurasia.2014.1024a>
- 2012 \*Schneider, R. M. (2013). Opportunities for Teacher Learning During Enactment of
- 2013 Inquiry Science Curriculum Materials: Exploring the Potential for Teacher
- 2014 Educative Materials. *Journal of Science Teacher Education*, 24(2), 323–346.
- 2015 <https://doi.org/10.1007/s10972-012-9309-9>
- 2016 Scholl, D. (2018). Metatheorie der allgemeinen Didaktik: Ein systemtheoretisch
- 2017 begründeter Vorschlag. *Klinkhardt Forschung*. Julius Klinkhardt.
- 2018 <https://elibrary.utb.de/doi/book/10.35468/9783781556539>
- 2019 Scholl, D., Küth, S., Flath, M., Lathan, H., Schwarz, B., Wolters, P., Rheinländer, K., &
- 2020 Schüle, C. (2019). Zum Konstrukt der Planungskompetenz in allgemein- und
- 2021 fachdidaktischen Ansätzen. In D. Scholl, S. Wernke, & D. Behrens (Eds.), *Jahrbuch*
- 2022 *für Allgemeine Didaktik* (pp. 75–94). Schneider Verlag Hohengehren.
- 2023 Scholl, D., Küth, S., & Schüle, C. (2022). Interdependentes Entscheiden in der
- 2024 Unterrichtsplanung – Entwicklung eines generischen Rahmenmodells und eines
- 2025 vignettenbasierten Fähigkeitstests. *Zeitschrift Für Erziehungswissenschaft*, 25(4),
- 2026 895–916. <https://doi.org/10.1007/s11618-022-01117-9>
- 2027 Schröder, J., Riese, J., Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin,
- 2028 M., Kulgemeyer, C., Reinhold, P., & Schecker, H. (2020). Die Messung der
- 2029 Fähigkeit zur Unterrichtsplanung im Fach Physik mit Hilfe eines standardisierten
- 2030 Performanztests. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 26(1), 103–
- 2031 122. <https://doi.org/10.1007/s40573-020-00115-w>

- 2032 Shavelson, R. J., Cadwell, J., & Izu, T. (1977). Teachers' Sensitivity to the Reliability of  
2033 Information in Making Pedagogical Decisions. *American Educational Research*  
2034 *Journal*, 14(2), 83–97. <https://doi.org/10.3102/00028312014002083>
- 2035 Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (2005). Windows into the mind.  
2036 *Higher Education*, 49(4), 413–430. <https://doi.org/10.1007/s10734-004-9448-9>
- 2037 Shulman, L. S. (1986). Those who understand: knowledge growth in teaching.  
2038 *Educational Researcher*, 15(2), 4–14.
- 2039 Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard*  
2040 *Educational Review*, 57(1), 1–22.
- 2041 \*So, W., & Watkins, D. A. (2005). From Beginning Teacher Education to Professional  
2042 Teaching: A Study of the Thinking of Hong Kong Primary Science Teachers.  
2043 *Teaching and Teacher Education: An International Journal of Research and Studies*,  
2044 21(5), 525-541 (17 Seiten). <https://doi.org/10.1016/j.tate.2005.03.003>
- 2045 \*So, W. M. (1997). A study of teacher cognition in planning elementary science lessons.  
2046 *Research in Science Education*, 27(1), 71–86. <https://doi.org/10.1007/BF02463033>
- 2047 Sorge, S., Kröger, J., Petersen, S., & Neumann, K. (2019). Structure and Development of  
2048 Pre-Service Physics Teachers' Professional Knowledge. *International*  
2049 *Journal of Science Education*, 41(7), 862-889 (28 Seiten).  
2050 <https://doi.org/10.1080/09500693.2017.1346326>
- 2051 Sorge, S., Stender, A., & Neumann, K. (2019). The Development of Science Teachers'  
2052 Professional Competence. In A. Hume, R. Cooper, & A. Borowski (Eds.),  
2053 *Repositioning pedagogical content knowledge in teachers' knowledge for teaching*  
2054 *science* (pp. 151–166). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_6](https://doi.org/10.1007/978-981-13-5898-2_6)

- 2055 Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.  
 2056 (2019). Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der  
 2057 Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019.  
 2058 [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf)  
 2059 [004\\_12\\_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf)
- 2060 Starkey, L. (2010). Teachers' pedagogical reasoning and action in the digital age.  
 2061 *Teachers and Teaching*, 16(2), 233–244.  
 2062 <https://doi.org/10.1080/13540600903478433>
- 2063 \*Stender, A., Brückmann, M., & Neumann, K. (2017). Transformation of Topic-Specific  
 2064 Professional Knowledge into Personal Pedagogical Content Knowledge through  
 2065 Lesson Planning. *International Journal of Science Education*, 39(12), 1690-1714.  
 2066 <https://doi.org/10.1080/09500693.2017.1351645>
- 2067 Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H. L., Levac, D., Moher,  
 2068 D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C.,  
 2069 McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., .  
 2070 . . Straus, S. E. (2018). Prisma Extension for Scoping Reviews (PRISMA-ScR):  
 2071 Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467–473.  
 2072 <https://doi.org/10.7326/M18-0850>
- 2073 Tyler, R. W. (1950). *Basic principles of curriculum development*. University of Chicago.
- 2074 van der Valk, T. A. E., & Broekman, H. H. G. B. (1999). The lesson preparation method:  
 2075 a Way of Investigating Pre-Service Teachers' Pedagogical Content Knowledge.  
 2076 *European Journal of Teacher Education*, 22(1), 11–22.

- 2077 van Eck, N. J., & Waltman, L. (2010). Software survey: Vosviewer, a computer program  
 2078 for bibliometric mapping. *Scientometrics*, 84(2), 523–538.  
 2079 <https://doi.org/10.1007/s11192-009-0146-3>
- 2080 van Eck, N. J., & Waltman, L. (2014). Visualizing Bibliometric Networks. In Y. Ding,  
 2081 R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact* (pp. 285–320).  
 2082 Springer International Publishing. [https://doi.org/10.1007/978-3-319-10377-8\\_13](https://doi.org/10.1007/978-3-319-10377-8_13)
- 2083 \*Verma, G. (2008). Using Scholarship of Teaching and Learning (SoTL) to Inquire into  
 2084 Pre-Service Teachers' Science Lesson Planning Considerations. *International*  
 2085 *Journal for the Scholarship of Teaching and Learning*, 2(2), iten).
- 2086 Vogelsang, C., Kulgemeyer, C., & Riese, J. (2022). Learning to plan by learning to  
 2087 reflect? - Exploring relations between professional knowledge, reflection skills, and  
 2088 Planning skills of preservice physics teachers in a one-semester field experience.  
 2089 *Education Sciences*, 12(7), 479. <https://doi.org/10.3390/educsci12070479>
- 2090 Vogelsang, C., & Riese, J. (2017). Wann ist eine Unterrichtsplanung ‚gut‘? -  
 2091 Planungsperformanz in Praxisratgebern zur Unterrichtsplanung. In S. Wernke & K.  
 2092 Zierer (Eds.), *Die Unterrichtsplanung: ein in Vergessenheit geratener*  
 2093 *Kompetenzbereich?!*: Status Quo und Perspektiven aus Sicht der empirischen  
 2094 *Forschung* (pp. 47–61). Julius Klinkhardt.
- 2095 Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general  
 2096 pedagogical/psychological knowledge: Test construction and validation. *Journal of*  
 2097 *Educational Psychology*, 103(4), 952–969. <https://doi.org/10.1037/a0025125>
- 2098 \*Wahbeh, N., & Abd-El-Khalick, F. (2014). Revisiting the Translation of Nature of  
 2099 Science Understandings into Instructional Practice: Teachers' Nature of Science



2100 Pedagogical Content Knowledge. *International Journal of Science Education*,  
2101 36(3), 425-466 (42 Seiten). <https://doi.org/10.1080/09500693.2013.786852>

2102 \*Weitzel, H., & Blank, R. (2020). Pedagogical Content Knowledge in peer dialogues  
2103 between pre-service biology teachers in the planning of science lessons. Results of  
2104 an intervention Study. *Journal of Science Teacher Education*, 31(1), 75–93.  
2105 <https://doi.org/10.1080/1046560X.2019.1664874>

2106 Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher*  
2107 *Education*, 42(4), 292–305.

2108 \*Wyner, Y. (2013a). A Conceptual Model for Teaching the Relationship of Daily Life  
2109 and Human Environmental Impact to Ecological Function. *International Journal of*  
2110 *Environmental and Science Education*, 8(4), 561-586 (26 Seiten).

2111 \*Wyner, Y. (2013b). The Impact of a Novel Curriculum on Secondary Biology Teachers’  
2112 Dispositions Toward Using Authentic Data and Media in Their Human Impact and  
2113 Ecology Lessons. *Journal of Science Teacher Education*, 24(5), 833–857.  
2114 <https://doi.org/10.1007/s10972-013-9335-2>

2115 \*Yildiz, E. (2021). Developing Knowledge and Skills of Prospective Science Teachers  
2116 on Teaching Strategies, Models, Methods and Techniques: An Action Research.  
2117 *International Journal of Curriculum and Instruction*, 13(1), 794-815 (22 Seiten).

2118 Yinger, R. J. (1980). A Study of teacher planning. *Elementary School Journal*, 80(3),  
2119 107–127.

2120 Yoon, H.-G., Joung, Y. J., & Kim, M. (2012). The Challenges of Science Inquiry  
2121 Teaching for Pre-Service Teachers in Elementary Classrooms: Difficulties on and  
2122 under the Scene. *Research in Science Education*, 42(3), 589–608.  
2123 <https://doi.org/10.1007/s11165-011-9212-y>

- 2124 Zahorik, J. A. (1975). Teachers' planning models. *Educational Leadership*, 33, 134–139.
- 2125 Zaragoza, A., Seidel, T., & Hiebert, J. (2021). Exploring preservice teachers' abilities to  
2126 connect professional knowledge with lesson planning and observation. *European*  
2127 *Journal of Teacher Education*, 1–20.  
2128 <https://doi.org/10.1080/02619768.2021.1996558>
- 2129 Zaragoza, A., Seidel, T., & Santagata, R. (2023). Lesson analysis and plan template:  
2130 scaffolding preservice teachers' application of professional knowledge to lesson  
2131 planning. *Journal of Curriculum Studies*, 1–15.  
2132 <https://doi.org/10.1080/00220272.2023.2182650>
- 2133 Zeichner, K. M., & Tabachnick, B. R. (1981). Are the Effects of University Teacher  
2134 Education 'Washed Out' by School Experience? *Journal of Teacher Education*,  
2135 32(3), 7–11. <https://doi.org/10.1177/002248718103200302>
- 2136 \*Zimmermann, F., Melle, I., & Huwer, J. (2021). Developing Prospective Chemistry  
2137 Teachers' TPACK—A Comparison between Students of Two Different Universities  
2138 and Expertise Levels Regarding Their TPACK Self-Efficacy, Attitude, and Lesson  
2139 Planning Competence. *Journal of Chemical Education*, 98(6), 1863–1874.  
2140 <https://doi.org/10.1021/acs.jchemed.0c01296>

## Beitrag 2

**Großmann, L.** & Krüger, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen, *Erkenntnisweg Biologiedidaktik*, 19, 21-39.

# Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen

Leroy Großmann & Dirk Krüger

leroy.grossmann@fu-berlin.de – dirk.krueger@fu-berlin.de

Freie Universität Berlin: Didaktik der Biologie

Schwendenerstraße 1, 14195 Berlin

---

## Zusammenfassung

*In der empirischen naturwissenschaftsdidaktischen Forschung lassen sich zunehmend methodische Ansätze beobachten, mit denen das fachdidaktische Wissen als konstitutiver Teil der professionellen Kompetenzen von Lehrkräften beispielsweise zur Untersuchung der Denkprozesse im Unterrichtshandeln erfasst wird. Bislang wenig Beachtung findet dabei die Planungspraxis von Biologie-Lehrkräften. Im Rahmen dieser Arbeit wird die Entwicklung eines Kategoriensystems dargestellt, mit dem Staatsexamensentwürfe von Referendar\*innen qualitativ inhaltsanalytisch untersucht werden, was exemplarisch an einer Fallanalyse mit dem entwickelten Kategoriensystem gezeigt wird. Das Potential dieses Vorgehens für die biologie-didaktische Forschung und Implikationen für die Lehrkräftebildung im Fach Biologie werden diskutiert.*

## Abstract

*Pedagogical content knowledge has proven constitutive for teachers' expertise and is regarded as crucial in the process of planning, teaching and reflecting lessons. There is an increasing number of methodical approaches in empirical research on science teacher education to capture science teachers' pedagogical content knowledge. However, there is still a lack of findings on what cognitive resources are accessed by the teachers, particularly when planning biology lessons. Therefore, a category system has been developed and tested for a qualitative content analysis of biology trainee teachers' written lesson plans. In this article, we describe the development of the category system and demonstrate its application in one purposefully selected case. Finally, we discuss the potential of this methodical approach in the context of biology teacher education.*

## 1 Einleitung

Die Planung von Unterricht gilt als Kernaufgabe von Lehrkräften (KMK, 2005) und damit als konstitutiver Bestandteil ihrer professionellen Kompetenzen (BAUMERT & KUNTER, 2006). Planungshandeln kann als Grundlage für eine funktionale Durchführung und eine kritische Reflexion von Unterricht betrachtet werden (WAHL, 2002), da es für die Entwicklung professioneller Kompetenzen im Allgemeinen und fachdidaktischen Wissens im Besonderen eine wesentliche Rolle spielt (BAYLOR, 2002) und somit als Schlüsselstelle unterrichtlichen Handelns gilt (ZIERER, WERNER & WERNKE, 2015). Die Komplexität von Lehr-Lern-Prozessen, mögliche Schwierigkeiten in der praktischen Umsetzung und Handlungsalternativen werden dabei antizipiert (LOUGHRAN, KEAST & COOPER, 2016), wofür insbesondere fachdidaktisches Wissen benötigt wird (COE, ALOISI, HIGGINS & MAJOR, 2014).

In den vergangenen zehn Jahren wurden insbesondere in der naturwissenschafts-didaktischen Forschung einige Versuche unternommen (CHAN & HUME, 2019), fachdidaktisches Wissen von Lehrkräften in vielfältigen methodischen Ansätzen quantitativ (z.B. GROßSCHEDL, WELTER & HARMS, 2019; JÜTTNER et al., 2013) und qualitativ (z.B. ALONZO & KIM, 2016; BROVELLI, BÖLSTERLI, REHM & WILHELM, 2013) zu erfassen – nur selten jedoch mit explizitem Fokus auf die Unterrichtsplanung (z.B. KÖNIG, BUCHHOLTZ & DOHMEN, 2017).

Bislang kaum beachtet wurde die Frage, auf welche kognitiven fachdidaktischen Wissensbestände Lehrkräfte bei der Planung von Unterricht zurückgreifen und wie sich diese in schriftlichen Unterrichtsplanungen manifestieren. Zwar liegen für das Fach Biologie inzwischen kompetenzbereichsspezifische Studien vor, die das fachdidaktische Wissen von Biologielehrkräften mittels leitfadengestützter Interviews im Bereich der Bewertungskompetenz (POHLMANN, 2019) oder bei der Planung von Experimentierstunden in einem standardisierten Setting im Bereich der Erkenntnisgewinnung (TARDENT KUSTER, 2020) untersuchen, allerdings fehlt es an einer allgemeineren „kritischen Bestandsaufnahme der gegenwärtigen Planungspraxen von Lehrkräften“ (WEINGARTEN, 2019, S.28), aus der sich Hinweise ableiten ließen, welche Denkprozesse bei der Planung guten Biologieunterrichts besonders berücksichtigt werden müssten.

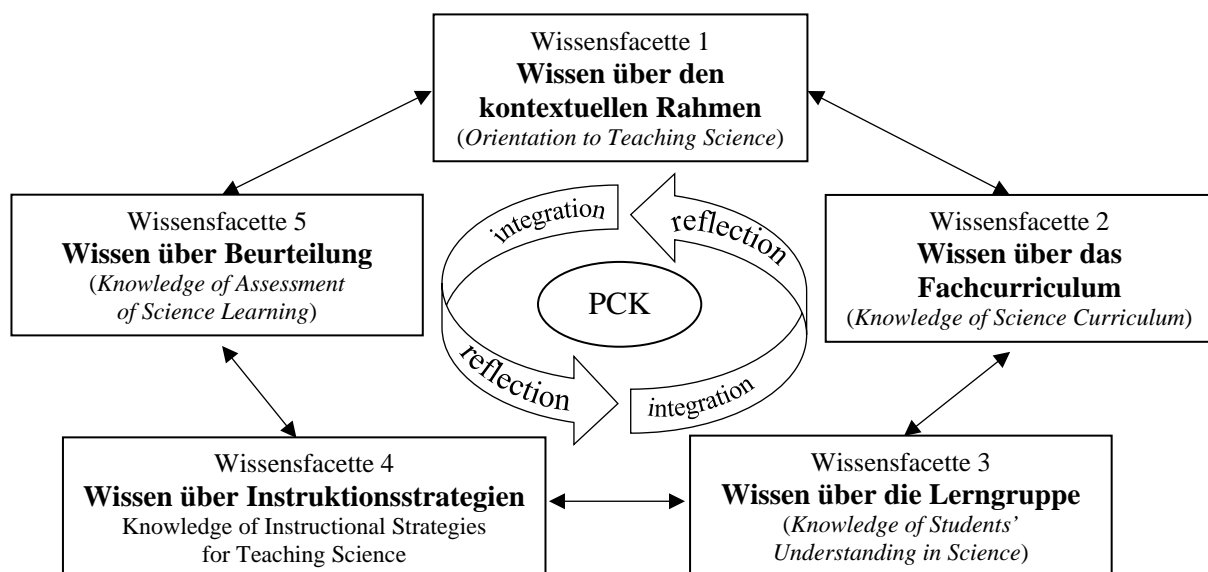
Um die bei der Planung von Biologieunterricht aktivierten fachdidaktischen Wissensbestände in ihrer Wechselwirkung untereinander untersuchen zu können, wurde auf der Grundlage einer Modellierung fachdidaktischen Wissens (PARK & OLIVER, 2008) ein Kategoriensystem entwickelt und auf die Unterrichtsplanungen von Referendar\*innen im Rahmen der Prüfung zum Zweiten Staatsexamen angewendet.

Dieser Artikel stellt den Entwicklungsprozess des Kategoriensystems dar und erläutert anhand eines exemplarischen Falls, inwiefern das Instrument zur Analyse fachdidaktischen Wissens geeignet ist und welche Implikationen sich aus der gegenwärtigen Planungspraxis von angehenden Biologie-Lehrkräften ergeben.

## 2 Theorie

### 2.1 Fachdidaktisches Wissen von Lehrkräften

Neben pädagogischem Wissen (*pedagogical knowledge*, PK) und Fachwissen (*content knowledge*, CK) stellt fachdidaktisches Wissen (*pedagogical content knowledge*, PCK) die wichtigste Wissenskomponente der Expertise von Lehrkräften dar (BAUMERT & KUNTER, 2006; SHULMAN, 1986) und gilt als „Kernstück [ihres] hochgradig unterrichtsrelevanten Professionswissens“ (NEUWEG, 2011). PCK umfasst dabei das Wissen und die Fähigkeiten, die benötigt werden, um ein „*particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes*“ (GESS-NEWSOME, 2015, S.36) zu unterrichten. Forschungsvorhaben zum PCK gehen der Frage nach, wie dieses Wissen entwickelt werden kann und wie sich dies auf den Lernerfolg von SuS auswirkt (KIND & CHAN, 2019). Eine Konzeptualisierung des fachdidaktischen Wissens von Lehrkräften naturwissenschaftlicher Fächer, die aufgrund ihrer fünf diskreten Wissensfacetten eine gute Operationalisierbarkeit bietet (POHLMANN, 2019), schlagen PARK und OLIVER (2008) vor (Abb. 1):



**Abb. 1:** Pentagon-Modell des fachdidaktischen Wissens (PCK) von Lehrkräften naturwissenschaftlicher Fächer (nach: PARK & OLIVER, 2008).

CARLSON und DAEHLER (2019) unterscheiden dabei zwischen persönlichem PCK (*personal PCK*; pPCK) als „*reservoir of knowledge and skills that the teacher can draw upon during the practice of teaching*“ (Ebd., S.85) und dem diesem untergeordneten, situationsspezifischen und sich im konkreten Handeln manifestierenden PCK (*enacted PCK*; ePCK). Ausgeprägtes fachdidaktisches Wissen in der Planung (ePCK<sub>p</sub>; ALONZO, BERRY & NILSSON, 2019) von Unterricht besteht demnach darin, Planungsentscheidungen unter Berücksichtigung aller dieser Facetten zu treffen.

Dies steht in unmittelbarem Zusammenhang mit der als *pedagogical reasoning* (LOUGHRAN, KEAST & COOPER, 2016) bezeichneten Fähigkeit, diese Entscheidungen zu begründen, Lernschwierigkeiten zu antizipieren und Alternativen einzuplanen. Sie wird als konstitutiver Teil des PCK aufgefasst (CHAN et al., 2019) und ist elementar für die Qualität der Unterrichtsplanung von Lehrkräften (VOGELSANG & RIESE, 2017). Da viele Studien im Design und der Interpretation nicht unterscheiden, welche Art von PCK gemessen wird, dies aber wichtig ist (CHAN & HUME, 2019), gilt für die folgenden Ausführungen: Untersucht wird das sich im konkreten Planungshandeln manifestierende ePCK<sub>p</sub> als Teil des fachdidaktischen Wissens einer Lehrkraft, d.h. diejenigen Wissenskomponenten, die in der Planung von Biologieunterricht aktiviert werden und folglich in schriftlichen Unterrichtsentwürfen nachweisbar sind.

## 2.2 Unterrichtsplanung

Anders als beim fachdidaktischen Wissen fehlt beim Konstrukt Planungskompetenz bislang eine Operationalisierung, so dass kein entsprechendes Kompetenzmodell als Referenzrahmen für empirische Untersuchungen (WERNKE & ZIERER, 2017) herangezogen werden kann. Aufgrund der heterogenen Begriffsdefinitionen und der diversen methodischen Zugänge gestaltet sich die Verallgemeinerbarkeit von Befunden schwierig und steht einer konsistenten Beschreibung des Planungshandelns von Lehrkräften im Wege (WEINGARTEN, 2019, S.23). Gleichwohl besteht weitgehend Konsens hinsichtlich der Bedeutung der Unterrichtsplanung für erfolgreiche Lehr-Lern-Prozesse (HATTIE & YATES, 2015). Sie gilt als mentales Probehandeln unter Einbeziehung von kognitiven Dispositionen (BROMME & SEEGER, 1979), in dem planungsrelevante Aspekte (Thema, Lerngruppe, Lehrplan usw.) zueinander in Beziehung gesetzt werden und der Lernprozess antizipiert wird.

Daher argumentiert WEINGARTEN (2019), dass die Qualität der Planung von Unterricht von den kognitiven Dispositionen der Lehrkraft, d.h. dem Vernetzungsgrad der Wissensbestände, abhängt. Bisherige Studien zeigen, dass, während

Experten auf mentale Schemata zurückgreifen können und daher vor allem Hürden antizipieren und andere triviale Aspekte nicht berücksichtigen, Novizen ein viel höheres Auflösungs-niveau zeigen, d.h. Wichtiges seltener von Unwichtigem trennen und folglich viele Entscheidungen penibel begründen, die für den Lernprozess von nachrangiger Bedeutung sind (LIVINGSTON & BORKO, 1989). Novizen greifen bei der Planung auf weniger Wissensbestände zurück, die untereinander zudem weniger stark vernetzt sind als bei Experten (WESTERMAN, 1991). Durch die Integration der verschiedenen Wissensbestände (vgl. NEUMANN, KIND & HARMS, 2019) werden diese im Planungshandeln somit zum ePCK<sub>p</sub> verknüpft (GASSMANN, 2013).

### 3 Fragestellungen

Das Ziel dieser Studie ist es, das für die Unterrichtsplanung von Biologie-Lehrkräften relevante fachdidaktische Wissen qualitativ zu erfassen und daraus Qualitätskriterien für gelingendes Planungshandeln abzuleiten. Die vorliegende Arbeit fokussiert die Entwicklung und Anwendung eines Kategoriensystems und adressiert die beiden folgenden Fragestellungen:

- F1: Wie lässt sich auf der Grundlage des Pentagon-Modells ein Kategoriensystem entwickeln, das den Vernetzungsgrad des ePCK<sub>p</sub> von Biologie-Lehrkräften erfasst und eine valide Analyse der vorliegenden Staatsexamensentwürfe erlaubt?
- F2: Inwiefern ist das entwickelte Kategoriensystem geeignet, um das ePCK<sub>p</sub> von Lehrkräften in Staatsexamensentwürfen zu identifizieren?

### 4 Methodik

Das Pentagon-Modell fachdidaktischen Wissens (Abb. 1) dient als theoretischer Rahmen für die Untersuchung. Die Analyse orientiert sich am PCK-Map-Approach (PARK & SUH, 2019), bei dem aus dem Datenmaterial mittels qualitativer Inhaltsanalyse das fachdidaktische Wissen der Lehrperson in Form sogenannter PCK-Episoden rekonstruiert und in Form einer pentagonalen PCK-Map visualisiert wird. In dieser wird zum einen die Summe der kodierten Stellen aller fünf Wissensfacetten in den Ecken des Pentagons sowie die Anzahl der Verknüpfungen auf den Linien zwischen den Wissensfacetten dargestellt (Abb. 2). Die Beachtung dieses etablierten Vorgehens (PARK & OLIVER, 2008; POHLMANN, 2019) gewährleistet intersubjektive Nachvollziehbarkeit (MAYRING, 2020).



#### 4.1 Datenmaterial und Stichprobe

Von den von der Senatsverwaltung für Bildung, Jugend und Familie des Landes Berlin zur Verfügung gestellten Staatsexamensentwürfen ( $N=118$ ) wurden sieben nicht berücksichtigt, weil sie Unterricht in der Primarstufe adressierten oder unvollständig waren. Von den übrigen 111 Entwürfen wurde im Rahmen der hier dargestellten Untersuchung gleichverteilt nach Kompetenzbereichen (KMK, 2005) eine zufällige Stichprobe ( $n = 16$ ) gezogen. Die Entwürfe wurden im Rahmen der Staatsprüfung für das Lehramt im Land Berlin (Zweites Staatsexamen) gemäß den Vorgaben (SENBJF 2017) von Referendar\*innen erstellt und sind demnach hinsichtlich des Umfangs und der Grundstruktur vergleichbar (Tab. 1).

**Tab. 1:** Aufbau eines kompetenzorientierten Unterrichtsentwurfs (nach: SENBJF, 2017, S.54).

<b>Bestandteil</b>	<b>Hinweise</b>
Fachlich-inhaltlicher Schwerpunkt	– Sachstruktur – Aufgabenanalyse
Standards des Rahmenlehrplans	– angestrebte Kompetenzentwicklung – Bezug zu Rahmenlehrplan
Längerfristige Kompetenzentwicklung	– Orientierung an spezifischen Lernvoraussetzungen der Lerngruppe – inhaltsbezogener, kumulativ-vernetzter Aufbau der Kompetenzentwicklung in Bezug auf Unterrichtsreihe
Konkretisierung der Standards	– Anwendung der Standards des Rahmenlehrplans auf den zu planenden Unterrichtsausschnitt – Grundlage für Evaluation der Lernerfolgs
Kompetenzentwicklung der SuS	– Diagnose der Heterogenität der SuS – Analyse des Kompetenzstandes z.B. von drei exemplarischen SuS – Bedarf an Binnendifferenzierung
Begründung der Lehr-Lern-Struktur	– didaktische Entscheidungen und methodische Umsetzung – Begründungen für die Entscheidungen
Konkretisierung der Lehr-Lern-Prozesse	– Verlaufsplan inkl. Impulsen, Aufgaben, erwarteten SuS-Leistungen
Anlagen	– Arbeitsblätter, antizipiertes Tafelbild usw.

Aufgrund der vollständigen Anonymisierung durch die Senatsverwaltung konnten keinerlei Kontextvariablen erhoben werden.

#### 4.2 Datenanalyse: Qualitative Inhaltsanalyse

Als methodischer Zugang zur Analyse der Staatsexamensentwürfe wurde die qualitative Inhaltsanalyse (MAYRING, 2015) gewählt. Da in Form des Pentagon-Modells fachdidaktischen Wissens (PARK & OLIVER, 2008; Abb.1) das zu untersuchende Konstrukt bereits theoretisch konzeptualisiert ist, konnten a priori Kategorien entwickelt werden, die in einer Exploration induktiv ergänzt oder modifiziert wurden, so dass die Passung zwischen Untersuchungsinstrument und Datenmaterial erhöht werden konnte (SCHREIER, 2014).

### 4.3 Entwicklung des Kategoriensystems

Die fünf Wissensfacetten des fachdidaktischen Wissens (Abb. 1) dienen bei der Gestaltung des Kategoriensystems als Hauptkategorien, die jeweils unter Einbezug spezifisch biologiedidaktischer (z.B. GROPENGEIER, HARMS & KATTMANN, 2019) und z. T. allgemein-pädagogischer (z.B. HATTIE & TIMPERLEY, 2007) oder lernpsychologischer Literatur (z.B. ANDERSON & KRATWOHL, 2001) ausdifferenziert wurden:

*Wissen über den kontextuellen Rahmen* (Wissensfacette 1) umfasst fünf Kategorien, die im Sinne kompetenzorientierten Biologieunterrichts (KMK, 2004) all jene Überlegungen umfassen, die auf eine Kompetenzförderung (Kat. 1, 3a, 3b, 4) abzielen und die nachweisbar machen (Kat. 2).

*Wissen über das Fachcurriculum* (Wissensfacette 2) umfasst vier Kategorien, die im Sinne der fachlichen Klärung im Rahmen des Modells der didaktischen Rekonstruktion (GROPENGEIER & KATTMANN, 2019) auf den Fachinhalt als solchen (Kat. 5a, 6) sowie seine didaktischen Vermittlungspotentiale (Kat. 5b, 7) abzielen.

*Wissen über die Lerngruppe* (Wissensfacette 3) umfasst sechs Kategorien, die Aspekte der Lerngruppenanalyse aus dem Pentagon-Modell (Kat. 8-13) aufgreifen und die im Sinne der Lernpotential-Diagnose (GROPENGEIER & KATTMANN, 2019) auf Aspekte abzielen, die bei der Planung von Lehr-Lern-Prozessen beachtet werden sollten, beispielsweise Schülervorstellungen (Kat. 9) oder themenspezifische Lernschwierigkeiten (Kat. 10), aber auch allgemeinere Aspekte (Kat. 11-13).

*Wissen über Instruktionsstrategien* (Wissensfacette 4) umfasst 13 Kategorien, die bis auf Kat. 20 und 21 jeweils binär Beschreibungen und Begründungen wesentlicher Vermittlungsstrategien adressieren. Dabei werden auch Aspekte berücksichtigt, die nicht genuin biologiedidaktisch sind, z.B. Maßnahmen der Binnendifferenzierung (Kat. 19a, 19b), die aber dennoch für den Biologieunterricht eine große Rolle spielen und in der Planung berücksichtigt werden sollen.

*Wissen über Beurteilung* (Wissensfacette 5) umfasst vier Kategorien, die auf eine Evaluation des Lernprozesses im Sinne formativen Assessments durch die SuS (Kat. 22a, 22b) oder die Lehrperson selbst (23a, 23b) abzielen und deren Berücksichtigung eine Sicherung, Konsolidierung oder gar Reflexion der Ergebnisse bzw. des Kompetenzzuwachses ermöglicht.

Dabei wurde auf der Grundlage des initialen, deduktiv entwickelten Kategoriensystems anhand der ersten vier Staatsexamensentwürfe geprüft, ob alle bisherigen Kategorien im Datenmaterial nachweisbar sind und ob sich Aspekte identifizieren lassen, die vom Kategoriensystem noch nicht erfasst werden. Diese wurden als neue Sub-Kategorien aufgenommen oder die Beschreibung einer bereits existierenden Subkategorie erweitert, so dass das Kategoriensystems empirisch valide Interpretationen zulässt (SCHREIER, 2014). Sub-Kategorien, die nicht kodiert wurden, wurden aus dem Kategoriensystem entfernt.

Um eine valide Interpretation der Untersuchungsergebnisse zu gewährleisten, wurden Maßnahmen zur Qualitätssicherung der qualitativen Inhaltsanalyse ergriffen (GÖHNER & KRELL, 2020). So wurde in der Entwicklung des Kategoriensystems die semantische Gültigkeit (MAYRING, 2015, S.126) des Kodierleitfadens sichergestellt, indem aus allen Entwürfen sämtliche einer Subkategorie zugeordneten Textstellen darauf hin analysiert wurden, ob die Beschreibungen, Kodierhinweise und Ankerbeispiele konsistent sind und alle Textstellen somit homogen der jeweiligen Subkategorie zuzuordnen sind.

Zusätzlich wurde das Kategoriensystem durch Diskussionen mit Fachdidaktikern und einer Lehrkraft in einem Peer-Debriefing diskutiert (vgl. STEINKE, 2013), wodurch vor allem die Trennschärfe zwischen den Kategorien nachgebessert und aussagekräftigere Ankerbeispiele für die Sub-Kategorien eingefügt werden konnten.

#### **4.4 Anwendung des Kategoriensystems**

Die 16 Staatsexamensentwürfe wurden auf der Grundlage eines Kodierleitfadens (MAYRING, 2015) von zwei Kodierern unabhängig voneinander mit der Software MAXQDA2020 kodiert. Dabei wurde als Analyseinheit ein schriftlicher Unterrichtsentswurf ohne Anhang, als Kodiereinheit ein mindestens vollständiger Teilsatz und als Kontexteinheit der Anhang eines schriftlichen Unterrichtsentswurfs definiert (SCHREIER, 2014). Als Maß für die Intra- und die Interrater-Übereinstimmung wurde Cohen's Kappa ( $\kappa$ ) berechnet. Ein gutes  $\kappa$  ( $0,60 \leq \kappa \leq 0,74$ ; WIRTZ & CASPAR, 2002) für die Intrarater-Übereinstimmung würde ausreichende Stabilität des Untersuchungsverfahrens anzeigen (MAYRING, 2015, S.127), für die Interrater-Übereinstimmung ließe sich auf ausreichende Objektivität schließen (Ebd., S.124).

Nachdem der Erstkodierer für jeden Entwurf auf der Basis zweier Kodiervorgänge eine Konsens-Fassung erstellt hat, wurde diese Fassung mit derjenigen des Zweitkodierers verglichen und Verständigung erzielt (konsensuelles Kodieren; KUCKARTZ, 2016, S.211f.). Auf diese Weise wurde einerseits gewährleistet, dass die Auswertung jedes einzelnen Entwurfs eine valide Interpretation

ermöglicht, andererseits wurden die Kategorienbeschreibungen und Kodierhinweise im Kodierleitfaden optimiert.

In der resultierenden finalen Version wurden vom Erstkodierer PCK-Episoden identifiziert (PARK & CHEN, 2012; POHLMANN, 2019), also Stellen, in denen mindestens zwei der fünf Wissensfacetten in einem Sinnabschnitt gemeinsam auftreten (z.B. in mehreren Zeilen im Fließtext oder innerhalb einer Tabelle). Dies führte zu einer Querverbindung zwischen den entsprechenden Ecken des Pentagons. Die Summe der Verbindungen aus allen PCK-Episoden resultiert in einer PCK-Map. Sie bildet den Vernetzungsgrad des ePCK<sub>p</sub> der Lehrkraft insofern ab, als zum einen ersichtlich wird, welche Wissensfacetten überhaupt miteinander in Beziehung stehen und welche nicht, und als zum anderen die unterschiedlichen Stärken dieser Beziehungen sichtbar werden.

Exemplarisch wurde der Staatsexamensentwurf des Probanden (P) 40 mit besonders stark vernetztem Wissen ausgewählt (PATTON, 1990), um die Anwendung des Kategoriensystems sowie den vollständigen PCK-Mapping-Prozess darstellen zu können. In dem Entwurf zur Stoffwechselphysiologie im 11. Jahrgang (Grundkurs) mit dem Thema „*Seine Leiche war komplett erstarrt... – Ein Fall für die Gerichtsmedizin*“ sollen die SuS einen adressatengerechten Text produzieren, in dem sie den Querbrückenzyklus bei der Muskelkontraktion im Kontext der Leichenstarre erklären.

## 5 Ergebnisse

### 5.1 Entwicklung des Kategoriensystems

Das ursprüngliche, deduktiv entwickelte Kategoriensystem umfasste 28 Kategorien. Drei Sub-Kategorien konnten im Datenmaterial nicht nachgewiesen werden und wurden entfernt, drei Sub-Kategorien wurden induktiv ergänzt (4, 11, 21) und eine weitere (Didaktische Strukturierung) wurde in mehrere Subkategorien unterteilt, so dass v.a. die Wissensfacetten 1 und 4 ausdifferenziert wurden und die Kategorien einen vergleichbaren Körnungsgrad aufweisen.

Tabelle 2 zeigt das Kategoriensystem, mit dem die bislang untersuchten Unterrichtsentwürfe analysiert worden sind und mit dem das ePCK<sub>p</sub> erfasst wird.

### 5.2 Anwendung des Kategoriensystems

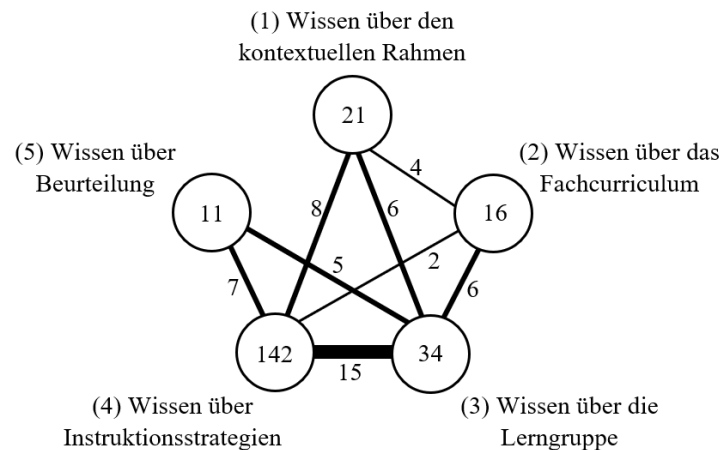
Für die Anwendung des Kategoriensystems auf die 16 bislang untersuchten Staatsexamensentwürfe zeigen sich insgesamt hohe Mittelwerte für die Intrarater-

Tab.2: Kategoriensystem zur Analyse des fachdidaktischen Wissens

PCK	Subkategorie	Beschreibung	Ankerbeispiel	Literaturbezug
1	Kompetenzzuwachs	L. wendet Standard aus dem Lehrplan oder den Richtlinien der KMK auf die geplante Stunde an.	<i>SuS aufgestellten Hypothese zu planen. (P28)</i>	KMK, 2005; Mayer, 2019
		L. nennt Indikatoren, anhand derer ein Kompetenzzuwachs nachgewiesen werden kann.	<i>Die [SuS] können die Fortbewegung des Regenwurms unter Einbeziehung seines [...] Aufbaus erklären. (P5)</i>	Kiper & Mischke, 2009
	3a Sequenz	Beschreibung L. beschreibt den langfristigen Kompetenzerwerb, z.B. in Form einer tabellarischen Sequenzplanung.	<i>Die Kompetenzschwerpunkte [...] der Reihe unterliegen [einem] Wechsel von Fachwissen und Kommunikation. (P40)</i>	Freiman, 2001; Leisen, 2014
		3b Begründung	L. begründet die Struktur des langfristigen Kompetenzerwerbs und somit die Anordnung der Lernschritte.	<i>Ich habe mich für die regelmäßigen Kommunikationsstunden entschieden, weil die [SuS Defizite] in der Textproduktion aufweisen. (P40)</i>
2	4 Unterrichtsprinzipien	L. nimmt Bezug auf Unterrichtsprinzipien, die für den Unterricht eine Leitfunktion darstellen.	<i>Der Unterrichtsgang folgt dem problemorientierten Unterrichtsverfahren ohne Hypothesenbildung. (P40)</i>	Kattmann, 2019
		L. beschreibt die Stellung des Stundenthemas innerhalb des fachlichen Themenkomplexes.	<i>[Es] soll sich mit der durchschnittlichen Pulsfrequenz des Menschen auseinandergesetzt werden. (P6)</i>	Kattmann, 2019; Duit et al., 2012; Gropengießer & Kattmann, 2019
	5a Themenwahl	L. begründet die Auswahl des Themas (ggf. Verweis auf Relevanzkriterien oder Exemplarisches Prinzip).	<i>Zum anderen stellt auch der Tod aus biologischer Sicht ein für Lernende relevantes Thema dar. (P40)</i>	Kattmann, 2019; Duit et al., 2012; Gropengießer & Kattmann, 2019
	5b Sachstrukturanalyse	L. analysiert den Fachinhalt aus fachwissenschaftlicher Perspektive.	<i>Die Temperaturabhängigkeit der Enzyme spielt in der Forensik zur Bestimmung der Todeszeit eine große Rolle. (P40)</i>	Kattmann, 2019; Duit et al., 2012; Gropengießer & Kattmann, 2019
3	7 Didaktische Reduktion	L. beschreibt und begründet Reduktionsentscheidungen.	<i>[...] aus diesem Grund habe ich mich dafür entschieden die neuesten Forschungsergebnisse in reduzierter Form anzugeben. (StEx_S2_F_Ö_04)</i>	Duit et al., 2012; Duit & Treagust, 2012; Hammann & Asshoff, 2014; Kattmann, 2015
		L. beschreibt fachlichen Voraussetzungen der Lerngruppe.	<i>Die Rolle von ATP in dem Querbrückenzyklus wurde bisher im Unterricht nicht thematisiert. (P40)</i>	Duit et al., 2012; Duit & Treagust, 2012; Hammann & Asshoff, 2014; Kattmann, 2015
	9 SuS-Vorstellungen	L. beschreibt Vorstellungen der SuS zu einem Phänomen (i.S. von Alltagsvorstellungen/Präkonzepten)	<i>Die SuS verstehen Evolution als teleologischen Prozess, der abgeschlossen ist, sobald das angestrebte Ziel erreicht ist (P43)</i>	Duit & Treagust, 2012; Hammann & Asshoff, 2014; Kattmann, 2015; Park & Oliver, 2008
		10 Lernschwierigkeiten	L. beschreibt themenspezifische oder kompetenzbedingte Lernschwierigkeiten der SuS.	<i>[...], dass es den Schüler*innen schwerfällt, wesentliche von unwesentlichen Informationen zu unterscheiden, wobei die Schwierigkeit mit der Komplexität und der Abstraktionsebene der Darstellungsform zunimmt (P87)</i>
11 methodische Vorkenntnisse	L. beschreibt die methodischen Voraussetzungen der Lerngruppe (Methoden, Sozialformen).	<i>Die Methode des Lerntempo-Duets wurde bisher ein Mal durchgeführt. (P4)</i>	Park & Oliver, 2008	
	L. beschreibt die sozialen Voraussetzungen der Lerngruppe (z.B. Gruppenstruktur, Konflikte).	<i>Das Klassenklima ist, wenn auch nicht auf den ersten Blick sichtbar, sehr angespannt und belastet. (P1)</i>	Park & Oliver, 2008	
	L. beschreibt die motivationalen Voraussetzungen der Lerngruppe (z.B. Motivation, Interesse, Mitarbeit).	<i>In der Klasse sind einige Schülerinnen und Schüler mit großem Interesse am Fach Biologie, [...]. (P1)</i>	Brophy, 1988; Hidi & Harackiewicz, 2000; Park & Oliver, 2008	

PKK	Subkategorie	Beschreibung	Ankerbeispiel	Literaturbezug	
Wissenfacette 4	14a	Beschreibung L. beschreibt die Struktur der Stunde und Abfolge der einzelnen Unterrichtsschritte.	Der eigentliche Einstieg erfolgt als narrativ-historischer Einstieg in Verbindung mit einem Bildimpuls zu Mendel. (P5)		
	14b	Begründung L. begründet die Struktur der Stunde und Abfolge der einzelnen Unterrichtsschritte.	[...] gibt es eine Zwischensicherung, um sicherzustellen, dass die SuS die Fachbegriffe korrekt verstanden haben, ehe sie weiterarbeiten. (P5)	Gropengießer et al., 2019	
	15a	Beschreibung L. beschreibt den Aufbau und den Einsatz der Aufgabenstellung für den Lernprozess.	Die Aufgaben sind progressiv aufgebaut. (P14)	Anderson & Kratwohl, 2001; Höfle & Jahnke, 2010; Maier et al., 2010	
	15b	Begründung L. begründet den Einsatz der Aufgabenstellung für den Lernprozess.	Außerdem fördert die Aufgabe nochmals auf andere Weise die Kernkompetenz dieser Stunde. (P17)		
	16a	Methode L. beschreibt den Einsatz von Methoden zum Erarbeiten, Sichern, Üben usw.	Die Begriffe „Biotop“ und „Biozönose“ werden in Form eines Lernprodukts erarbeitet. (P49)		
	16b	Begründung L. begründet den Einsatz und die Auswahl von Methoden.	Die Mystery-Methode des soll die Kommunikationsfreudigkeit des Kurses für das Unterrichtsgeschehen nutzen. (P77)	Spöhrase & Ruppert, 2018	
	17a	Beschreibung L. beschreibt den Einsatz von Medien und ihre Gestaltung.	Die SuS untersuchen den Regenwurm mit einer Lupe, einer Glasplatte und Filterpapier. (P7)		
	17b	Begründung L. begründet den Einsatz und die Auswahl von Medien.	Der Einstieg über das Bild der Arztpraxis wurde ausgewählt, da dies einen Bezug zum Alltag der SuS aufweist. (P108)	Gropengießer et al., 2019	
	18a	Beschreibung L. formuliert verbale (operationalisierte) Impulse (oder W-Fragen), die den Unterrichtsgang steuern.	Erklären Sie den vorgefundenen Zustand der Leiche. (P40)		
	18b	Begründung L. begründet den Einsatz von Impulsen unter Berücksichtigung der damit verbundenen Steuerungsfunktion.	Durch den Impuls [...] werden die SuS schnell die naheliegende Problemfrage formulieren können. (P14)	Leisen, 2014	
	19a	Beschreibung L. beschreibt die ausgewählten Differenzierungsmaßnahmen.	SuS, die sehr schnell zu einem Ergebnis in ihrer Gruppe kommen, erhalten ein weiteres Fallbeispiel. (P1)	Hardy et al., 2019;	
	19b	Begründung L. begründet den Einsatz ausgewählter Differenzierungsmaßnahmen.	Damit die B-Schüler ihre Kompetenzen in der Verschriftlichung von biologischen Sachverhalten erweitert, (...). (P40)	Krüger & Meyfarth, 2009	
	20	Antizipation von SuS-Verhalten L. antizipiert Verhalten, konkrete Äußerungen sowie vermutete Schwierigkeiten der SuS.	Ich nehme an, die Mehrzahl der SuS würde spontan einen rezessiven Erbgang vermuten. (P14)	Bromme, 1981	
	21	Alternativplanung L. entwickelt eine alternative Handlungsstrategie als Vorbereitung auf ein nicht wie erwartet verlaufendes Unterrichtsgeschehen.	Bei Zeitknappheit werden die SuS aufgefordert die Problemfrage zu beantworten und die Todeszeit von Josef K. ggfs. den Zustand seiner Leiche zu begründen.	Glodowski & Funke, 1990	
	Wissenfacette 5	22a	Beschreibung L. beschreibt das Feedback der SuS zur Aufgabenstellung, zum Lernprozess oder zur Selbstregulation.	Die anderen Gruppen können die Präsentationen mit ihren Ergebnissen [...] ergänzen. (P1)	
		22b	Begründung L. begründet den Einsatz und die Form des Schülerfeedbacks hinsichtlich ihres Zwecks/Nutzens.	[...], damit alle Lernprodukte gewürdigt werden. (P28)	Gropengießer et al., 2019;
		23a	Beschreibung L. beschreibt das Feedback der Lehrkraft zum Lernprozess bzw. Lernprodukten.	Anhand eines Lösungswortes können die Lernenden sich selbst überprüfen. (P40)	Hattie & Timperley, 2007;
		23b	Begründung L. begründet den Einsatz und die Form des Lehrerfeedbacks hinsichtlich ihres Zwecks/Nutzens.	Da eine schriftliche Sicherung zu lange dauern würde, [...] (P14)	Harlen, 1999

( $M_{\kappa} = 0,76$ ) und die Interrater-Übereinstimmung ( $M_{\kappa} = 0,67$ ) (WIRTZ & CASPAR, 2002). Die Untersuchung des im Folgenden dargestellten Staatsexamensentwurfs ergab eine sehr gute Intrarater- ( $\kappa = 0,72$ ) und eine mäßige Interrater-Übereinstimmung ( $\kappa = 0,47$ ). In ihrer Planung greift die Lehrkraft auf vernetztes fachdidaktisches Wissen zu (Abb. 2).



**Abb. 2:** PCK-Map von P40 (Zahlen in Kreisen geben die Summe kodierter Stellen aller Subkategorien je Wissensfacette an; Zahlen an Linien geben die Anzahl der PCK-Episoden an, in denen je zwei Wissensfacetten gemeinsam kodiert wurden; dünne Linien 0-4 Mal, mittlere Linien 5-9 Mal, dicke Linien > 10 Mal)

Alle Wissensfacetten wurden in diesem Entwurf mehrfach kodiert, am häufigsten die Instruktionsstrategien. Dabei werden vorwiegend Maßnahmen der didaktischen Strukturierung beschrieben und begründet. Im Gegensatz zu allen anderen bisher untersuchten Entwürfen wird hier jedoch auch explizit Bezug auf die Aufgabenstellungen (Kat. 15a, 15b) genommen und die Gestaltung des Lernprozesses somit begründet:

*In der zweiten Aufgabe sollen die [SuS] anhand einer Info-Box ein gerichtsmedizinisches Gutachten schreiben, womit sie den vorgefundenen Zustand der Leiche erklären und den Todeszeitpunkt begründen können.<sup>15a</sup> [...] Um den Spannungsbogen und die Motivation der Lernenden im Unterrichtsverlauf aufrechtzuerhalten<sup>15b</sup>, [...] ermitteln [sie] die Sterbezeit der gefundenen Leiche. Anhand dieser Information können sie die Problemfrage beantworten [...] <sup>15a</sup>*

Auf diese Weise hebt sich der vorliegende Entwurf von den anderen Staatsexamensentwürfen ab und verbleibt nicht vorwiegend auf einer deskriptiven Ebene, in der der Verlauf des Unterrichts, Methoden, Medien usw. geschildert werden, sondern begründet die zentrale Lernhandlung der SuS in Form der Aufgabenstellungen unter Einbeziehung fachlicher und motivationaler Faktoren.

Besonders auffällig an der PCK-Map ist, dass die Schülerkognition mit allen Wissensfacetten stark vernetzt ist. Planungsentscheidungen werden also insbesondere mit Blick auf die spezifische Lerngruppe getroffen, so dass der Unterricht auf der Grundlage einer fundierten Analyse der Lerngruppe geplant wird:

*Der Schwerpunkt der heutigen Stunde soll auf der Textproduktion liegen<sup>1</sup>, da ich in den letzten Wochen feststellte, dass die Lernenden Defizite bei der Formulierung eines Fließtextes aufwiesen.<sup>8</sup> Mit Hilfe eines Ausschnitts eines Zeitungsartikels [...] sollen die Lernenden [...] die Leichenstarre erklären.<sup>17a</sup> Im Hinblick auf die Kursphase erachte ich es als notwendig den Lernenden geeignete Strategien zur Formulierung von aufgabenzentrierten Fließtexten an die Hand zu geben.<sup>3b</sup>*

Ganz im Sinne eines an die Voraussetzungen der spezifischen Lerngruppe anknüpfenden, kompetenzorientierten Biologieunterrichts wird hier eine langfristige Entwicklung der Schreibkompetenz verfolgt. Die Gestaltung der Lernumgebung wird nicht als Selbstzweck verstanden, sondern ist argumentativ miteinander verwoben: Die Leichenstarre und ihre Ursachen als fachlicher Inhalt, die Vorentlastung durch den Zeitungsartikel und das gerichtsmedizinische Gutachten als motivierendes Textformat mit seinen formalen Kriterien werden hier aufeinander abgestimmt und können somit als funktional für die Förderung der Schreibkompetenz erachtet werden. Dies spiegelt sich auch in der Überlegung wider, Lernzuwachs durch Feedback sichtbar zu machen:

*Die Erarbeitung ist [...] in Partnerarbeit vorge[se]hen.<sup>16a</sup> Ich habe mich für diese kooperative Sozialform entschieden, um den Schreibprozess zu entlasten<sup>16b</sup> und damit sich in der Sicherung jeweils vier Lernende einer Tischreihe gegenseitig eine schriftliche Rückmeldung geben können.<sup>22a</sup> Um den Zustand [...] der Leiche [...] erklären zu können, müssen die Lernenden auf den Querbrückenzyklus zurückgreifen.<sup>6</sup> Diese haben die Lernenden in der letzten Woche erarbeitet<sup>8</sup>, aber noch nicht schriftlich festgehalten.<sup>3a</sup>*

Zwar wird hier nicht ausgeführt, in welcher Form und mit welchem Ziel das Schülerfeedback stattfinden soll, gleichwohl zeigt sich, dass die Lehrkraft hier berücksichtigt, dass es zu den Lernprodukten eine schülerzentrierte, metakognitive Reflexionsphase geben soll, in der Lernzuwachs sichtbar und Lernpotential definiert werden könnte. Diese Lehrkraft verknüpft also verschiedene fachdidaktische Wissensbestände miteinander und legt damit die didaktische Begründung für die Gestaltung des Lehr-Lern-Prozesses dar.

## 6 Diskussion

### 6.1 Entwicklung des Kategoriensystems

Das Ziel dieser Studie war es, ein Kategoriensystem zu entwickeln, mit dem das situationsbezogene fachdidaktische Wissen von Lehrkräften in schriftlichen Unterrichtsplanungen erfasst und analysiert werden kann.

Ausgehend von dem Befund, dass in vielen Studien zum fachdidaktischen Wissen nicht dargestellt wird, welche Art von PCK erfasst werden soll, und dass ein Vergleich empirischer Befunde aufgrund heterogener Konzeptualisierungen des Konstrukts PCK nahezu unmöglich sei, entwickeln CHAN et al. (2019) eine Schablone, wie Kategoriensysteme zur Erfassung von PCK gestaltet sein können. Von der Empfehlung, für alle Kategorien Niveaustufen zu definieren und damit im Sinne SCHREIERS (2014) eine evaluative Inhaltsanalyse vorzunehmen, wurde aus



zeitökonomischen Gründen abgesehen. Stattdessen wurde dem *pedagogical reasoning* (LOUGHRAN et al., 2016) große Bedeutung beigemessen – nicht als isolierte Kategorie, sondern integriert in die jeweiligen Wissensfacetten. Das angewandte Vorgehen hat folglich den Nachteil, keine numerischen Aussagen über die Ausprägung des fachdidaktischen Wissens treffen zu können. Es wird daher aktuell geprüft, ob sich zumindest für alle Begründungskategorien Niveaustufen einführen lassen, die als Indikator für die professionellen Kompetenzen von Lehrkräften dienen können (LIVINGSTON & BORKO, 1989).

Da in einem deduktiven Ansatz a priori Kategorien auf der Grundlage des Pentagon-Modells (PARK & OLIVER, 2008) definiert wurden, die sich folglich innerhalb eines etablierten konzeptionellen, theoretischen Rahmens bewegen, können die hier dargestellten Analyseergebnisse mit Befunden anderer Studien verglichen werden. Aufgrund der induktiven Nachschärfungen (KUCKARTZ, 2016) kann von einer hohen Passung zwischen dem Kategoriensystem und den schriftlichen Unterrichtsplanungen ausgegangen werden, so dass eine valide Interpretation der Untersuchungsergebnisse ermöglicht wird.

Unter Beachtung der oben aufgeführten Maßnahmen zur Sicherung der Qualität des Kategoriensystems werden insgesamt hohe Intrarater- ( $M_{\kappa} = 0,76$ ) und Interrater-Übereinstimmungswerte ( $M_{\kappa} = 0,67$ ) (WIRTZ & CASPAR, 2002) erreicht, was auf hohe Stabilität des Auswertungsverfahrens und hinreichende Auswertungsobjektivität (MAYRING, 2020) hinweist. Insbesondere die Prüfung auf semantische Gültigkeit (MAYRING, 2015) und das Peer-Debriefing (STEINKE, 2013) haben zur weiteren Ausdifferenzierung und somit zu trennscharfen Kategorien (v.a. 1-4, 10-12 und 16-27) geführt. Die mäßige Interrater-Übereinstimmung des analysierten Entwurfs ( $\kappa = 0,47$ ) lässt sich damit erklären, dass zu diesem Zeitpunkt noch Modifikationen am Kategoriensystem vorgenommen wurden.

## 6.2 Anwendung des Kategoriensystems

Ein weiteres Ziel dieser Studie war es, anhand eines exemplarischen Staatsexamensentwurfs die Anwendbarkeit des Kategoriensystems zu demonstrieren und Aussagen über das sich darin manifestierende fachdidaktische Wissen zu treffen. Im Einklang mit WESTERMAN (1991) deutet sich anhand des hier dargestellten Staatsexamensentwurfs an, dass die Referendar\*innen als Personen am Übergang zwischen Novizen- und Expertentum zum Teil bereits über elaborierte kognitive Schemata verfügen, die sie in der Planung anwenden. Die Planung von Biologieunterricht ist auf die Lerngruppe ausgerichtet (KÖNIG et al., 2015), so dass der Unterrichtsentwurf hier als Werkzeug genutzt wird. Dennoch ist auffällig, dass zahlreiche Planungsentscheidungen beschrieben und z. T. begründet werden, die nicht unmittelbar für den Lernprozess relevant sind. Hier

wird ein hohes Auflösungs-niveau gezeigt (LIVINGSTON & BORKO, 1989), was jedoch auch am Kontext (Staatsexamensprüfung) liegen kann, im Rahmen dessen der Entwurf verfasst worden ist.

Einschränkend sei erwähnt, dass zur prognostischen Validität des fachdidaktischen Wissens für die Unterrichtsqualität keine Aussagen getroffen werden können, da keine Daten im Rahmen der Unterrichtsdurchführung vorliegen.

### 6.3 Methodenkritik

Die Darstellung des exemplarischen Falls zeigt, dass der PCK-Map-Approach eine differenzierte Erfassung des fachdidaktischen Wissens möglich macht. Zwei methodische Schwächen sind bei der Analyse der Staatsexamensentwürfe diesbezüglich anzumerken: Erstens vereinfacht der methodische Zugang das komplexe Konstrukt PCK insofern, als es keine Gewichtung innerhalb der PCK-Maps gibt und suggeriert wird, dass alle Verknüpfungen gleich wichtig seien (PARK & CHEN, 2012). POHLMANN (2019) folgend wird hier jedoch der Standpunkt vertreten, dass dieses Design einen leichten Zugang zum fachdidaktischen Wissen erlaubt, der die methodischen Schwächen überwiegt. Zweitens folgen die Staatsexamensentwürfe einer im Wesentlichen vorgegebenen Struktur (SENBJF, 2017): Durch die Kapitelfolge werden bestimmte Aspekte oft in den entsprechenden Kapiteln erwähnt und dann zur Vermeidung von Redundanzen in anderen Kapiteln nicht wieder aufgegriffen, was die Wahrscheinlichkeit minimiert, dass z.B. Überlegungen zum Fachcurriculum und zur Beurteilung des Kompetenzzuwachses gemeinsam auftreten (vgl. Abb. 2). Dies wiederum führt zu einer geringeren Vernetzung bestimmter Wissensfacetten innerhalb der PCK-Map, was die Aussagekraft über das ePCK<sub>p</sub> etwas mindert. Um ein so komplexes Konstrukt wie das fachdidaktische Wissen nicht zu sehr zu vereinfachen, wird daher bei der Analyse der Entwürfe zusätzlich zur Kategorienorientierung auch fallorientiert vorgegangen (KUCKARTZ, 2019), so dass neben der Quantifizierung in Form der PCK-Maps zusätzlich für jeden Fall eine hermeneutische Textanalyse der Staatsexamensentwürfe vorgenommen wird.

Hinsichtlich der Staatsexamensentwürfe als Datenmaterial gilt es zwei weitere Besonderheiten bei der Interpretation der Daten zu berücksichtigen: Zum einen ist fraglich, ob in unserem Ansatz tatsächlich *enacted PCK* als individuelle Disposition und nicht zumindest in Teilen auch *collective PCK* als kanonisches Expertenwissen (CARLSON & DAEHLER, 2019) erfasst wird, schließlich werden Entwürfe im Rahmen einer Staatsprüfung gewiss von zahlreichen anderen, auch didaktisch geschulten Personen gelesen, so dass deren Ansichten miteinfließen. Darüber hinaus visualisiert die PCK-Map keineswegs zwangsläufig die Qualität der Entwürfe. Erfasst wird bloß, ob zwei Wissensfacetten gemeinsam auftreten,

nicht jedoch, ob die entsprechenden Planungsentscheidungen funktional zur Förderung der ausgewählten Kompetenzen sind. Zwar deutet eine relativ hohe Vernetzung des ePCK<sub>p</sub> auf ein elaborierteres fachdidaktisches Wissen hin (LIVINGSTON & BORKO, 1989; WESTERMAN, 1991), allerdings spiegelt sich darin weder die Qualität der Unterrichtsplanung im Allgemeinen noch die Qualität der Begründungen wider (LOUGHRAN et al., 2016; CHAN et al., 2019). Daher wird im weiteren Forschungsprozess der hier vorgestellte analytische Ansatz zur Textbeurteilung durch eine holistische Qualitätsbeurteilung der Entwürfe in Form eines Expertenratings ergänzt und untersucht, welche Aspekte des ePCK<sub>p</sub> als Prädiktoren für hohe Unterrichtsplanungsqualität gelten können.

## 7 Fazit

Die vorliegende Studie beschreibt die Entwicklung eines Kategoriensystems zur Erfassung des fachdidaktischen Wissens von Biologie-Lehrkräften und erläutert anhand der exemplarischen Analyse eines Staatsexamensentwurfs seine Anwendbarkeit. Es hat sich bei der bisherigen Analyse als tragfähig erwiesen. Es bleibt zu prüfen, ob eine Erweiterung um Niveaustufen sinnvoll ist, um die Qualität der Begründungen zu erfassen (CHAN et al., 2019). Zunächst werden weitere Staatsexamensentwürfe mit dem Kategoriensystem analog zum hier dargestellten Fall analysiert. Mittelfristig wird das Kategoriensystem außerdem auf die Unterrichtsreflexion von Referendar\*innen (ePCK<sub>r</sub>) angewendet werden, um weitere Evidenzen für die valide Interpretation der vorliegenden Ergebnisse zu sammeln.

## Danksagung

Das Projekt *K2Teach* wird im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1802 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren. Die Autoren danken der Senatsverwaltung für Bildung, Jugend und Familie des Landes Berlin für die Bereitstellung der Staatsexamensentwürfe.

## Zitierte Literatur

- ALONZO, A. C., BERRY, A. & NILSSON, P. (2019). *Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK*. In: HUME, A., COOPER, R., & BOROWSKI, A. [Hrsg.]: *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, Singapur, 271-286.
- ALONZO, A. C. & KIM, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, 53(8), 1259-1286.

- ANDERSON, L.W. & KRATWOHL, D. (2001). A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. New York.
- BAUMERT, J. & KUNTER, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469-520.
- BAYLOR, A. L. (2002). Expanding preservice teachers' metacognitive awareness of instructional planning through pedagogical agents. *Educational Technology Research and Development*, 50(2), 5-22.
- BROMME, R. (1981). Das Denken von Lehrern bei der Unterrichtsvorbereitung: eine empirische Untersuchung zu kognitiven Prozessen von Mathematiklehrern. Beltz.
- BROMME, R. & SEEGER, F. (1979). *Unterrichtsplanung als Handlungsplanung. Eine psychologische Einführung in die Unterrichtsvorbereitung*. Scriptor, Königstein.
- BROPHY, J. (1988). Educating teachers about managing classrooms and students. *Teaching and teacher Education*, 4(1), 1-18.
- BROVELLI, D., BÖLSTERLI, K., REHM, M. & WILHELM, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettentest mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat. *Unterrichtswissenschaft*, 41 (4), 306-329.
- CARLSON, J. & DAEHLER, K. R. (2019). *The refined consensus model of pedagogical content knowledge in science education*. In: HUME, A., COOPER, R., & BOROWSKI, A. [Hrsg.]: *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, Springer, Singapur, 77-92.
- CHAN, K. K. H. & HUME, A. (2019). *Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies*. In: HUME, A., COOPER, R., & BOROWSKI, A. [Hrsg.]: *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, Springer, Singapur, 3-76.
- CHAN, K. K. H., ROLLNICK, M. & GESS-NEWSOME, J. (2019). *A Grand Rubric for Measuring Science Teachers' Pedagogical Content Knowledge*. In: HUME, A., COOPER, R., & BOROWSKI, A. [Hrsg.]: *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, Springer, Singapur, 251-269.
- COE, R., ALOISI, C., HIGGINS, S., & MAJOR, L. E. (2014). What makes great teaching? Review of the underpinning research. Retrieved from <http://dro.dur.ac.uk/13747/1/13747.pdf> (24.04.2020)
- DUIT, R., GROPENIEBER, H., KATTMANN, U., KOMOREK, M., & PARCHMANN, I. (2012). The model of educational reconstruction – A framework for improving teaching and learning science. In *Science education research and practice in Europe*, Brill Sense, 13-37.
- DUIT, R. & TREAGUST, D. F. (2012). How can conceptual change contribute to theory and practice in science. In B. J. Fraser, K. Tobin & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 107–118). Dordrecht, The Netherlands: Springer.
- FREIMAN, T. (2001): Kumulatives Lernen im Biologieunterricht, in: *Praxis der Naturwissenschaften - Zeitschrift für den experimentellen Unterricht*, Nr.50/7, S.1-2.
- GASSMANN, C. (2013). *Erlebte Aufgabenschwierigkeit bei der Unterrichtsplanung: Eine qualitativ-inhaltsanalytische Studie zu den Praktikumsphasen der universitären Lehrerbildung*. Springer, Wiesbaden.
- GESS-NEWSOME, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In: BERRY, A., FRIEDRICHSEN, P. J. & LOUGHRAN, J. [Hrsg.]: *Re-examining pedagogical content knowledge in science education*. Routledge, New York, 28-42.
- GŁODOWSKI, A. S. & FUNKE, J. (1990). Planen und Problemlösen: Überlegungen zur neuropsychologischen Diagnostik von Basiskompetenzen beim Planen. *Zeitschrift für Neuropsychologie*, 1, 139-148.
- GÖHNER, M. & KRELL, M. (2020). Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review. *Zeitschrift für Didaktik der Naturwissenschaften*, Springer (<https://doi.org/10.1007/s40573-020-00111-0>) (24.04.2020).
- GROPENIEBER, H., HARMS, U. & KATTMANN, U. [Hrsg.] (2019). *Fachdidaktik Biologie*, 11. Auflage, Aulis, Hallbergmoos.
- GROPENIEBER, H. & KATTMANN, U. (2019): Didaktische Rekonstruktion. In: Gropengießer, H., Harms, U. & Kattmann, U.: *Fachdidaktik Biologie*, 11. Auflage, Hallbergmoos 2019, S.16-23.
- GROBSCHEDL, J., WELTER, V. & HARMS, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402-439.
- HAMMANN, M., & ASSHOFF, R. (2014). Schülervorstellungen im Biologieunterricht. Ursachen für Lernschwierigkeiten. Stuttgart.

- HARDY, I., DECRISTAN, J., & KLIEME, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for educational research online*, 11(2), 169-191.
- HARLEN, W. (1999). Effective Teaching of Science. A Review of Research. Using Research Series, 21. Scottish Council for Research in Education, 15 St. John Street, Edinburgh EH8 8JR, Scotland.
- HATTIE, J. & TIMPERLEY, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- HATTIE, J. & YATES, G. C. R. (2015). *Lernen sichtbar machen*. Überarbeitete deutschsprachige Ausgabe von „Visible Learning“ besorgt von Wolfgang BEYWL und Klaus ZIERER. Schneider Hohengehren, Baltmannsweiler.
- HIDI, S., & HARACKIEWICZ, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70(2), 151-179.
- HÖBLE, C., & JAHNKE, L. (2010). Gute Lernaufgaben für den Biounterricht? – Eine große Herausforderung. *Lernaufgaben und Lernmaterialien im kompetenzorientierten Unterricht*, 167-178.
- HUME, A., COOPER, R. & BOROWSKI, A. [Hrsg.] (2019). *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*, Springer, Singapur.
- JÜTTNER, M., BOONE, W., PARK, S. & NEUHAUS, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, 25(1), 45-67.
- KATTMANN, U. (2019): Auswahl und Verknüpfung der Lerninhalte. In: Gropengießer, H., Harms, U. & Kattmann, U.: *Fachdidaktik Biologie*, 11. Auflage, Hallbergmoos 2019, S.29-38.
- KATTMANN, U. (2015). Schüler besser verstehen. Alltagsvorstellungen im Biologieunterricht, Aulis-Verlag, Köln.
- KIND, V. & CHAN, K. K. (2019). Resolving the amalgam: connecting pedagogical content knowledge, content knowledge and pedagogical knowledge. *International Journal of Science Education*, 41(7), 964-978.
- KIPER, H., & MISCHKE, W. (2009). *Unterrichtsplanung*. Beltz.
- KÖHLER, K. (2013): Nach welchen Prinzipien kann Biologieunterricht gestaltet werden? In: Spörhase, U. (Hrsg.): *Biologie-Didaktik. Praxishandbuch für die Sekundarstufe I und II*, 6. Auflage. Cornelsen, Berlin: 112-129.
- KONFERENZ DER KULTUSMINISTER DER LÄNDER IN DER BUNDESREPUBLIK DEUTSCHLAND (KMK) [HRSG.] (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Beschluss vom 16.12.2004. Wolters Kluwer, München und Neuwied.
- KÖNIG, J., BUCHHOLTZ, C., & DOHMEN, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18(2), 375-404.
- KUCKARTZ, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (3. Aufl.). Beltz Juventa, Weinheim.
- KUCKARTZ, U. (2019). Qualitative Inhaltsanalyse: von Kracauers Anfängen zu heutigen Herausforderungen. *Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research*, 20(3), Art. 12, <http://dx.doi.org/10.17169/fqs-20.3.3370>. (24.04.2020)
- LEISEN, J. (2014): Wie soll ich denn meinen Unterricht planen? Lehr-Lern-Prozesse planen am Beispiel Elektrizitätslehre in Physik. In: Maier, U. (Hrsg.): *Lehr-Lernprozesse in der Schule: Referendariat. Praxiswissen für den Vorbereitungsdienst*, Bad Heilbrunn, S.102-117.
- LIVINGSTON, C. & BORKO, H. (1989). Expert-novice differences in teaching: A cognitive analysis and implications for teacher education. *Journal of teacher education*, 40(4), 36-42.
- LOUGHRAN, J., KEAST, S. & COOPER, R. (2016). Pedagogical reasoning in teacher education. *International handbook of teacher education*, 387-421.
- MAIER, U., KLEINKNECHT, M., METZ, K., & BOHL, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 28(1), 84-96.
- MAYRING, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Beltz, Weinheim.
- MAYRING P. (2020). *Qualitative Inhaltsanalyse*. In: MEY, G. & MRUCK, K. [Hrsg.]: *Handbuch Qualitative Forschung in der Psychologie*. Springer, Wiesbaden.
- NEUMANN, K., KIND, V. & HARMS, U. (2019). Probing the amalgam: the relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, 41(7), 847-861.

- NEUWEG, G. H. (2011). *Das Wissen der Wissensvermittler. Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerbewusstsein*. In: *Handbuch der Forschung zum Lehrerberuf*. Waxmann, Münster, 583-614.
- PARK, S. & CHEN, Y. C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922-941.
- PARK, S. & OLIVER, J. S. (2008). National board certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45(7), 812-834.
- PARK, S. & SUH, J. K. (2019). *The PCK Map Approach to Capturing the Complexity of Enacted PCK (ePCK) and Pedagogical Reasoning in Science Teaching*. In: HUME, A., COOPER, R., & BOROWSKI, A. [Hrsg.]: *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, Springer, Singapur, 185-197.
- PATTON, M. Q. (1990). *Qualitative evaluation and research methods*. Sage, Beverly Hills (CA).
- POHLMANN, M. (2019). *Förderung ethischer Bewertungskompetenz: der Einfluss ausgewählter Lerngelegenheiten auf die inhaltliche Ausdifferenzierung und die Kohärenz der Komponenten des fachdidaktischen Wissens von Biologielehrkräften*, Oldenburg.
- ROSENBERGER, K. (2018). *Unterrichten: Handeln in kontingenten Situationen*. Beltz, Weinheim/Basel.
- SCHREIER, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten, *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 15(1), Art. 18, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1401185> (24.04.2020).
- SENATSV ERWALTUNG FÜR BILDUNG, JUGEND UND FAMILIE (2017). *Handbuch Vorbereitungsdienst. Materialien für den reformierten Berliner Vorbereitungsdienst* (6. Aufl.). Berlin.
- SHULMAN, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(4), 4-14.
- SPÖRHASE-EICHMANN, U., & RUPPERT, W. (HRSG.). (2018). *Biologie-Methodik: Handbuch für die Sekundarstufe I und II*, 5. Auflage, Cornelsen.
- STEINKE, I. (2013). Gütekriterien qualitativer Forschung. In: FLICK, U., VON KARDORFF, E. & STEINKE, I. [Hrsg.]: *Qualitative Forschung: Ein Handbuch*. Rowohlt, Reinbek bei Hamburg.
- TARDENT KUSTER, J. (2020). Unterrichtsplanungen von angehenden Lehrpersonen zum experimentellen Handeln. Pädagogische Hochschule Heidelberg, Heidelberg. <https://opus.ph-heidelberg.de/frontdoor/index/index/docId/353> (24.04.2020).
- VOGELSANG, C. & RIESE, J. (2017). *Wann ist eine Unterrichtsplanung 'gut'? - Planungsperformanz in Praxisratgebern zur Unterrichtsplanung*. In: WERNKE, S. & ZIERER, K. [Hrsg.]: *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Bad Heilbrunn*, 47-60.
- WAHL, D. (2002). Mit Training vom trägen Wissen zum kompetenten Handeln? *Zeitschrift für Pädagogik*, 48 (2), 227-241.
- WEINGARTEN, J. (2019). *Wie planen angehende Lehrkräfte ihren Unterricht?: Empirische Analysen zur kompetenzorientierten Gestaltung von Lernangeboten*. Waxmann Verlag, Münster.
- WERNKE, S., & ZIERER, K. (2017). *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! In: WERNKE, S. & ZIERER, K. [Hrsg.]: Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung*. Klinkhardt, Bad Heilbrunn. 7-16.
- WESTERMAN, D. A. (1991). Expert and novice teacher decision making. *Journal of teacher education*, 42(4), 292-305.
- WIRTZ, M., & CASPAR, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität [Rater agreement and rater reliability]. Hogrefe, Göttingen.
- ZIERER, K., WERNER, J. & WERNKE, S. (2015). Besser planen? Mit Modell! Empirisch basierte Überlegungen zur Entwicklung eines Planungskompetenzmodells. *DDS – Die Deutsche Schule*, 107 (4), 375-395.



### Beitrag 3

**Großmann, L.** & Krüger, D. (2022). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, 28, 4. <https://doi.org/10.1007/s40573-022-00141-w>





# Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe?

Leroy Großmann<sup>1</sup> · Dirk Krüger<sup>1</sup>

Eingegangen: 15. Juli 2021 / Angenommen: 9. März 2022  
© Der/die Autor(en) 2022

## Zusammenfassung

Die Unterrichtsplanung ist eine der Kernaufgaben von Lehrkräften. Insbesondere schriftliche Unterrichtsentwürfe gehören zum festen Bestandteil der Lehrkräftebildung. In ihnen zeigt sich das fachdidaktische Wissen (PCK), das in der Entwicklung und Begründung des Unterrichts genutzt wird und das als Prädiktor für hohe Unterrichtsqualität einerseits und für Lernerfolg von Schüler\*innen andererseits gilt. Empirische Untersuchungen zu Unterrichtsentwürfen im Fach Biologie sowie zur Bedeutung des PCK in der Planung sind bislang rar, obwohl eine Untersuchung der Planungsprodukte angehender Biologie-Lehrkräfte Aufschluss darüber geben könnte, in welchen Bereichen die Ausbildungsinstitutionen der ersten und zweiten Phase der Lehrkräftebildung noch stärker unterstützend einwirken könnten. In der vorliegenden Studie werden Staatsexamensentwürfe im Fach Biologie ( $N=46$ ) untersucht, indem PCK-Elemente und ihre Vernetzungen mittels einer qualitativen Inhaltsanalyse identifiziert werden und analysiert wird, in welchen Aspekten sich gelungene von weniger gelungenen Unterrichtsentwürfen unterscheiden. Die Ergebnisse liefern Hinweise, dass die Qualität eines Unterrichtsentwurf mit der Vernetztheit des PCK korreliert. Allerdings zeigt ein qualitativer Fallvergleich, dass auch sehr gute Entwürfe eher schwach ausgeprägtes PCK enthalten können und vice versa, so dass die Anzahl der PCK-Verbindungen allein keinen hinreichenden Qualitätsindikator darstellt. Anhand ausgewählter Beispiele wird beschrieben und diskutiert, inwiefern sich in der Berücksichtigung von PCK-Elementen qualitative Unterschiede identifizieren lassen und welche Aspekte in Unterrichtsentwürfen besonders berücksichtigt werden sollten.

**Schlüsselwörter** Biologieunterricht · Unterrichtsplanung · Planungskompetenz · Fachdidaktisches Wissen (PCK) · Referendariat · Qualitative Inhaltsanalyse

---

✉ Leroy Großmann  
leroy.grossmann@fu-berlin.de

Dirk Krüger  
dirk.krueger@fu-berlin.de

<sup>1</sup> Didaktik der Biologie, Freie Universität Berlin, Schwendenerstraße 1, 14195 Berlin, Deutschland

## What's the Role of Trainee Biology Teachers' Pedagogical Content Knowledge for the Quality of their Written Lesson Plans?

### Abstract

Lesson planning is a core part of teachers' expertise. Particularly written lesson plans play a major role in teacher training. They reflect the pedagogical content knowledge (PCK) that teachers use to develop and justify their lessons and that has proven to be a predictor of teaching quality on the one hand and of students' achievement on the other hand. Yet, empirical research on written lesson plans in biology as well as the role of PCK in lesson planning is scarce. In fact, an exploration on biology teachers' lesson planning could help to identify issues that prospective teachers struggle with and that need to be addressed more intensively in teacher training. In this study, lesson plans from the final exam of teacher training in Germany ( $N=46$ ) are analyzed by means of qualitative content analysis. PCK elements and their interconnections are identified. Based on that, it is shown in what aspects good lesson plans differ from poor lesson plans. The results indicate that lesson plan quality correlates with the interconnectedness of PCK, e.g. in good lesson plans, the relation between curricular standards and knowledge of students' understanding is significantly more frequently addressed. However, a comparative case analysis shows that good lesson plans can contain rather poorly interconnected PCK and vice versa. Hence, the sole amount of PCK connections is not a sufficient indicator of lesson plan quality. Using different examples, we show qualitative differences in addressing PCK elements and discuss which aspects might be particularly important in written lesson plans.

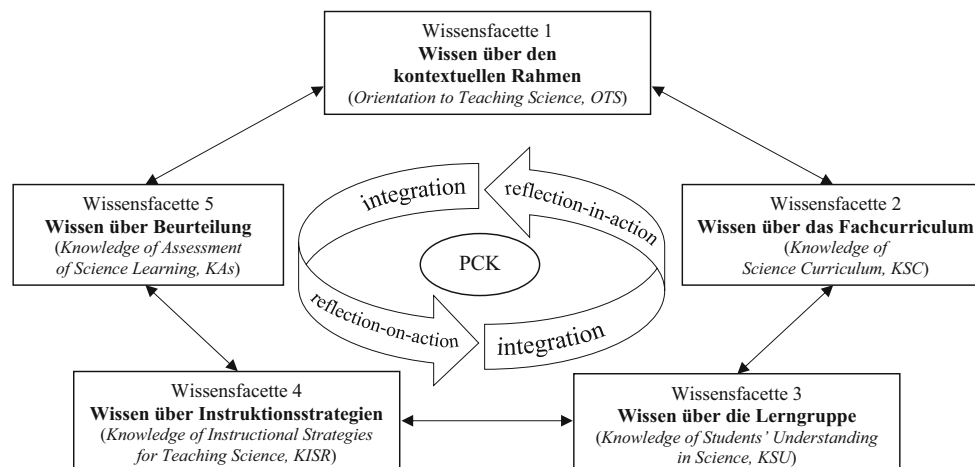
**Keywords** Biology education · Lesson planning · Planning competence · Pedagogical content knowledge (PCK) · In-service teacher training · Qualitative content analysis

### Einleitung

Die Unterrichtsplanung gilt als eine der Kernaufgaben von Lehrkräften. Dies spiegelt sich sowohl in den Standards für die Lehrerbildung (KMK 2019), in allgemeindidaktischen Ansätzen zur Strukturierung professioneller Kompetenz (Baumert und Kunter 2006) als auch in naturwissenschafts-didaktischen Modellen zum fachdidaktischen Wissen von Lehrkräften (Carlson und Daehler 2019) wider. Angesichts dieser Relevanz in Theorie und Praxis überrascht es, dass die wenigen aktuellen empirischen Untersuchungen zur Unterrichtsplanung einhellig eine defizitäre Forschungslage verzeichnen (z.B. Gassmann 2013; Stender 2014; Weingarten 2019; Zaragoza et al. 2021). Bezüglich der Fähigkeit von Lehrkräften zur Unterrichtsplanung sprechen Wernke und Zierer (2017) sogar von einem „in Vergessenheit geratenen Kompetenzbereich“. Aktuelle Studien zur Unterrichtsplanung stützen sich im Wesentlichen immer noch auf die kognitionspsychologischen Untersuchungen der 1970er und 1980er-Jahre, die oft in verschiedenen Unterrichtsfächern mit (sehr) kleinen Stichproben und unterschiedlichen methodischen Zugängen arbeiteten (vgl. Gassmann 2013; Weingarten 2019). Derzeit liegt weder ein etabliertes allgemeindidaktisches Planungskompetenzmodell (Rothland 2021; Werner et al. 2017) noch eine deskriptive biologiedidaktische Exploration des in der Planung genutzten fachdidaktischen Wissens oder eine „kritisch[e] Bestandsaufnahme der gegenwärtigen Planungspraxen von Lehrkräften“ (Weingarten 2019, S. 28) vor. Insbesondere letztere wären jedoch aus zweierlei Gründen bedeutsam:

Erstens gehören in der zweiten Phase der Lehrkräftebildung Unterrichtsbesuche durch Seminar- und Schulleitungen zum konstitutiven Teil der Ausbildung. Zu diesem Zweck müssen Unterrichtsentwürfe verfasst werden, die in vielen Bundesländern auch durch die Auszubildenden beurteilt werden (Strietholt und Terhart 2009). Die dabei zugrunde gelegten Beurteilungskriterien werden von einem beträchtlichen Anteil an Referendar\*innen und Auszubildenden als uneinheitlich empfunden (Döbrich und Abs 2008; Käerner et al. 2019). Evidenzbasierte Kriterien zur Einschätzung solcher Unterrichtsentwürfe hätten daher wichtige Implikationen für die Transparenz der Leistungserwartungen und -beurteilungen in der Ausbildung. Zweitens wurde gezeigt, dass die Planungskompetenz von Referendar\*innen stark mit der Qualität durchgeführten Unterrichts korreliert und die Unterrichtsplanung somit als Prädiktor für die professionelle Durchführung von Unterricht gelten kann (Weingarten 2019). Dabei bildet das fachdidaktische Wissen für die Unterrichtsplanung eine wichtige Ressource, weil es mit wichtigen Variablen der Unterrichtsqualität (z.B. der kognitiven Aktivierung) positiv korreliert (Keller et al. 2017; Kunter et al. 2013; Park et al. 2011). Aus einer Analyse authentischer Entwürfe ließen sich Kriterien ableiten, die nicht bloß zur Bewertung von Unterrichtsentwürfen durch die Ausbildungsinstitutionen, sondern auch von angehenden Lehrkräften selbst genutzt werden können, um die Qualität ihrer Entwürfe einzuschätzen (Brookhart 2018). Somit könnten Gelingensbedingungen wichtiger Aspekte der Unterrichtsplanung im Allgemeinen (z. B. Progression des Kompetenzförderung, Transparenz) sowie im

**Abb. 1** Pentagon-Modell des fachdidaktischen Wissens (PCK) von Lehrkräften naturwissenschaftlicher Fächer (nach Park und Oliver 2008)



Fach Biologie im Besonderen (z. B. die Berücksichtigung von Schülervorstellungen, Hammann und Asshoff 2014; die Reflexion über Arbeitsweisen wie das Experimentieren, Gropengießer 2020) transparent gemacht werden. Dies könnte einen positiven Einfluss auf die Qualität der Lehr-Lern-Prozesse und damit auf den Lernerfolg der Lernenden haben. Schröder et al. (2020) stellen einen dringenden Bedarf an Instrumenten fest, mit denen die schriftliche Planung angehender Lehrkräfte objektiv und reliabel mithilfe transparenter Qualitätskriterien eingeschätzt werden kann. Ziel dieser Studie ist es unter diesem Gesichtspunkt, authentische Unterrichtsentwürfe angehender Lehrkräfte zu untersuchen und Facetten fachdidaktischen Wissens zu identifizieren, deren Berücksichtigung zur Qualität von Unterrichtsentwürfen beiträgt.

## Unterrichtsplanung als fachdidaktisches Problemlösen

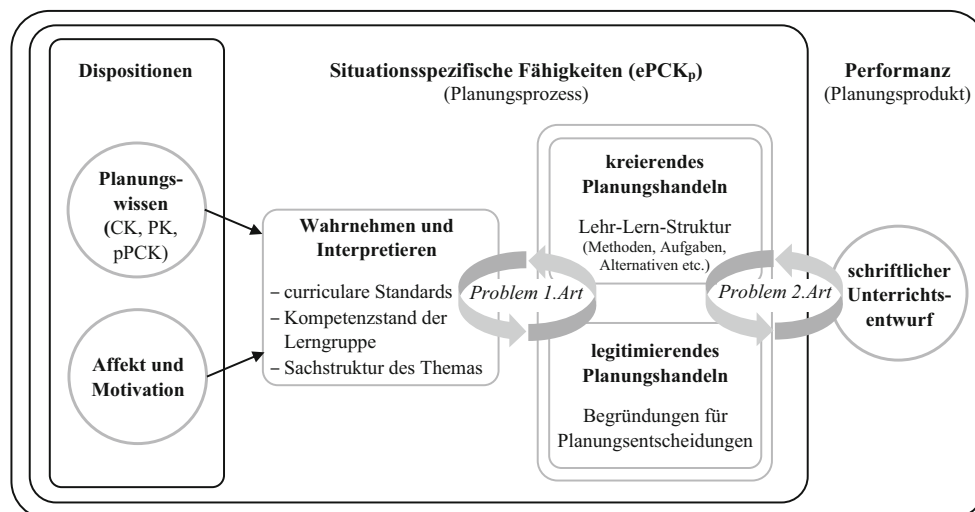
Im Prozess der Unterrichtsplanung werden verschiedene Wissensbestände integriert und in unterrichtliches Handeln überführt (König et al. 2017; Stender 2014; Westerman 1991). Hinsichtlich der Topologie des Professionswissen innerhalb der professionellen Kompetenz von Lehrkräften wird zwischen Fachwissen (*content knowledge*, CK), pädagogischem Wissen (*pedagogical knowledge*, PK) und fachdidaktischem Wissen (*pedagogical content knowledge*, PCK) unterschieden (Baumert und Kunter 2006; Shulman 1986). Diese Differenzierung wird inzwischen auch in der biologiedidaktischen Forschung durch empirische Befunde gestützt (Großschedl et al. 2015; Jüttner und Neuhaus 2013; Neuhaus 2021). Das PCK gilt u. a. deshalb als besonders wichtige Ressource (Carlson und Daehler 2019; Chan und Hume 2019), weil es zum einen positiv mit Unterrichtsqualität im Allgemeinen (z. B. Park et al. 2011) und ihren Teilaspekten wie der kognitiven Aktivierung

(Keller et al. 2017) oder der individuellen Unterstützung von Lernenden (Kunter et al. 2013) korreliert. Zum anderen gibt es Hinweise auf einen Zusammenhang zwischen dem PCK-Niveau von Lehrkräften und den Leistungen ihrer Lernenden (Förtsch et al. 2016; Jin et al. 2015; Mahler et al. 2017). PCK umfasst das Wissen und die Fähigkeiten zum Unterrichten eines „*particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes*“ (Gess-Newsome 2015, S. 36). Das kennzeichnende Charakteristikum von PCK ist die Integration und Verknüpfung verschiedener Facetten zu einer dynamischen, flexiblen und anwendungsbezogenen Wissensbasis, die im Pentagon-Modell fachdidaktischen Wissens zum Ausdruck kommt (Abb. 1; Park und Oliver 2008).

Demnach entwickelt sich PCK aus der Integration der fünf Wissensfacetten, die in den Denkprozessen während des Unterrichts (*reflection-in-action*) und in der Planung vor bzw. der Reflexion nach dem Unterricht (*reflection-on-action*) miteinander verknüpft werden. Die Qualität des PCK drückt sich dabei weniger in der Berücksichtigung der fünf distinkten Wissensfacetten, sondern in deren Vernetzung aus (z. B. Chan und Hume 2019; Park und Suh 2019; Reynolds und Park 2021).

Eine weitergehende Differenzierung des PCK nehmen Carlson und Daehler (2019) vor, die zwischen dem allgemeingültigen Kanonwissen innerhalb der – in diesem Fall – biologiedidaktischen Forschung und Lehre (*collective PCK*, cPCK), dem persönlichen PCK (*personal PCK*, pPCK) als „*reservoir of knowledge and skills that the teacher can draw upon during the practice of teaching*“ (ebd., S. 85) und dem sich in einer konkreten Handlungssituation manifestierenden PCK (*enacted PCK*, ePCK) unterscheiden. Alonzo et al. (2019) unterscheiden zusätzlich zwischen dem ePCK<sub>p</sub> (*planning*), ePCK<sub>t</sub> (*teaching*) und ePCK<sub>r</sub> (*reflecting*), da in allen drei Phasen unterrichtlichen Handelns unterschiedliches fachdidaktisches Wissen genutzt wird.

**Abb. 2** Heuristisches Kompetenzmodell der schriftlichen Planungskompetenz (nach Blömeke et al. 2015; König et al. 2017; vgl. Großmann und Krüger 2022)



Während Baumert und Kunter (2006) allgemeindidaktisches Planungswissen und fachübergreifende Prinzipien der Unterrichtsplanung als Facetten von PK definieren und Werner et al. (2017) Planungskompetenz als Synthese aus CK, PK und PCK verstehen, adressieren wir mit dem ePCK<sub>p</sub> also explizit die fachdidaktische Dimension als einen der wesentlichen Aspekte von Planungskompetenz (Abb. 2).

Das Verfassen eines Unterrichtsentwurfs ist eine kognitiv herausfordernde Aufgabe (König et al. 2021) und kann insofern als doppeltes Problemlösen aufgefasst werden, als es die Überwindung eines Defizits zwischen einem Ausgangszustand und einem angestrebten Endzustand mithilfe geeigneter Operationen darstellt (Funke 2003): Erstens muss unter Integration verschiedener Dispositionen ein Unterrichtsziel auf der Grundlage curricularer Standards definiert, der Kompetenzstand der Lerngruppe daraufhin analysiert und die Sachstruktur des Fachinhalts geklärt werden (Karlström und Hamza 2021). Auf dieser Grundlage wird eine Lehr-Lern-Struktur entwickelt, in der funktionale interdependente Planungsentscheidungen (Gassmann 2013) den Lernenden einen Kompetenzzuwachs ermöglichen sollen (Abb. 2: *Problem 1. Art*). Hierbei wird ePCK<sub>p</sub> als fachspezifische Fähigkeit genutzt. Zweitens muss die geplante Lehr-Lern-Struktur in einen strukturierten, sprachlich nachvollziehbaren Unterrichtsentwurf als Planungsprodukt überführt werden, in dem die Planungsentscheidungen begründet und die Denkprozesse des Planungshandelns transparent gemacht werden (Abb. 2: *Problem 2. Art*). Planungsqualität drückt sich demnach darin aus, dass beide Probleme angemessen gelöst werden. Das bedeutet, dass die Lehr-Lern-Struktur funktionale Planungsentscheidungen umfasst, die der spezifischen Lerngruppe Kompetenzzuwachs ermöglicht, wobei diese Entscheidungen schriftlich gut begründet und nachvollziehbar dargestellt werden. Letzteres fällt angehenden Lehrkräften oft besonders schwer (Zaragoza et al. 2021).

Die Abstimmung der Unterrichtsgestaltung [KISR] auf die Lerngruppenvoraussetzungen [KSU] gilt als besonders relevante fachdidaktische Fähigkeit (Förtsch et al. 2016; Jin et al. 2015; König et al. 2015; Park und Chen 2012; Reynolds und Park 2021). Dies spiegelt sich in dem in den Naturwissenschaftsdidaktiken etablierten *Modell der didaktischen Rekonstruktion* (vgl. Kattmann et al. 1997), allgemeindidaktischen Planungsmodellen (vgl. Weingarten 2019) sowie Planungsratgebern (vgl. Vogelsang und Riese 2017) wider. Demnach sollte auch der Fachinhalt [KSC] eine wichtige Rolle spielen, da beispielsweise die zentralen Ideen bzw. Konzepte eines Themas planungsleitend sein können, doch wird diese Wissensfacette selten mit anderen verknüpft (Park und Chen 2012). Dies gilt auch für eine formative Rückmeldung zum Lernzuwachs [KAs], die zwar vor allem im angloamerikanischen Raum (z. B. Wiggins und McTighe 2005) verbreitet ist, von angehenden Lehrkräften aber zumeist nicht berücksichtigt wird (Aydin und Boz 2013; Bravo und Cofré 2016; Chan und Yung 2018; Hanuscin et al. 2018; Mthethwa-Kunene et al. 2015; Park und Chen 2012). Stattdessen liegt der Fokus in der Sicherungsphase oft auf der Fixierung der erarbeiteten Begriffe und Konzepte (Beyer und Davis 2012; Weitzel und Blank 2020).

## Der Unterrichtsentwurf

Der Unterrichtsentwurf lässt sich als schriftliches Planungsprodukt definieren, in dem in einer bestimmten Kapitelfolge das Ergebnis des Planungsprozesses (Abb. 2) dargestellt wird. Zu den dabei zu berücksichtigenden planungsrelevanten Aspekten gehören die Festlegung von Unterrichtszielen, die Bereitstellung von strukturierten Inhalten, die Analyse der Lerngruppenvoraussetzungen, die Entwicklung von problemhaltigen Aufgaben inklusive der Arbeitsmate-

**Tab. 1** Aufbau eines kompetenzorientierten Unterrichtsentwurfs (SenBJF 2017, S. 54)

Bestandteil	Hinweise
Fachlich-inhaltlicher Schwerpunkt	Sachstruktur Aufgabenanalyse
Standards des Rahmenlehrplans	Angestrebte Kompetenzentwicklung Bezug zu Rahmenlehrplan
Längerfristige Kompetenzentwicklung	Orientierung an spezifischen Lernvoraussetzungen der Lerngruppe Inhaltsbezogener, kumulativ-vernetzter Aufbau der Kompetenzentwicklung in Bezug auf Unterrichtsreihe
Konkretisierung der Standards	Anwendung der Standards des Rahmenlehrplans auf den zu planenden Unterrichtsausschnitt Grundlage für Evaluation der Lernerfolgs
Kompetenzentwicklung der SuS	Diagnose der Heterogenität der SuS Analyse des Kompetenzstandes z. B. von drei exemplarischen SuS Bedarf an Binnendifferenzierung
Begründung der Lehr-Lern-Struktur	Didaktische Entscheidungen und methodische Umsetzung Begründungen für die Entscheidungen
Konkretisierung der Lehr-Lern-Prozesse	Verlaufsplan inkl. Impulsen, Aufgaben, erwarteten SuS-Leistungen
Anlagen	Arbeitsblätter, antizipiertes Tafelbild usw

rialien und eines tabellarischen Stundenverlaufsplans sowie die Rückmeldung über den Lernerfolg an die Lernenden (Karlström und Hamza 2021; Abb. 2: *Problem 1. Art*). Dazu gehören eine adressatengerechte Verschriftlichung dieser Aspekte und die fundierte Begründung der Planungsentscheidungen (Abb. 2: *Problem 2. Art*). Diese Aspekte sind deckungsgleich mit den in den Standards für die Lehrerbildung formulierten Zielvorstellungen zur Unterrichtsplanung (KMK 2019) und den formalen Vorgaben des Landes Berlin (Tab. 1), in dem die hier untersuchten Unterrichtsentwürfe verfasst worden sind (SenBJF 2017).

Insbesondere Noviz\*innen haben Schwierigkeiten, den Lerngegenstand zu analysieren (Sachstrukturanalyse), curriculare Standards an ein konkretes Thema anzupassen (Kompetenzen/Lernziel), den Kompetenzstand der Schüler\*innen fundiert zu analysieren (Lerngruppenanalyse), Lernzuwachs zu diagnostizieren und neben der methodisch-medialen Gestaltung der Lernprozesse (Oberflächenstruktur) in der didaktischen und methodischen Analyse auch die kognitiven Prozesse der Lernenden (Tiefenstruktur) zu berücksichtigen (Borko und Livingston 1989; Chizhik und Chizhik 2018; Haas 1998; Westerman 1991; Abb. 2: *Problem 1. Art*). Selbst wenn die kognitiven Prozesse berücksichtigt werden, werden sie oft nicht verschriftlicht, so dass schriftliche Unterrichtsentwürfe nur einen Teil der Unterrichtsplanung als mentales Probehandeln widerspiegeln (Morine-Dershimer 1979; Abb. 2: *Problem 2. Art*).

Mit Unterrichtsentwürfen befassen sich nur wenige aktuellere Studien im deutschsprachigen Raum, allerdings keine mit naturwissenschaftsidaktischer Ausrichtung. Gassmann (2013) identifizierte in einer qualitativen Interviewstudie zusätzlich zur Analyse von Unterrichtsentwürfen die Schwierigkeiten, die angehende Lehrkräfte ( $N=25$ ) unterschiedlicher Fächer beim Verfassen von Unterrichtsentwürfen hatten. Insbesondere die Vernetzung der verschiedenen zu berücksichtigenden Aspekte (Abb. 2: *Problem 1. Art*) und die Verschriftlichung und Begründung dessen, was intuitiv geplant wird (Abb. 2: *Problem 2. Art*), fiel den Studierenden schwer. König et al. (2015) wiesen in einer Längsschnittstudie für Referendar\*innen ( $N=106$ ) im Fach Deutsch zwischen dem Beginn und dem Ende des Vorbereitungsdienstes eine signifikante Zunahme didaktisch adaptiven Unterrichts nach, d. h. eine stärkere und angemessenere Abstimmung der Aufgabenstellungen auf den Kompetenzstand der Lerngruppe und darauf basierende Maßnahmen der Binnendifferenzierung (Abb. 2: *Problem 1. Art*). Weingarten (2019) untersuchte inhaltsanalytisch Staatsexamensentwürfe ( $N=180$ ) in den Fächern Mathematik, Deutsch, Politik und Musik. Demnach korreliert die Planungskompetenz von Referendar\*innen stark mit der Qualität durchgeführten Unterrichts in Form der Staatsexamensnote. Die Unterrichtsplanung kann somit als Prädiktor für die professionelle Durchführung von Unterricht gelten. Da die wenigen Studien zur Unterrichtsplanung vorwiegend deskriptiv ausgerichtet sind und keine Bewertungen der Qualität anstreben, bleibt weiterhin unklar, wann eine Unterrichtsstunde als gut geplant gelten kann (Stender 2014; Vogelsang und Riese 2017).

## Ziel und Forschungsfragen

Die vorliegende Studie zielt darauf ab, zur von Weingarten (2019) geforderten „kritische[n] Bestandsaufnahme der gegenwärtigen Planungspraxis von Lehrkräften“ (S. 28) beizutragen. Dazu wurden Unterrichtsentwürfe analysiert, die im Rahmen der Prüfung zum zweiten Staatsexamen am Ende des Vorbereitungsdienstes entwickelt worden sind und die daher einen Einblick in die Planungskompetenz am Ende der Lehrkräftebildung geben. Insbesondere in der zweiten Phase der Lehrkräftebildung bilden Unterrichtsentwürfe einen wichtigen Teil der Ausbildung, da sie für regelmäßig stattfindende Unterrichtsbesuche verfasst werden müssen und in manchen Bundesländern sogar Teil der Beurteilung sind (Strietholt und Terhart 2009). Die vorliegende Studie kann dazu beitragen, ePCK<sub>p</sub>-Aspekte zu identifizieren, die Biologie-Referendar\*innen bereits häufig und sinnvoll in ihren Entwürfen berücksichtigen, aber auch solche, die für das Schreiben guter Entwürfe unter Umständen noch stärker berücksichtigt werden sollten.

Als konzeptionellen Rahmen nutzen wir das Pentagon-Modell fachdidaktischen Wissens (Abb. 1; Park und Oliver 2008) und identifizieren Wissensfacetten, die angehende Lehrkräfte am Ende ihrer Ausbildung zur Unterrichtsplanung (Alonzo et al. 2019) anwenden. Im Sinne von Park (2019) geht es dabei insbesondere darum nachzuvollziehen, wie diese Wissensfacetten miteinander vernetzt werden. Über den Vergleich der identifizierten Wissensfacetten und ihrer Vernetzungen mit einer ganzheitlichen Beurteilung der Entwürfe als Qualitätsindikator gehen wir der Frage nach, welche fachdidaktischen Elemente in besonderer Weise die Qualität von Unterrichtsentwürfen auszeichnen. Dabei ist es möglich und erwartbar, dass es qualitative Unterschiede in der Berücksichtigung von ePCK<sub>p</sub>-Facetten und ihren Vernetzungen gibt, d.h. dass in zwei Entwürfen zwar jeweils eine Verbindung zwischen dem Kompetenzstand der Schüler\*innen und den Standards des Lehrplans hergestellt wird, allerdings unterschiedlich elaboriert. Drei Forschungsfragen werden zu diesem Zweck in Unterrichtsentwürfen von Biologie-Referendar\*innen am Ende ihrer Ausbildung untersucht:

- **Forschungsfrage 1:** Inwiefern unterscheidet sich die Häufigkeit von ePCK<sub>p</sub>-Facetten in elaboriert und wenig elaboriert eingeschätzten Entwürfen?
- **Forschungsfrage 2:** Inwiefern unterscheidet sich die Häufigkeit der Vernetzung von ePCK<sub>p</sub>-Facetten in elaboriert und wenig elaboriert eingeschätzten Entwürfen?
- **Forschungsfrage 3:** Inwiefern drückt sich in der Berücksichtigung von ePCK<sub>p</sub>-Facetten und in deren Vernetzung Planungsqualität aus?

## Methoden

### Beschreibung der Stichprobe

Die vorliegende Studie ist in der zweiten Phase der deutschen Lehrkräftebildung (Referendariat) verortet. Für die Analyse wurden Unterrichtsentwürfe genutzt (N=46), die im Rahmen der Prüfung zum zweiten Staatsexamen von Referendar\*innen im Fach Biologie entwickelt und vom zuständigen Kultusministerium im Land Berlin anonymisiert zur Verfügung gestellt worden sind. Aufgrund der Anonymisierung sind keine weiteren Personenmerkmale bekannt, auch nicht die Note der Examenslehrprobe auf der Basis des Unterrichtsentwurfs für die gehaltene Stunde. Da die

**Tab. 2** Stichprobe

	Fachwissen	Erkenntnisgewinnung	Kommunikation	Bewertung	Gesamt
Sek I	4	4	4	2	14
Sek II	3	12	10	7	32
Gesamt	7	16	14	9	46

explizite Zustimmung der Verfasser\*innen zur Verwendung der Entwürfe zu Forschungszwecken nicht vorliegt, werden in diesem Artikel keine direkten Zitate aus den Staatsexamensentwürfen verwendet.

In der Stichprobe sind alle vier Kompetenzbereiche (KMK 2005) und die beiden Sekundarstufen vertreten (Tab. 2).

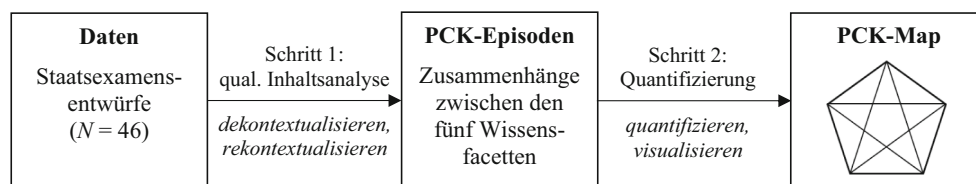
### Analyse des fachdidaktischen Wissens (ePCK<sub>p</sub>): PCK map approach

Zur Analyse des fachdidaktischen Wissens (ePCK<sub>p</sub>), das sich in den Staatsexamensentwürfen manifestiert, wurde der sogenannte *PCK map approach* (Abb. 3; Park und Suh 2019) verwendet, der in die deutschsprachige Naturwissenschaftsdidaktik erstmalig von Pohlmann (2019) eingeführt worden ist. Dabei werden über eine qualitative Inhaltsanalyse – in diesem Fall nach Mayring (2015) – *PCK-Maps* erstellt, die das ePCK<sub>p</sub> visualisieren und einer qualitativen Analyse zugänglich machen.

In den Staatsexamensentwürfen wurden diejenigen Passagen identifiziert, in denen auf Aspekte Bezug genommen wird, die sich einer der fünf Facetten fachdidaktischen Wissens des Pentagon-Modells (Abb. 1; Park und Oliver 2008) zuordnen lassen. Zu diesem Zweck wurde vom Erstautor und einer geschulten studentischen Hilfskraft eine qualitative Inhaltsanalyse (Mayring 2015) mittels eines deduktiv-induktiv entwickelten und zuvor erprobten Kategoriensystems durchgeführt (Tab. 3, Großmann und Krüger 2020) und die Staatsexamensentwürfe mit der Software MAXQDA 2020 (VERBI Software 2019) kodiert.

Als Analyseeinheit wurde der schriftliche Unterrichtsentwurf ohne Anhang kodiert, der Anhang wurde nur herangezogen, wenn Passagen in der Analyseeinheit nicht selbsterklärend waren (Schreier 2014). Zur Bestimmung der Intra- und Inter-Rater-Übereinstimmung hat der Erstautor 16 Entwürfe in einem Abstand von 14 Tagen jeweils zweimal kodiert. Als Maß für die Intra- und die Interrater-Übereinstimmung wur-

**Abb. 3** Analyseschritte des *PCK map approaches* (nach Park 2019)



**Tab. 3** Kategoriensystem zur Analyse des fachdidaktischen Wissens in Unterrichtsentwürfen (vgl. Großmann und Krüger 2020)

Kategorie	Beschreibung
<i>Wissensfacette 1 – Wissen über den kontextuellen Rahmen (OTS)</i>	
1 Angestrebter Kompetenzzuwachses	Formulierung eines Ziels für die Stunde basierend auf einem curricularen Standard
2 Indikatoren	Indikatoren zum Nachweis eines Kompetenzzuwachses
3 Unterrichtssequenz (Beschreibung)	Tabellarische Sequenzplanung zum langfristigen Kompetenzerwerb
4 Unterrichtssequenz (Begründung)	Begründung der Anordnung der Lernschritte in der Unterrichtssequenz
5 Unterrichtsprinzipien	Nennung von didaktischen Prinzipien (z. B. Problemorientierung)
<i>Wissensfacette 2 – Wissen über das Fachcurriculum (KSC)</i>	
6 Themenauswahl (Beschreibung)	Stellung des Stundenthemas innerhalb des fachlichen Themenkomplexes
7 Themenauswahl (Begründung)	Begründung der Auswahl des Themas (z. B. Verweis auf Relevanzkriterien)
8 Sachstrukturanalyse	Analyse des Fachinhalts aus fachwissenschaftlicher Perspektive
9 Didaktische Reduktion	Beschreibung und Begründung von Reduktionsentscheidungen
<i>Wissensfacette 3 – Wissen über die Lerngruppe (KSU)</i>	
10 Kompetenzstand	Analyse des aktuellen Kompetenzstandes der Lerngruppe
11 Schülervorstellungen	Beschreibung von Alltagsvorstellungen der SuS zum Stundenthema
12 Lernschwierigkeiten	Beschreibung von Aspekten, die den SuS Schwierigkeiten bereiten können
13 Methodenkenntnisse	Beschreibung der methodischen Voraussetzungen (z. B. Sozialformen)
14 Soziales	Beschreibung der sozialen Voraussetzungen der Lerngruppe (z. B. Konflikte)
15 Motivation	Beschreibung der motivationalen Voraussetzungen (z. B. Mitarbeit)
<i>Wissensfacette 4 – Wissen über fachspezifische Instruktionsstrategien (KISR)</i>	
16 Phasenstruktur (Beschreibung)	Beschreibung der Abfolge der einzelnen Unterrichtsschritte
17 Phasenstruktur (Begründung)	Begründung der Abfolge der einzelnen Unterrichtsschritte
18 Aufgabenstellungen (Beschreibung)	Beschreibung des Aufbaus und des Einsatzes der Aufgabenstellungen
19 Aufgabenstellungen (Begründung)	Begründung des Aufbaus und des Einsatzes der Aufgabenstellungen
20 Methoden (Beschreibung)	Beschreibung des Einsatzes der Methoden
21 Methoden (Begründung)	Begründung des Einsatzes der Methoden
22 Medien (Beschreibung)	Beschreibung der Gestaltung und des Einsatzes der Medien (z. B. Arbeitsblätter)
23 Medien (Begründung)	Begründung der Gestaltung und des Einsatzes der Medien (z. B. Arbeitsblätter)
24 Impulse (Beschreibung)	Nennung ausformulierter Impulse zur Steuerung des Lernprozesses
25 Impulse (Begründung)	Begründung der Formulierung der Impulse und ihrer Steuerungsfunktion
26 Binnendifferenzierung (Beschreibung)	Beschreibung der Gestaltung und des Einsatzes der Binnendifferenzierungsmaßnahmen
27 Binnendifferenzierung (Begründung)	Begründung der Gestaltung und des Einsatzes der Binnendifferenzierungsmaßnahmen
28 Antizipation von Schülerverhalten	Beschreibung möglicher SuS-Äußerungen und mögliche Bruchstellen
29 Alternativplanung	Beschreibung alternativer Handlungsstrategien als Reaktion auf mögliche Bruchstellen im Lernprozess
<i>Wissensfacette 5 – Wissen über Beurteilung (KAs)</i>	
30 Schülerfeedback (Beschreibung)	Beschreibung, wie die SuS eine Rückmeldung zum Lernprodukt oder zum Lernprozess geben
31 Schülerfeedback (Begründung)	Begründung, wie die SuS eine Rückmeldung zum Lernprodukt oder zum Lernprozess geben
32 Lehrerfeedback (Beschreibung)	Beschreibung, wie die Lehrkraft den SuS eine Rückmeldung zum Lernprodukt oder zum Lernprozess gibt
33 Lehrerfeedback (Begründung)	Begründung, wie die Lehrkraft den SuS eine Rückmeldung zum Lernprodukt oder zum Lernprozess gibt

de Cohen's Kappa ( $\kappa$ ) berechnet (Wirtz und Caspar 2002). Mittels konsensuellen kodierens wurden unterschiedlich kodierte Stellen zwischen Erst- und Zweitkodierer diskutiert und Konsens erzielt (Kuckartz 2016, S. 211f.). Die Berechnung der Intrarater- ( $M_\kappa = 0,76$ ) und Interrater-Übereinstimmungswerte ( $M_\kappa = 0,67$ ) weist auf eine hinreichende Stabilität des Auswertungsverfahrens und eine zufriedenstellende Auswertungsobjektivität hin (Wirtz und Caspar 2002). Auf diese Weise sind in allen Entwürfen zunächst isolier-

te Facetten fachdidaktischen Wissens identifiziert worden (Abb. 2: *Dekontextualisierung*).

Der *PCK map approach* geht davon aus, dass fachdidaktisches Wissen als „*integration of two or more components*“ (Park und Chen 2012, S. 928) verstanden werden muss. Daher wurden im nächsten Schritt diejenigen Sinnabschnitte in den Entwürfen identifiziert, in denen Codes aus mindestens zwei der fünf Wissensfacetten gemeinsam vergeben worden sind und somit eine Interaktion zwischen den Wis-

**Tab. 4** Beispiel einer PCK-Episode (nach Park und Chen 2012; Reynolds und Park 2021)

PCK-Episode	ePCK <sub>p</sub> -Facette	PCK-Map
[...] Da die Chromosomen in der Metaphase als dicht gepackte Körperchen vorliegen, wird die Gefahr reduziert, dass die DNA-Fäden verkleben [KSC, Kat. 8]. In den vergangenen Stunden hatten die SuS große Probleme, den Ablauf der Mitose zu verstehen und zu erklären, offenbar weil das Thema zu abstrakt ist [KSU, Kat. 12]. Ich habe mich daher für einen spielerischen Einstieg entschieden, in dem die SuS mithilfe von Spielkarten den Verlauf der Mitose selbst durchlaufen sollen [...] [KISR, Kat. 20]. Auf diese Weise werden sowohl die Begriffe (z. B. Metaphase) als auch der Prozess (z. B. Anordnung in Äquatorialebene) wiederholt und anschaulich gemacht [KISR, Kat. 21]. Insbesondere die Begriffe „Ein-/Zwei-Chromatid-Chromosom“ waren problematisch [KSU, Kat. 10], ich nutze daher in dieser Stunde auf der Anleitung für das Spiel im Einstieg Kattmann (2015) folgend die Begriffe „Einzel-“ bzw. „Doppelchromosom“ [KISR, Kat. 22]	<p>OTS – KSC ✓ KSU ✓ KISR ✓ KAs –</p>	

sensfacetten besteht (Abb. 3: *Rekontextualisierung*). Derartige Stellen werden *PCK-Episoden* genannt (Park und Chen 2012; Park und Suh 2019). PCK-Episoden sind von variabler Länge, und fokussieren jeweils einen Aspekt (z. B. die Beschreibung bzw. Begründung einer Methode im Einstieg). Sobald ein weiterer Aspekt angesprochen wird (z. B. die Beschreibung bzw. Begründung eines Impulses, Mediums im Einstieg), wird dies als neue PCK-Episode aufgefasst. Tab. 4 zeigt ein Beispiel für eine PCK-Episode aus einer Sachstrukturanalyse und die visualisierende Umsetzung als PCK-Map.

Im letzten Schritt (Abb. 3) wurden sämtliche PCK-Episoden eines Unterrichtsentwurfs summiert. Auf diese Weise wurden PCK-Maps entwickelt, die das in einem gesamten Staatsexamensentwurf identifizierbare ePCK<sub>p</sub> der angehenden Biologie-Lehrkräfte visualisieren. Daran ist der Vernetzungsgrad in Form der Anzahl an Verbindungen abzulesen: Je mehr Verbindungen zwischen den fünf Wissensfacetten einer PCK-Map und je häufiger diese bedient werden, desto elaborierter ist das ePCK<sub>p</sub> der Lehrkraft (Park und Chen 2012).

### Einschätzung der Planungsqualität

Zur Einschätzung der Planungsqualität wurden alle 46 Staatsexamensentwürfe von einem Biologie-Fachseminarleiter, der in der zweiten Phase der Lehrkräftebildung Referendar\*innen 21 Jahre lang ausgebildet und als Mitglied von Prüfungskommissionen Staatsexamensprüfungen durchgeführt hat, gelesen. Auf einer Notenskala von 1 (sehr gut) bis 5 (mangelhaft) wurde eingeschätzt, inwiefern die Planungsentscheidungen funktional zum Erreichen des angestrebten Unterrichtsziels sind (Abb. 2: *Problem 1. Art*) und inwiefern die Planungsentscheidungen nachvollziehbar dargestellt und fundiert begründet wurden (Abb. 2: *Problem 2. Art*). Alle 46 Staatsexamensentwürfe wurden in mehreren Interview-Sitzungen besprochen, in denen der Fachseminarleiter seine Urteilsfindung begründete. Dieser

Schritt wurde nötig, da die zuständige Senatsverwaltung im Interesse der beteiligten Lehrkräfte zwar die anonymisierten Staatsexamensentwürfe, nicht aber die Noten der Staatsexamensprüfung preisgeben konnte. Vier der Entwürfe wurden zudem von einem zweiten Fachseminarleiter bewertet, um die intersubjektive Nachvollziehbarkeit (Göhner und Krell 2020; Mayring 2015) der Einschätzungen zu gewährleisten. Beide Fachseminarleiter waren zum Zeitpunkt der Untersuchung im Land Berlin tätig und keiner von beiden stand in Kooperation mit der Universität. Der theoretische Hintergrund der Untersuchung (Abb. 1: Pentagon-Modell fachdidaktischen Wissens; Park und Oliver 2008) war beiden Fachseminarleitern nicht bekannt, so dass ihre Einschätzung davon nicht beeinflusst war.

### Auswertung der Daten in Bezug auf die Forschungsfragen

Zur Beantwortung der drei Forschungsfragen wurden die 46 Staatsexamensentwürfe anhand der Qualitätsbeurteilung in ein oberes Quartil Q<sub>.75</sub> (Note 1) und ein unteres Q<sub>.25</sub> (Noten 4–5) unterteilt (Reynolds und Park 2020). Für sämtliche Berechnungen wurden nicht die absoluten Werte der Kodierungen genutzt, sondern diese in Relation zur Wortanzahl der Entwürfe (Anzahl Kodierung/Wortanzahl \* 3407) gestellt, da der Umfang der 46 Entwürfe sehr breit streut (M = 3407 Wörter; Min = 1485 Wörter; Max = 7764 Wörter). Auf diese Weise relativieren wir den Einfluss des Umfangs der Entwürfe, da in der qualitativen Inhaltsanalyse alle Äußerungen kodiert wurden, auch wenn sie sich in längeren Entwürfen wiederholten. Dass ein wenig konziser, eher repetitiver Schreibstil bereits einen Indikator für Planungsqualität darstellt, erscheint nicht plausibel. Durch die Normierung der Wortzahl wird dieser Effekt für die statistische Auswertung relativiert.

Zur Beantwortung der Forschungsfragen 1 und 2 wurden Mann-Whitney-U-Tests einschließlich Alphafehler-Korrektur (Dunn-Bonferroni-Posthoc-Test) durchgeführt, um für



**Tab. 5** Exemplarische PCK-Maps aus  $Q_{.75}$  und  $Q_{.25}$  der eingeschätzten Planungsqualität. Vier Extremfälle, d. h. der am stärksten bzw. am schwächsten vernetzte Entwurf des  $Q_{.75}$  und des  $Q_{.25}$

Planungsqualität	PCK-Maps	
	Stark vernetzt	Schwach vernetzt
<b>Hoch (<math>Q_{.75}</math>)</b> ( $n = 11$ )	<p><b>a</b></p> <p>Acht Fälle, hier: Entwurf 42</p>	<p><b>b</b></p> <p>Drei Fälle, hier: Entwurf 38</p>
<b>Gering (<math>Q_{.25}</math>)</b> ( $n = 11$ )	<p><b>c</b></p> <p>Zwei Fälle, hier: Entwurf 33</p>	<p><b>d</b></p> <p>Neun Fälle, hier: Entwurf 22</p>

die fünf ePCK<sub>p</sub>-Elemente (Forschungsfrage 1) und die zehn möglichen ePCK<sub>p</sub>-Verbindungen (Forschungsfrage 2) laut Pentagon-Modell (Abb. 1; Park und Oliver 2008) jeweils das obere und untere Quartil (jeweils  $n = 11$ ) miteinander zu vergleichen. Zusätzlich wurde eine Korrelationsanalyse mit der gesamten Stichprobe ( $N = 46$ ) durchgeführt, um den Zusammenhang zwischen der Häufigkeit von ePCK<sub>p</sub>-Elementen und -Verbindungen auch auf der Basis der Gesamtdaten zu untersuchen. Zur Absicherung der Güte der Einschätzungen des Fachseminarleiters wurden diejenigen vier Entwürfe von einem zweiten Fachseminarleiter analysiert, die jeweils am stärksten bzw. am schwächsten im oberen bzw. im unteren Quartil vernetzt waren (*purposeful sampling*; Patton 1990). Die Einschätzungen der beiden Fachseminarleiter korrelierten stark ( $r = 0,71$ ). Zur Beantwortung der Forschungsfrage 3 wurden ein typischer Entwurf mit schwach vernetztem ePCK<sub>p</sub> aus dem oberen Quartil der Planungsqualität (Tab. 5b) mit einem Entwurf mit stark vernetztem ePCK<sub>p</sub> aus dem unteren Quartil der Planungsqualität (Tab. 5c) in einer kontrastiven Fallanalyse verglichen. Da davon auszugehen ist, dass gute Entwürfe stark vernetztes ePCK<sub>p</sub> bzw. schwache Entwürfe schwach vernetztes ePCK<sub>p</sub> aufweisen, bieten derartige Fälle wenig Erkenntnisgewinn in Bezug auf die Forschungsfrage. Aus der Analyse

eines Entwurfs mit schwach vernetztem ePCK<sub>p</sub>, der als besonders gelungen bewertet worden ist, bzw. eines Entwurfs mit stark vernetztem ePCK<sub>p</sub>, der als wenig gelungen bewertet worden ist, lässt sich hingegen ableiten, welche Aspekte eines Unterrichtsentwurfs über die bloße Häufigkeit hinaus einen Einfluss auf die Planungsqualität haben könnten. Dabei werden Besonderheiten in der Berücksichtigung der einzelnen Wissensfacetten oder ihrer Verknüpfungen miteinander anhand von Paraphrasen dargestellt und Aspekte identifiziert, die eine gelungene von einer optimierbaren Unterrichtsplanung unterscheiden. Für diese Analyse wurden auch die Einschätzungen des Fachseminarleiters aus den Interview-Sitzungen einbezogen.

## Ergebnisse

### Forschungsfrage 1: Planungsqualität und Einzelaspekte des ePCK<sub>p</sub>

Insgesamt wurden in den Entwürfen des  $Q_{.75}$  für alle fünf Wissensfacetten des ePCK<sub>p</sub> mehr Aussagen kodiert als in den Entwürfen des  $Q_{.25}$ ; dies gilt ebenso für die Summe der Kodierungen pro Entwurf (Tab. 6a).

**Tab. 6** Deskriptive Statistik der Kodierung der fünf ePCK<sub>p</sub>-Facetten sowie ihrer Verbindungen und Ergebnisse des Mann-Whitney-*U*-Tests zum Vergleich der Entwürfe des Q<sub>.75</sub> und des Q<sub>.25</sub>

<b>a) Berücksichtigung der ePCK<sub>p</sub>-Facetten</b>									
ePCK <sub>p</sub> -Facetten	Q <sub>.75</sub> (n = 11)		Q <sub>.25</sub> (n = 11)		Unterschied				
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>	<i>U</i>	<i>z</i>	<i>p</i>	<i>r</i>	
OTS	24	11	16	9	59,000	-0,098	0,949	-0,02	
KSC	14	9	8	4	60,000	-0,033	1,000	-0,01	
KSU	24	14	16	14	53,000	-0,492	0,652	-0,10	
KISR	49	22	21	11	36,000	-1,609	0,116	-0,34	
KAs	3	5	1	2	46,000	-0,963	0,365	-0,20	
Summe	121	42	64	23	43,000	-1,149	0,270	-0,24	
<b>b) Berücksichtigung der ePCK<sub>p</sub>-Verbindungen</b>									
ePCK <sub>p</sub> -Verbindungen	Q <sub>.75</sub> (n = 11)		Q <sub>.25</sub> (n = 11)		Unterschied				
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>	<i>U</i>	<i>z</i>	<i>p</i>	<i>r</i>	
OTS-KSC	2	3	0	1	36,000	-1,635	0,116	-0,35	
OTS-KSU	8	6	2	3	22,500	-2,496	0,010	-0,53	
OTS-KISR	3	6	1	1	35,000	-1,684	0,101	-0,36	
OTS-KAs	0	0	0	0	55,000	-1,000	0,748	-0,21	
KSC-KSU	2	4	0	2	42,500	-1,241	0,243	-0,26	
KSC-KISR	1	2	0	2	44,000	-1,122	0,300	-0,24	
KSC-KAs	0	0	0	0	-	-	-	-	
KSU-KISR	7	5	2	5	41,000	-1,281	0,217	-0,27	
KSU-KAs	0	1	0	0	47,000	-1,072	0,401	-0,23	
KISR-KAs	0	1	0	1	50,000	-0,697	0,519	-0,15	
Summe	25	21	9	7	27,000	-2,200	0,028	-0,47	

Q<sub>.75</sub> Entwürfe mit hoher Planungsqualität, Q<sub>.25</sub> Entwürfe mit geringer Planungsqualität, *Mdn* Median, *IQR* Interquartilsabstand

Besonders häufig wird zu Aspekten der Kompetenzorientierung (OTS), zu den Lerngruppenvoraussetzungen (KSU) und vor allem auf Instruktionsstrategien (KISR) Bezug genommen. Auf der Ebene der Subkategorien sind diesbezüglich einige Befunde bemerkenswert: Hinsichtlich der Lerngruppenvoraussetzungen wird überwiegend auf Vorwissen (Tab. 3, Kat. 10) und methodische Fähigkeiten (ebd., Kat. 13) sowie auf soziale (ebd., Kat. 15) und motivationale Aspekte (ebd., Kat. 16) hingewiesen. Auf Schülervorstellungen (ebd., Kat. 11) wird hingegen nur in zwei Entwürfen der gesamten Stichprobe jeweils einmal Bezug genommen, ohne dass es sich dabei jedoch um eine Diagnose für die konkrete Lerngruppe handeln würde und ohne dass daraus konkrete Planungsentscheidungen abgeleitet werden. Hinsichtlich der Instruktionsstrategien fällt auf, dass in 45 der 46 Entwürfe Binnendifferenzierungsmaßnahmen (ebd., Kat. 26, 27) beschrieben und z. T. begründet wurden. Demgegenüber wurden andere wichtige Aspekte wie Aufgabenstellungen nur in der Hälfte der Entwürfe begründet. In beiden Gruppen wird selten Wissen über Beurteilung (KAs, z. B. zur Rückmeldung über den Kompetenzzuwachs der Schüler\*innen) genutzt. Dies zeigt sich insbesondere in Sicherungsphasen, in denen fast in der gesamten Stichprobe Unterrichtsgespräche geplant werden, in denen ein\*e Schüler\*in ein Ergebnis präsentiert. Wie diagnostiziert werden

kann, ob die Lerngruppe tatsächlich Kompetenzzuwachs erfahren hat, bleibt oft unklar.

## Forschungsfrage 2: Planungsqualität und Vernetzung des ePCK<sub>p</sub>

Insgesamt wurden in den Entwürfen des Q<sub>.75</sub> für alle möglichen Verbindungen zwischen den Wissensfacetten des ePCK<sub>p</sub> mehr PCK-Episoden identifiziert als in den Entwürfen des Q<sub>.25</sub>; dies gilt ebenso für die Summe der Vernetzungen pro Entwurf (Tab. 6b). Eine Ausnahme stellt die Verbindung zwischen dem Wissen über das Fachcurriculum [KSC] und dem Wissen über Beurteilung [KAs], die in keinem Entwurf dieser Stichprobe gefunden wurde.

Am häufigsten werden insbesondere in den als elaboriert eingeschätzten Entwürfen (Q<sub>.75</sub>) Verbindungen zwischen Aspekten der Kompetenzorientierung und den Lerngruppenvoraussetzungen [OTS-KSU] sowie zwischen den Lerngruppenvoraussetzungen und der Gestaltung der Unterrichtsstunde [KSU-KISR] hergestellt. Mann-Whitney-*U*-Tests zeigen, dass nur die Verbindung zwischen der Kompetenzentwicklung [OTS] und den Lerngruppenvoraussetzungen [KSU] in den Entwürfen des Q<sub>.75</sub> mit großem Effekt (Tab. 6b) häufiger als in den Entwürfen des Q<sub>.25</sub> berücksichtigt wurde. Der Unterschied für die Summe aller

Verbindungen pro Entwurf ist mit mittlerem Effekt in den Entwürfen des  $Q_{.75}$  signifikant höher als in den Entwürfen des  $Q_{.25}$ .

Diese Befunde ließen sich durch eine Korrelationsanalyse unter Berücksichtigung der gesamte Stichprobe ( $N=46$ ) sowohl für die Verbindung OTS-KSU ( $r=-0,35$ ; mittlerer Effekt) und für die Summe der Verbindungen ( $r=-0,39$ ) bestätigen. Unter Berücksichtigung einer Alphafehler-Korrektur über den Dunn-Bonferroni-Posthoc-Test sind die beiden signifikanten Ergebnisse des Mann-Whitney- $U$ -Tests unter dem adjustierten Signifikanzniveau  $p^*$  nicht mehr als signifikant anzusehen.

Vergleicht man die PCK-Maps elaboriert und der wenig elaboriert eingeschätzte Entwürfe hinsichtlich ihres Vernetzungsgrades (Tab. 5) miteinander, fällt auf, dass Unterrichtsentwürfe mit stark vernetztem  $ePCK_p$  von geringer Planungsqualität sein können und umgekehrt. Auffällig ist, dass selbst in den am stärksten vernetzten Entwürfen beider Quartile (Tab. 5a,c) die Wissensfacetten KSC und KAs relativ schwach und die Trias aus OTS-KSU-KISR relativ stark vernetzt ist.

### Forschungsfrage 3: Qualitative Unterschiede im $ePCK_p$

Im Folgenden werden zwei Fälle analysiert, um einen qualitativen Einblick in die Planungskompetenz der Biologie-Referendar\*innen zu gewinnen. Dabei wird das Ziel verfolgt, Unterschiede in der Vernetztheit des  $ePCK_p$  anhand ausgewählter Aspekte qualitativ zu beschreiben und Unterschiede in der Planungsqualität zwischen elaboriert und wenig elaboriert eingeschätzten Unterrichtsentwürfen zu identifizieren. Zu diesem Zweck wurden zwei erwartungswidrige Fälle ausgewählt (vgl. Tab. 5b,c; Patton 1990), die nicht der Annahme entsprechen, dass hohe Planungsqualität und Häufigkeit der  $ePCK_p$ -Verbindungen positiv korrelieren (Reynolds und Park 2021). Beide Staatsexamensentwürfe adressieren die Förderung von Bewertungskompetenz in der Sekundarstufe II.

#### Fall 1: Hohe Qualität – schwach vernetztes $ePCK_p$ (Entwurf 38)

Der angestrebte Kompetenzzuwachs in Entwurf 38 (Tab. 5b) besteht darin, dass die Lernenden eines 11. Jahrgangs (Grundkurs) im Themenfeld Ökologie die Bedeutung des Graskarpfens im Ökosystem See reflektieren und mögliche Handlungsoptionen zum Umgang mit dieser invasiven Art im Rahmen einer Pro-Contra-Debatte diskutieren.

Die Stunde fügt sich in ein langfristiges Konzept zur Kompetenzentwicklung im Bereich der Bewertungskompetenz ein, demzufolge die SuS verschiedene Positionen anhand objektiver Kriterien kritisch reflektieren können sol-

len, um eine angemessene Handlungsoption zu wählen, die biologische, wirtschaftliche und gesellschaftliche Gesichtspunkte berücksichtigt (Tab. 3: Kat. 4). Dies spiegelt sich in der methodischen Entscheidung einer Pro-Contra-Debatte wider, in der ein Umweltpolitiker, ein Zoologe, ein Angler und ein Inhaber eines Strandbades die verschiedenen Perspektiven zum Umgang mit dem Graskarpfen vertreten (Tab. 3: Kat. 20). Die methodische Entscheidung ergibt sich hier insbesondere aus einer Diagnose des Kompetenzstandes:

*Die Lernenden können die Relevanz und Überzeugungskraft ihrer Argumente noch nicht gewichten und in ihre persönliche Entscheidungsfindung nur ihre eigene und keine anderen Perspektiven miteinbeziehen [KSU, Kat. 10]. Um das Reflektieren von möglichen Handlungsoptionen zu üben [OTS, Kat. 1], wird daher die Pro-Kontra-Debatte als Methode gewählt. (Zitat paraphrasiert)*

Das Unterrichtskonzept erscheint tragfähig, da es sich auf eine fundierte und differenzierte Lerngruppenanalyse stützt und auf ein präzise formuliertes Unterrichtsziel abgestimmt ist. Viele Planungsentscheidungen, die die Lehr-Lern-Struktur der daraufhin geplanten Stunde betreffen, können nur vor diesem Hintergrund verstanden und als gut begründet betrachtet werden. So werden beinahe sämtliche Entscheidungen bezüglich der Phasenstruktur und der medialen und methodischen Umsetzung [KISR] begründet, allerdings ohne einen expliziten Bezug zu den anderen Wissensfacetten herzustellen. So erhalten beispielsweise in der Sicherungsphase:

*diejenigen Schüler\*innen, die nicht aktiv an der Debatte teilnehmen, den Hörauftrag, Argumente der Debattierenden zu notieren und ihre Überzeugungskraft sowie Fachsprachlichkeit einzuschätzen [KISR, Kat. 18]. (Zitat paraphrasiert)*

Diese Entscheidung ist funktional zur Kompetenzförderung, wird aber nicht erneut explizit damit begründet und ergibt sich stattdessen aus den vorherigen Ausführungen. Bemerkenswert ist hingegen der Rückgriff aus dem angestrebten Kompetenzzuwachs in der Sicherungsphase.

*Hier werden die Schüler\*innen durch einen Impuls gebeten, auf der Grundlage ihrer Mitschriften zum Hörauftrag die Pro-Contra-Debatte zu reflektieren [KISR, Kat. 24]. Dazu sollen sie Stellung beziehen, welche der Argumente in der Pro-Contra-Debatte sie als überzeugend einschätzen, und dies begründen [KAs, Kat. 30]. Anhand des Feedbacks, das die Schüler\*innen geben, kann die Lehrkraft diagnostizieren, ob das angestrebte Unterrichtsziel erreicht worden ist [OTS, Kat. 1 + KISR, Kat. 21]. (Zitat paraphrasiert)*

Hier wird die Sicherungsphase über eine Verknüpfung verschiedener Wissensfacetten legitimiert, so dass ein Kompetenzzuwachs der Lernenden nachweisbar wird. Können die Lernenden die Argumente der Pro-Contra-Debatte hinsichtlich ihrer Überzeugungskraft nicht begründet einschätzen, kann diese Diagnose als Anlass für die Planung der darauffolgenden Stunde genutzt werden.

Zusammenfassend zeigt sich, dass zentrale Planungsentscheidungen wie die Integration des Unterrichtsvorhabens in ein langfristiges Konzept zur Kompetenzförderung unter Einbeziehung des Kompetenzstandes der Lernenden sowie die Prüfung des Kompetenzzuwachses in der Sicherungsphase mit voraussichtlich geeigneten methodischen Verfahren über vernetzte Wissensfacetten begründet werden, eine Vielzahl der Planungsentscheidungen zur methodischen Umsetzung aber nicht wiederholt unter Hinzuziehung anderer Wissensfacetten begründet wird. In diesem Entwurf werden also unterschiedliche Planungsentscheidungen getroffen, die in sich schlüssig sind und sich beispielsweise unmittelbar aus der Analyse des Kompetenzstandes der Schüler\*innen ergeben. Die Passung dieser Entscheidungen zueinander wird sprachlich aber nur selten durch Verknüpfungen fachdidaktischer Wissensfacetten transparent gemacht.

### Fall 2: Geringe Qualität – stark vernetztes ePCK, (Entwurf 33)

Der angestrebte Kompetenzzuwachs in Entwurf 33 (Tab. 5c) besteht darin, dass die Lernenden eines 12. Jahrgangs (Grundkurs) im Themenfeld Genetik die Chancen und Risiken des gentechnischen Verfahrens *Gene Drive* in der Malaria-Bekämpfung in Form einer Podiumsdiskussion reflektieren und begründet einen eigenen ethischen Standpunkt entwickeln.

Statt eines langfristigen Konzepts zur Förderung von Bewertungskompetenz enthält die Sequenzplanung eine Aufzählung genetischer Fachinhalte. Zudem fehlt eine Lerngruppenanalyse, in der die Lernenden differenziert und präzise hinsichtlich ihrer bereits vorhandenen Fähigkeiten, einen eigenen ethischen Standpunkt zu entwickeln, analysiert werden. Stattdessen wird die Stunde eher kurzfristig in Bezug auf die vorhergehende Stunde eingebettet [OTS, Kat. 3]:

*In einem Text sollten die Schüler\*innen als Hausaufgabe Argumente für und gegen den Einsatz des gentechnischen Verfahrens erarbeiten. Diese Argumente werden in der Podiumsdiskussion genutzt. Diese Methode wurde gewählt, weil die Schüler\*innen motiviert sind, im Biologieunterricht zu diskutieren [KSU, Kat. 15] und weil dieses Unterrichtsformat die*

*Schüler\*innen motivieren und aktivieren kann [KISR, Kat. 21]. (Zitat paraphrasiert)*

Im Entwurf wäre eine Lerngruppenanalyse notwendig, um die Angemessenheit des Settings adäquat einschätzen zu können. Es werden wiederholt Aspekte miteinander verknüpft, die die Planungsentscheidungen allerdings nicht hinreichend legitimieren:

*Mit dem Schwerpunkt der Stunde im Kompetenzbereich Reflexion [OTS, Kat. 1] werden die Schüler\*innen in Form eines Lehrervortrages zunächst in das Setting einer Konferenz der Vereinten Nationen eingeführt und mit der Problemfrage konfrontiert [KISR, Kat. 16], um einen motivierenden Lebensweltbezug zu bieten. Die Problemfrage ist der erste Schritt des naturwissenschaftlichen Erkenntnisweges und ermöglicht eine differenzierte Betrachtung des zu untersuchenden Problems [OTS, Kat. 1]. (Zitat paraphrasiert)*

Hier wird jedoch auf einen zweiten Kompetenzbereich (Erkenntnisgewinnung) Bezug genommen, der in dieser Stunde nicht gefördert werden soll. Stattdessen wäre es hilfreich zu begründen, warum dieses Setting gewählt wurde und warum es mit Blick auf die Kompetenzförderung, die das Entwickeln eines eigenen ethischen Standpunktes zum Ziel hat, geeignet ist. Dadurch bleibt fraglich, ob die gewählte Problemfrage („Ist der Einsatz genetischer Manipulation im Einsatz gegen Malaria vertretbar?“; Zitat paraphrasiert) in einem zehnminütigen Einstieg in Form eines Lehrervortrags eine kritische Reflexion vorbereiten kann. Ähnliches gilt für die Sicherungsphase, nachdem die Lernenden die in der Hausaufgabe gesammelten Argumente für ihre jeweilige Rolle (Forscherin, Entwicklungshelfer, Umweltschützer, Ärztin) ausgewählt und sich auf die Podiumsdiskussion vorbereitet haben:

*Als Auftrag sollen die Schüler\*innen vorgetragene Argumente der Rollen in einer Checkliste abhaken [KISR, Kat. 18] und nachfolgend ihre eigene Position darlegen [KAs, Kat. 30], bei verfügbarer Zeit in einer Abstimmung. Auf diese Weise wird Bewertungskompetenz gefördert, da die Schüler\*innen in dieser Stunde andere Perspektiven einnehmen, das Problem unter verschiedenen Perspektiven lösen und einen persönlichen Standpunkt einnehmen [OTS, Kat. 1]. (Zitat paraphrasiert)*

Es wird nicht deutlich, inwieweit die getroffenen Planungsentscheidungen tatsächlich einem präzisen Ziel verbunden sind, schließlich adressiert die Argumentation aus der Sicht einer bestimmten Rolle eine andere Teilkompetenz der Bewertungskompetenz als die Entwicklung eines eigenen Standpunktes. Ergänzend dazu wären wichtige Fragen zu

beantworten: Inwiefern trägt der Hörauftrag zur Förderung von Bewertungskompetenz bei? Wie wird sichergestellt, dass die Lernenden zwischen ihrer eigenen ethischen Position und der ihrer Rolle unterscheiden können? Inwiefern wird die eigentliche Podiumsdiskussion beispielweise bezüglich der Überzeugungskraft der Argumente reflektiert?

Zusammenfassend zeigt sich, dass zur Begründung einiger Planungsentscheidungen Wissensfacetten sinnvoll miteinander vernetzt werden. Dabei werden jedoch für die Kompetenzförderung relevante Aspekte wie die kompetenzbereichsspezifischen Lerngruppenvoraussetzungen oder die Konsolidierung des Kompetenzzuwachses in der Sicherungsphase kaum oder gar nicht berührt, so dass der Lehr-Lern-Prozess hier lückenhaft bleibt. In diesem Entwurf werden also unterschiedliche Planungsaspekte miteinander verknüpft, wobei wichtige legitimationsbedürftige fachdidaktische Planungsentscheidungen unberücksichtigt bleiben.

## Diskussion

Die vorliegende Studie zur Planungskompetenz von Biologie-Referendar\*innen im Rahmen der Prüfung zum zweiten Staatsexamen hat das Ziel, authentische Unterrichtsentwürfe angehender Biologie-Lehrkräfte zu untersuchen (Weingarten 2019) und den Zusammenhang zwischen Facetten fachdidaktischen Wissens und der Qualität von Unterrichtsentwürfen zu analysieren. Die Ergebnisse liefern Einsichten in die Probleme beim Verfassen elaborierter Unterrichtsentwürfe, insbesondere in die von Gassmann (2013) beschriebenen Schwierigkeiten beim Herstellen interdependenter Planungsüberlegungen und dabei vor allem bei der Nutzung fachdidaktischen Wissens zur Begründung von Planungsentscheidungen (Zaragoza et al. 2021).

## Limitationen

Ausführliche schriftliche Unterrichtsentwürfe gelten als Ausbildungsartefakte (Seel 2011), die aufgrund der heterogenen Anforderungen der jeweiligen Ausbildungsinstitutionen in der ersten und zweiten Phase der Lehrkräftebildung nur in begrenztem Maße verallgemeinernde Schlussfolgerungen über die Planungskompetenz angehender Lehrkräfte zulassen. Die Prüfung zum zweiten Staatsexamen als besonderer und daher unauthentischer Planungskontext (Karlström und Hamza 2021) kann sowohl den Planungsprozess der Referendar\*innen (z.B. Auswahl des Themas, der Methoden) als auch das Planungsprodukt (z.B. post-hoc-Legitimationen) beeinflusst haben (Weingarten 2019). Insbesondere letzteres ist bei der Aussagekraft der Untersuchung zu berücksichtigen: Ob Planungsentscheidungen aus didaktischen Erwägungen heraus getroffen und entspre-

chend begründet wurden oder ob sie eher intuitiv getroffen und nachträglich mit Argumenten begründet wurden, die im Rahmen einer Prüfungssituation womöglich als positiv bewertet werden, kann im Rahmen dieser Studie nicht beurteilt werden. Hier wurde das bloße Planungsprodukt, nicht aber der vorgelagerte Planungsprozess fokussiert. Die Lösung des Problems 2. Art (Abb. 2) hängt also auch vom Planungskontext ab. Bereits Weingarten (2019) diskutiert die Zweckgebundenheit von Prüfungsentwürfen mithilfe der Theorie des symbolischen Interaktionismus, nach der die Verfasser\*innen die Erwartungshaltung der Prüfenden in ihrer Planung miteinbezogen haben könnten. Hinzu kommt, dass aufgrund der Bedeutung der zweiten Staatsexamensprüfung die Referendar\*innen ihre Entwürfe möglicherweise von Kolleg\*innen haben Korrektur lesen lassen und dass die geplante Unterrichtsstunde zumindest in Teilen auf fachdidaktischem Wissen externer Personen beruht. Das kann bedeuten, dass wir nicht nur ePCK, sondern in Teilen cPCK (Carlson und Daehler 2019) erfasst haben. Andererseits macht die Staatsexamensprüfung als Planungskontext die untersuchten Entwürfe vergleichbar und erlaubt aufgrund ihrer ökologischen Validität trotz der methodischen Nachteile gegenüber Vignetten- oder Performanztests (Schröder et al. 2020) eine Bestandsaufnahme der Planungskompetenz angehender Lehrkräfte am Ende der Ausbildung (König et al. 2015; Weingarten 2019).

Die Stichprobengröße von  $N=46$  dieser Studie ist zwar wesentlich geringer als beispielsweise bei König et al. (2021) mit  $N=172$ , König et al. (2015) mit  $N=106$  oder Weingarten (2019) mit  $N=180$ , die jedoch vor allem eine statistische Auswertung anstrebten, in der die Entwicklung über mehrere Messzeitpunkte (König et al. 2021; König et al. 2015) oder der Einfluss verschiedener Variablen (z.B. Geschlecht, Schulform, Unterrichtsfach) auf die Planungskompetenz (Weingarten 2019) untersucht wurde. Im Vergleich zu anderen qualitativen Studien ( $N=25$ , Gassmann 2013;  $N=10$ , Chizhik und Chizhik 2018;  $N=10$ , Morine-Dershimer 1979) basiert die vorliegende Studie auf einer relativ großen Stichprobe.

Die fehlende Möglichkeit, zusätzliche Interviews mit den Referendar\*innen zu führen, erlaubt anders als bei Park und Chen (2012) oder Reynolds und Park (2021) keine Datentriangulation. Die ausschließliche Analyse von Unterrichtsentwürfen ohne weitere Datenquellen oder methodische Zugänge ist in der PCK-Forschung unüblich (Chan und Hume 2019), denn anders als in Interviews (z.B. Gassmann 2013) oder in Interventionsstudien (z.B. Weitzel und Blank 2020) konnten die Denkprozesse der Referendar\*innen nicht stimuliert werden, so dass Zusammenhänge zwischen Wissensfacetten, die möglicherweise vorhanden waren, aber schriftlich nicht dargelegt wurden (Morine-Dershimer 1979), nicht erfasst werden konnten (Abb. 2: Problem 2. Art). Allerdings würde dadurch auch

eher pPCK als ePCK (Carlson und Daehler 2019) erfasst, was nicht Ziel dieser Untersuchung war. Zudem konnten die Unterrichtsstunden, die in den Entwürfen geplant wurden, nicht beobachtet und in ihrer Qualität eingeschätzt werden, so dass der Datensatz keine Rückschlüsse auf die prognostische Validität erlaubt (König et al. 2015; Zaragoza et al. 2021). Da es in der vorliegenden Studie allerdings ausschließlich um die Kompetenz des Schreibens von Unterrichtsentwürfen und Aussagen über die Kompetenz allein anhand der Performanz, d.h. des Planungsprodukts, gemacht werden (Abb. 2), besteht eher Potenzial für weiterführende Studien, die mehr als die schriftliche Planungskompetenz in den Blick nehmen.

Aus forschungsökonomischen Gründen konnte nur für vier Entwürfe eine Zweitbewertung eingeholt werden. Die starke Korrelation ( $r=0,71$ ) der Bewertung der beiden Fachseminarleiter lässt auf eine gute Übereinstimmung schließen und stützt die intersubjektive Nachvollziehbarkeit (Göhner und Krell 2020; Mayring 2015) der Einschätzungen des ersten Fachseminarleiters, der alle 46 Entwürfe bewertet hat. Die in ihren Begründungen kritisierten Mängel der Entwürfe sind weitgehend deckungsgleich, Unterschiede in der Bewertung resultierten aus einer unterschiedlich starken Gewichtung dieser Mängel. Dass die Einschätzungen im Wesentlichen übereinstimmen, obwohl es keine allgemeingültigen Kriterien guter Unterrichtsentwürfe gibt (Döbrich und Abs 2008; König et al. 2020; Vogelsang und Riese 2017), kann in Teilen zufällig sein, kann jedoch auch damit erklärt werden, dass beide Fachseminarleiter zum Zeitpunkt der Datenerhebung im Land Berlin tätig waren und die dort im Rahmen der Staatsexamensprüfung geltenden Kriterien angelegt haben. Durch die Aufteilung der Stichprobe in Quartile (Reynolds und Park 2021) wird die Bedeutung von Bewertungsunterschieden zudem relativiert.

## Inhaltliche Diskussion

### Forschungsfrage 1: Planungsqualität und Einzelaspekte des ePCK<sub>p</sub>

Einige planungsrelevante Aspekte wie die Analyse des Kompetenzstandes der Lernenden oder Maßnahmen zur Binnendifferenzierung werden sowohl in elaboriert als auch in wenig elaboriert eingeschätzte Entwürfe berücksichtigt, so dass für diese Stichprobe der Versuch verzeichnet werden kann, didaktisch adaptiven Unterricht zu planen (König et al. 2015).

Erstaunlich ist, dass Schülervorstellungen [KSU] in dieser Stichprobe nicht berücksichtigt worden sind, obwohl sie als Teil einer themenspezifischen Analyse des Kompetenzstandes der Lernenden (Kattmann et al. 1997; Whittington et al. 2021) relevant sein können. Insbesondere in den Stunden zu den Kompetenzbereichen Fachwissen und Erkennt-

nisgewinnung wäre eine Bezugnahme auf Schülervorstellungen zu erwarten gewesen. Im Sinne der konstruktivistischen Lerntheorie könnten sie zu einer noch höheren didaktischen Adaptivität beitragen, indem in der Planung explizit die Vorstellungen der spezifischen Lerngruppe aufgegriffen und ggf. sogar im Sinne der Conceptual Change-Theorie (vgl. Krüger 2007) erweitert werden könnten. Dieses Ergebnis bestätigt die Befunde von Weitzel und Blank (2020) und wirft die Frage auf, warum nicht gerade bei der schriftlichen Unterrichtsplanung, die zeitlich nicht so stark limitiert ist wie andere Forschungsdesigns, auf das große Angebot an deutschsprachiger fachdidaktischer Literatur zu Schülervorstellungen (Hamann und Asshoff 2014; Kattmann 2015) Bezug genommen und entweder zur Gestaltung des Lehr-Lern-Prozesses (Abb. 2: *Problem 1. Art*) oder zumindest zur Legitimation bestimmter Planungsentscheidungen (Abb. 2: *Problem 2. Art*) genutzt wird. Hierin liegt allerdings auch eine forschungsmethodische Schwierigkeit: Das Pentagon-Modell fachdidaktischen Wissens (Park und Oliver 2008) suggeriert, dass alle Verbindungen zwischen den fünf Wissensfacetten gleichwertig sind. So wurden unter dem Wissen über die Lerngruppe neben Schülervorstellungen beispielsweise auch motivationale Aspekte der Lerngruppe kodiert (Tab. 3). Es ist davon auszugehen, dass beispielsweise Schülervorstellungen im Sinne der konstruktivistischen Lerntheorie im Zuge der didaktischen Rekonstruktion (Gropengießer und Kattmann 2020) besonders bedeutsam sind und in der Planung von Biologieunterricht noch viel stärker berücksichtigt werden sollten, als es allgemeindidaktische Planungsmodelle aufzeigen könnten (vgl. Rothland 2021). Insofern ist zu beachten, dass die Kodierhäufigkeiten allein nur eine eingeschränkte Aussage über die Qualität zulassen.

Neben vielen erfreulicherweise oft berücksichtigten Aspekten wie beispielsweise der Binnendifferenzierung könnte es hilfreich sein, v.a. auch die Aufgabenstellungen ausführlicher darzustellen und didaktisch zu begründen. Oft erschöpfen sich die methodisch-didaktischen Begründungen in einer Beschreibung des geplanten Stundenverlaufs, in dem vereinzelt wichtige Methodenentscheidungen oder Differenzierungsmaßnahmen begründet werden. Die Aufgabenstellungen, mit denen sich die Schüler\*innen aber hauptsächlich beschäftigen, werden trotz ihrer großen Bedeutung für erfolgreiche Lehr-Lern-Prozesse (Kang et al. 2016) oft nicht berücksichtigt, dabei gäbe es in der Gestaltung von Aufgaben einige wichtige Aspekte (z.B. die Abstraktheit, die sprachliche Komplexität) zu bedenken (Maier et al. 2014).

Darüber hinaus wird eine Überprüfung des Lernzuwachses [KAs] erwartungskonform (z.B. Aydin und Boz 2013; Park und Chen 2012; Weitzel und Blank 2020) selten berücksichtigt (Tab. 6). Gerade in einer Examenstunde wäre es möglich und sinnvoll, am Stundenende der Prüfungskommission den Lernzuwachs der Lernenden sichtbar zu

machen, um die Relevanz der eigenen Stunde zu legitimieren. Das weitgehende Fehlen von KAs kann im Rahmen dieser Studie nicht hinreichend damit erklärt werden, dass keine geeigneten Methoden bekannt sind (Bravo und Co-fré 2016; Hanuscin et al. 2018; Mthethwa-Kunene et al. 2015) oder dass eine transmissive Lehrhaltung eine Rückmeldung zum Lernzuwachs ausschließt (Chan und Yung 2018), schließlich wurde kompetenzorientierter Unterricht gefordert und die Stunden konnten Wochen im Voraus geplant und Methoden recherchiert werden. Insofern bleibt offen, ob die seltene Berücksichtigung von KAs beispielsweise auf die historisch gewachsene Didaktik im deutschsprachigen Raum zurückzuführen ist, die anders als im anglo-amerikanischen Raum (z. B. Wiggins und McTighe 2005) die formative Rückmeldung nicht zum konstitutiven Kern der Unterrichtsgestaltung hat, wie sich beispielsweise im Fehlen dieses Aspekts in deutschsprachigen Planungsratgebern zeigt (Vogelsang und Riese 2017).

### Forschungsfrage 2: Planungsqualität und Vernetzung des ePCK<sub>p</sub>

Ein eindeutiger Zusammenhang zwischen der Qualität der Unterrichtsentwürfe und der Vernetzung des fachdidaktischen Wissens konnte anders als bei Reynolds und Park (2021) in dieser Stichprobe nicht festgestellt werden (Tab. 4). Dass trotz Normierung der Wortzahlen der Entwürfe in den gut bewerteten Entwürfen mehr ePCK<sub>p</sub>-Verbindungen hergestellt werden als in den als schwach bewerteten Entwürfen, deutet dennoch darauf hin, dass sich Planungsqualität nicht in der Länge der Entwürfe, sondern unter anderem auch in der Vernetzung fachdidaktischer Wissensfacetten ausdrückt. Ein Grund für die fehlenden Korrelationen zwischen der Planungsqualität und den meisten ePCK<sub>p</sub>-Verbindungen könnte darin bestehen, dass mit den beiden eingesetzten Methoden (*PCK map approach*, Park und Suh 2019; Einschätzung der Planungsqualität durch Fachseminarleiter) Unterschiedliches erfasst wird: Die Kodierung des ePCK<sub>p</sub> ermöglicht nur die lokale Erfassung fachdidaktischer Wissensfacetten und ihrer Verknüpfungen. Für Fachseminarleitungen ist jedoch weniger relevant, ob lokal an vereinzelt Stellen im Entwurf fachdidaktische Wissensfacetten genutzt werden, sondern ob global über den gesamten Entwurf hinweg eine Passung zwischen den vielfältigen Analysen der Unterrichtsvoraussetzungen und den Planungsentscheidungen (Abb. 2: *Problem 1. Art*) besteht. Beim *PCK map approach* (Abb. 3) wird nicht bewertet, ob eine Aussage elaboriert ist oder nicht, während die Planungsqualität des gesamten Entwurfs in der Einschätzung von Seminarleitungen davon abhängig ist, ob die getroffenen Planungsentscheidungen in eine kompetenzfördernde Lehr-Lern-Struktur münden.

Dass die häufigsten ePCK<sub>p</sub>-Verbindungen in der Gesamtstichprobe die Trias aus Kompetenzentwicklung [OTS], Lerngruppenvoraussetzungen [KSU] und Instruktionsstrategien [KISR] betreffen, bestätigt die empirische Befundlage (Park und Chen 2012; Reynolds und Park 2021; Weitzel und Blank 2020) und deutet darauf hin, dass die Gestaltung der Lehr-Lern-Prozesse adaptiv an die Lerngruppe angepasst wird. Dass dabei jedoch die fachliche Klärung in Form einer Sachstrukturanalyse [KSC] kaum miteinbezogen wird (Tab. 6b), widerspricht dem Modell der didaktischen Rekonstruktion (Kattmann et al. 1997) und kann als zusätzlicher Hinweis gelten, dass die Anpassung des Lehr-Lern-Prozesses an eine didaktische Analyse des Themas schwerfällt (Aydin und Boz 2013) oder aber als nicht besonders relevant eingeschätzt wird.

In diesem Zusammenhang stellt sich die Frage, ob die vorliegenden Ergebnisse überhaupt fachspezifischen Charakter oder ob es sich um Befunde handelt, die in ähnlicher Weise auch für andere Fächer repliziert werden könnten. Es wurde für diese Untersuchung das vorrangig allgemeindidaktisch modellierte Konstrukt *Planungskompetenz* (Rothland 2021; Vogelsang und Riese 2017) herangezogen und mithilfe des naturwissenschaftsdidaktischen Konstrukts *Fachdidaktisches Wissen* bzw. *PCK* (Baumert und Kunter 2006; Park und Oliver 2008) operationalisiert. In der Unterrichtsplanung werden naturgemäß sowohl allgemeine als auch fachspezifische Wissensbestände genutzt (Zaragoza et al. 2021), letztere allerdings signifikant seltener, wie König et al. (2020) in ihrer Untersuchung von Unterrichtsentwürfen von Referendar\*innen im Fach Deutsch zeigen. Betrachtet man die Kodierungen losgelöst von ihrem Auftreten in einem Unterrichtsentwurf, ließen sich manche der Aussagen auch als allgemein-pädagogisches Wissen (pedagogical knowledge, PK; Baumert und Kunter 2006; Tamir 1988) auffassen. Im Sinne des Refined Consensus Models (Carlson und Daehler 2019) betrachten wir PK als distinkte Dimension des Professionswissens, ohne die elaboriertes PCK nicht entwickelt und angewendet werden kann. Insofern wirkt PK auf vielfältige fachdidaktische Planungsentscheidungen ein und wird zu deren Begründung genutzt. Anders als König et al. (2020) wurde in der vorliegenden Untersuchung für die Subkategorien (Tab. 3) nicht jeweils unterschieden, ob eine Äußerung allgemein oder fachspezifisch gemeint ist, wir haben sie in jedem Fall als PCK und somit fachdidaktisch aufgefasst, weil sie sich immer auf eine konkrete Unterrichtssituation und somit einen konkreten Unterrichtsinhalt und eine Kompetenz bezieht: „Only subject-specific planning decisions may enable in-depth elaborations, as student learning of a school subject takes place in a particular domain“ (König et al. 2020, S. 135). Insofern wäre dies auch eine mögliche Erklärung für die erwartungswidrigen Fälle in dieser Stichprobe, deren Vernetzungsgrad des ePCK<sub>p</sub> nicht mit der Planungsqualität korreliert (Tab. 5):

In Entwürfen mit vielen vernetzten Ausführungen kann eher auf einer allgemeinen Ebene argumentiert werden und vice versa (vgl. Forschungsfrage 3). Dieses Vorgehen deckt sich mit einer neuen biologiedidaktischen Perspektive auf die Unterrichtsqualitätsforschung, der zufolge es keine allgemeinen und fachspezifischen Qualitätsmerkmale gibt, sondern diese beiden Aspekte Pole eines Kontinuums bilden, auf dem sich die verschiedenen Merkmale von Unterrichtsqualität (z. B. auch von PCK) einordnen lassen (Neuhaus 2021).

### Forschungsfrage 3: Qualitative Unterschiede im ePCK<sub>p</sub>

Der Vergleich der zwei Entwürfe kann als Einblick in eine kritische Bestandsaufnahme der Planungspraxis angehender Biologie-Lehrkräfte (Weingarten 2019) gelten und zugleich als Ausgangspunkt für die Entwicklung von Kriterien guter Unterrichtsentwürfe, die sich erwartungswidrig (Reynolds und Park 2021) nicht in der bloßen quantitativen Berücksichtigung von ePCK<sub>p</sub>-Elementen, sondern in ihrer qualitativen Umsetzung zeigen (Chan und Yung 2018). Zwar ließe sich einwenden, dass die beiden untersuchten Entwürfe Bewertungskompetenz adressieren, mit deren Förderung Lehrkräfte ohnehin Schwierigkeiten haben (Hartmann-Mrochen 2012), weil ihnen prozedurales Wissen zur konkreten Förderung von Bewertungskompetenz fehlt (Alfs 2012) oder weil ihnen die Vernetzung verschiedener Facetten fachdidaktischen Wissens in diesem Bereich schwerfällt (Pohlmann 2019). Es sei betont, dass die in diesem Beitrag herausgearbeiteten Unterschiede in den Entwürfen sich gleichermaßen in Entwürfen der drei anderen Kompetenzbereiche identifizieren lassen und die Auswahl von zwei Fällen zum gleichen Kompetenzbereich der besseren Vergleichbarkeit dient.

Die Gestaltung eines kompetenzorientierten Lehr-Lern-Prozesses setzt eine dezidierte Analyse der Lerngruppenvoraussetzungen [KSU] mit Blick auf die angestrebte Kompetenzentwicklung [OTS] voraus: Während Entwurf 38 von einer differenzierten Diagnose in Bezug auf die Bewertungskompetenz ausgeht und diese im Rahmen eines langfristigen Konzepts über mehrere Stunden hinweg entwickelt, wird in Entwurf 33 ausschließlich Bezug auf motivationale Aspekte und bereits entwickeltes Vorwissen über den Fachinhalt genommen. Ein langfristiges Konzept zur Kompetenzentwicklung fehlt hier (Hanuscin et al. 2018). Zwar werden in beiden Fällen ePCK<sub>p</sub>-Elemente verknüpft [OTS-KSU], doch erst die explizit kompetenzorientierte Lerngruppenanalyse in Entwurf 38 ermöglicht eine adaptive Gestaltung des Lehr-Lern-Prozesses, insbesondere im Hinblick auf die Aufgabenstellungen [KISR] (König et al. 2015).

Entwurf 38 zeigt eine schlüssige Passung der verschiedenen ePCK<sub>p</sub>-Elemente, ohne dass diese dazu fortwährend

wieder aufgegriffen werden. Eines der von Gassmann (2013) identifizierten schwierigkeiterzeugenden Merkmale bei der Unterrichtsplanung – die Herstellung von Interdependenzen – ist hier also gelungen. Die Integration des vorhandenen ePCK<sub>p</sub> ließe sich jedoch durch eine deutlichere Bezugnahme auf die entsprechenden Wissensfacetten insbesondere in den Begründungen für Planungsentscheidungen (Vogelsang und Riese 2017) transparenter machen (Abb. 2: *Problem 2. Art*). Da Unterrichtsentwürfe in eine Kommunikationssituation mit den hospitierenden Personen (Seminarleitungen, Schulleitungen, Kolleg\*innen) eingebettet sind, würde eine transparentere Vernetzung des eigenen Wissens die Legitimationsstruktur der eigenen Planung stärker stützen, als es den Leser\*innen zu überlassen, kapitelübergreifend die Passung selbst zu erkennen. Da ein Großteil der Planung mental abläuft (Mörner-Dersheimer 1979), könnte eine stärkere Aufforderung, die ePCK<sub>p</sub>-Facetten expressis verbis darzustellen, dabei helfen, Leerstellen im Unterrichtsentwurf zu füllen.

Umgekehrt zeigt sich exemplarisch anhand der Sicherungsphase in Entwurf 33, dass eine starke Integration verschiedener ePCK<sub>p</sub>-Elemente nicht unbedingt bereits als Qualitätsindikator gelten kann. Dies stellt die methodologische Basis des *PCK map approaches* (Abb. 3; Park und Oliver 2008; Park und Suh 2019) zumindest teilweise infrage und kann Chan et al. (2019) folgend als Plädoyer für die Entwicklung von Niveaustufen zur Erfassung von ePCK<sub>p</sub> betrachtet werden. Die zur Förderung von Bewertungskompetenz als eher dysfunktional einzuschätzende Sicherungsphase zeigt, dass hier nicht bloß auf eine formative Rückmeldung für die Lernenden verzichtet wird (Beyer und Davis 2012; Weitzel und Blank 2020), sondern auch, dass die angewandte Methodik ungeeignet erscheint, um das Erreichen der im angestrebten Kompetenzzuwachs dargelegten Zielstellung zu prüfen (Bravo und Cofré 2016; Hanuscin et al. 2018; Mthethwa-Kunene et al. 2015; Park und Chen 2012). Insbesondere konkrete und messbare Indikatoren zum Nachweis des Lernzuwachses (Hanuscin et al. 2018) wären hilfreich, um die eigenen methodischen Entscheidungen auf Funktionalität hin zu prüfen.

## Implikationen

Die vorliegende Studie liefert neben einer Untersuchung authentischer Unterrichtsentwürfe (Weingarten 2019) angehender Biologie-Lehrkräfte Impulse für die weitere Forschung: Wenn die Planung als Kernaufgabe von Lehrkräften gilt (KMK 2019; Baumert und Kunter 2006; Carlson und Daehler 2019) und PCK zum einen positiv mit der Unterrichtsqualität korreliert (Keller et al. 2017; Kunter et al. 2013; Park et al. 2011) und zum anderen als Prädiktor des Lernerfolgs von Lernenden gilt (Förtsch et al. 2016;



Mahler et al. 2017; Jin et al. 2015), dann besteht offenkundig ein Bedarf, die beiden Konstrukte Planungskompetenz und fachdidaktisches Wissen zusammenzuführen und angehende Lehrkräfte beim Problemlösen in ihrer Unterrichtsplanung zu unterstützen. Für angehende Lehrkräfte erscheint es ratsam, sprachlich möglichst viele Interdependenzen zwischen ePCK<sub>p</sub>-Elementen herzustellen und transparent zu machen (Abb. 2: *Problem 2. Art*). Wenn angehende Lehrkräfte mit dieser Haltung ihren Entwurf schreiben, wären sie bereits im Planungsprozess dazu angehalten zu prüfen, ob eine bestimmte Planungsentscheidung (z. B. die Gestaltung einer Phase) sinnvoll mit anderen Planungsentscheidungen (z. B. einer Methode), den analysierten vorunterrichtlichen Bedingungen (z. B. aktueller Kompetenzstand der Schüler\*innen) und vor allem dem angestrebten konkreten Lernziel im Einklang steht (König et al. 2021). Diese Verknüpfungen in den Begründungen der Planungsentscheidungen explizit zu machen, könnte dabei helfen, den Rezipienten des Unterrichtsentwurfs (v. a. Seminarleitungen, Schulleitungen) die Qualität der eigenen fachdidaktischen Arbeit in Form eines Unterrichtsentwurfs nachvollziehbar zu machen.

Es ist davon auszugehen, dass angehende Lehrkräfte auch in anderen Fächern mit sehr ähnlichen Problemen beim Schreiben von Unterrichtsentwürfen konfrontiert sind. Hier besteht derzeit noch großer Forschungsbedarf – auch deshalb, weil auf diesem Wege fachspezifische Herausforderungen identifiziert und Maßnahmen in den jeweiligen Fachdidaktiken entwickelt und erprobt werden können. So wird derzeit ein biologiedidaktisches Kriterienraster entwickelt und geprüft, das angehenden Lehrkräften mithilfe transparenter Kriterien aufzeigt, wann beispielsweise eine Lerngruppenanalyse oder eine Begründung für Aufgabenstellungen gelungen ist (Großmann und Krüger 2022).

Der Vergleich zweier Entwürfe hat gezeigt, dass das alleinige Berücksichtigen von ePCK<sub>p</sub>-Elementen noch keinen hinreichenden Qualitätsindikator darstellt. Derartige Kriterien inklusive Niveauabstufungen lassen sich auf der Grundlage dieser Studie entwickeln, um die schriftliche Planungskompetenz objektiv, reliabel und valide einschätzen zu können (Schröder et al. 2020), was insbesondere im Zuge der zweiten Phase der Lehrkräftebildung ein Desiderat darstellt (Döbrich und Abs 2008; Kärner et al. 2019; Strietholt und Terhart 2009). Ein solches Instrument würde einerseits neue Forschungsperspektiven eröffnen (z. B. die Messung der längsschnittlichen Entwicklung von Planungskompetenz in der ersten und zweiten Phase der Lehrkräftebildung). Andererseits könnte es von Dozierenden und Seminarleitungen, aber auch Studierenden und Referendar\*innen genutzt werden (Brookhart 2018), um die Qualität von Unterrichtsentwürfen vorunterrichtlich einschätzen zu können. Für die Lehrkräftebildung scheint es notwendig zu sein, Aspekte wie die Berücksichtigung

von Schülervorstellungen, die Bedeutung der Sachstrukturanalyse oder die Notwendigkeit, Lernzuwachs transparent zu machen, stärker im Zuge der Unterrichtsplanung zu thematisieren und angehenden Biologie-Lehrkräften dabei zu helfen, diese verschiedenen Aspekte in ihren Vernetzungen mitzudenken (*Problem 1. Art*) und in den Unterrichtsentwürfen auf elaborierte Weise darzulegen (*Problem 2. Art*).

**Danksagung** Wir danken der Senatsverwaltung für Bildung, Jugend und Familie des Landes Berlin für die Kooperation und die Bereitstellung der anonymisierten Staatsexamensentwürfe.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

Das Projekt K2Teach (Know how to teach) wird im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1802 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Alfs, N. (2012). *Ethisches Bewerten fördern*. Hamburg.
- Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: enacted and personal PCK. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 271–286). Singapore: Springer.
- Aydin, S., & Boz, Y. (2013). The nature of integration among PCK components: a case study of two experienced chemistry teachers. *Chemistry Education Research and Practice*, 14(4), 615–624. <https://doi.org/10.1039/C3RP00095H>.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Beyer, C. J., & Davis, E. A. (2012). Learning to critique and adapt science curriculum materials: Examining the development of preservice elementary teachers' pedagogical content knowledge. *Science Education*, 96(1), 130–157. <https://doi.org/10.1002/sce.20466>.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>.

- Borko, H., & Livingston, C. (1989). Cognition and improvisation: differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473–498.
- Bravo, P., & Cofré, H. (2016). Developing biology teachers' pedagogical content knowledge through learning study: the case of teaching human evolution. *International Journal of Science Education*, 38(16), 2500–2527. <https://doi.org/10.1080/09500693.2016.1249983>.
- Brookhart, S.M. (2018). Appropriate criteria: key to effective rubrics. *Frontiers in Education*, 3(22), 1–12. <https://doi.org/10.3389/educ.2018.00022>.
- Carlson, J., & Daehler, K.R. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 77–92). Singapore: Springer.
- Chan, K.K.H., & Hume, A. (2019). Towards a consensus model: literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 3–76). Singapore: Springer.
- Chan, K.K.H., & Yung, B.H.W. (2018). Developing pedagogical content knowledge for teaching a new topic: more than teaching experience and subject matter knowledge. *Research in Science Education*, 48(2), 233–265. <https://doi.org/10.1007/s11165-016-9567-1>.
- Chan, K.K.H., Rollnick, M., & Gess-Newsome, J. (2019). A grand rubric for measuring science teachers' pedagogical content knowledge. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 253–271). Singapore: Springer.
- Chizhik, E.W., & Chizhik, A.W. (2018). Using activity theory to examine how teachers' lesson plans meet students' learning needs. *The Teacher Educator*, 53(1), 67–85. <https://doi.org/10.1080/08878730.2017.1296913>.
- Döbrich, P., & Abs, H.J. (2008). Evaluation der zweiten Phase der Lehrerbildung. Pädagogische Entwicklungsbilanzen mit Studienseminaren in Hessen. *Schulverwaltung. Hessen, Rheinland-Pfalz*, 13(3), 70–73.
- Förtsch, C., Werner, S., von Kotzebue, L., & Neuhaus, B.J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. <https://doi.org/10.1080/09500693.2016.1257170>.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Gassmann, C. (2013). *Erlebte Aufgabenschwierigkeit bei der Unterrichtsplanung: Eine qualitativ-inhaltsanalytische Studie zu den Praktikumsphasen der universitären Lehrerbildung*. Wiesbaden: Springer.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, P.J. Friedrichsen, & J. Loughran (Hrsg.), *Re-examining pedagogical content knowledge in science education* (S. 28–42). New York: Routledge.
- Göhner, M., & Krell, M. (2020). Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review. *Zeitschrift für Didaktik der Naturwissenschaften*, 26, 207–225. <https://doi.org/10.1007/s40573-020-00111-0>.
- Gropengießer, H. (2020). Experimentieren. In H. Gropengießer, U. Harms & U. Kattmann (Hrsg.), *Fachdidaktik Biologie* (S. 284–293). Seelze: Aulis.
- Gropengießer, H., & Kattmann, U. (2020). Didaktische Rekonstruktion. In H. Gropengießer, U. Harms & U. Kattmann (Hrsg.), *Fachdidaktik Biologie* (S. 16–23). Seelze: Aulis.
- Großmann, L. & Krüger, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen. *Erkenntnisweg Biologiedidaktik*, 19, 21–39.
- Großmann, L. & Krüger, D. (2022). Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen. *SEMINAR – Lehrerbildung und Schule*, 2022(1), 90–109.
- Großschedl, J., Harms, U., Kleickmann, T., & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: structure and learning opportunities. *Journal of Science Teacher Education*, 26(3), 291–318. <https://doi.org/10.1007/s10972-015-9423-6>.
- Haas, A. (1998). *Unterrichtsplanung im Alltag: Eine empirische Untersuchung zum Planungshandeln von Hauptschul-, Realschul- und Gymnasiallehrern*. Regensburg: Roderer. Zugl.: Weingarten, Pädag. Hochsch., Diss., 1998
- Hammann, M., & Asshoff, R. (2014). *Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten*. Seelze: Klett.
- Hanuscin, D.L., Cisterna, D., & Lipsitz, K. (2018). Elementary teachers' pedagogical content knowledge for teaching structure and properties of matter. *Journal of Science Teacher Education*, 29(8), 665–692. <https://doi.org/10.1080/1046560X.2018.1488486>.
- Hartmann-Mrochen, M. (2012). *Zwischen Notengebung und Urteilsfähigkeit: Einstellungen und Vorstellungen von Lehrkräften verschiedener Fachkulturen zum Kompetenzbereich Bewertung der Nationalen Bildungsstandards*. Hamburg: Univ., FB Erziehungswiss., Diss. 2012
- Jin, H., Shin, H., Johnson, M.E., Kim, J., & Anderson, C.W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching*, 52(9), 1269–1295. <https://doi.org/10.1002/tea.21243>.
- Jüttner, M., & Neuhaus, B.J. (2013). Das Professionswissen von Biologielehrkräften: Ein Vergleich zwischen Biologielehrkräften, Biologen und Pädagogen. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 31–49.
- Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching*, 53, 1316–1340. <https://doi.org/10.1002/tea.21329>.
- Karlström, M., & Hamza, K. (2021). How do we teach planning to preservice teachers—A tentative model. *Journal of Science Teacher Education*, 32, 664–685. <https://doi.org/10.1080/1046560X.2021.1875163>.
- Kärner, T., Bonnes, C., & Schölzel, C. (2019). Bewertungstransparenz im Referendariat. *Zeitschrift für Pädagogik*, 65(3), 378–400.
- Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht*. Hallbergmoos: Aulis.
- Kattmann, U., Duit, R., Gropengießer, H., & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion – Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3(3), 3–18.
- Keller, M.M., Neumann, K., & Fischer, H.E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, 54(5), 586–614. <https://doi.org/10.1002/tea.21378>.
- KMK (2019). Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf). Zugegriffen: 18.03.2022.
- KMK (2005). *Bildungsstandards im Fach Biologie für den mittleren Schulabschluss. Beschluss vom 16.12.2004*. München, Neuwied: Luchterhand.
- König, J., Buchholtz, C., & Dohmen, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen

- Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18(2), 375–404. <https://doi.org/10.1007/s11618-015-0625-7>.
- König, J., Bremerich-Vos, A., Buchholtz, C., Lammerding, S., Strauß, S., Fladung, I. & Schleiffer, C. (2017). Die Bedeutung des Professionswissens von Referendarinnen und Referendaren mit Fach Deutsch für ihre Planungskompetenz (PlanvoLL-D). In Wernke, S. & Zierer, K. (Hrsg.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 121–133). Bad Heilbrunn: Klinkhardt.
- König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I., & Glutsch, N. (2020). Pre-service teachers' generic and subject-specific lesson-planning skills: on learning adaptive teaching during initial teacher education. *European Journal of Teacher Education*, 43(2), 131–150. <https://doi.org/10.1080/02619768.2019.1679115>.
- König, J., Krepf, M., Bremerich-Vos, A. & Buchholtz, C. (2021). Meeting Cognitive Demands of Lesson Planning: Introducing the CODE-PLAN Model to Describe and Analyze Teachers' Planning Competence. *The Teacher Educator*, 56(4), 466–487. <https://doi.org/10.1080/08878730.2021.1938324>.
- Krüger, D. (2007). Die Conceptual Change-Theorie. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 81–92). Berlin Heidelberg: Springer.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Weinheim, Basel: Beltz Juventa.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *The journal of educational psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>.
- Mahler, D., Großschedl, J., & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, 39(2), 213–237. <https://doi.org/10.1080/09500693.2016.1276641>.
- Maier, U., Bohl, T., Driike-Noe, C., Hoppe, H., Kleinknecht, M., & Metz, K. (2014). Das kognitive Anforderungsniveau von Aufgaben analysieren und modifizieren können: Eine wichtige Fähigkeit von Lehrkräften bei der Planung eines kompetenzorientierten Unterrichts. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32(2), 340–358.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Beltz.
- Morine-Dershimer, G. (1979). *Teacher plan and classroom reality: The South Bay Study, part IV*. Research series, Bd. 60. East Lansing: Institute for Research on Teaching College of Education Michigan State University.
- Mthethwa-Kunene, E., Onwu, G.O., & de Villiers, R. (2015). Exploring biology teachers' pedagogical content knowledge in the teaching of genetics in Swaziland science classrooms. *International Journal of Science Education*, 37(7), 1140–1165. <https://doi.org/10.1080/09500693.2015.1022624>.
- Neuhaus, B. J. (2021). Unterrichtsqualität aus der Perspektive der Biologiedidaktik. *Unterrichtswissenschaft*, 49(2), 273–283. <https://doi.org/10.1007/s42010-021-00114-x>.
- Park, S. (2019). Reconciliation between the refined consensus model of PCK and extant PCK models for advancing PCK research in science. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 119–130). Singapore: Springer.
- Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941. <https://doi.org/10.1002/tea.21022>.
- Park, S., & Oliver, J. S. (2008). National board certification (NBC) as a catalyst for teachers' learning about teaching: the effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45(7), 812–834. <https://doi.org/10.1002/tea.20234>.
- Park, S., & Suh, J. K. (2019). The PCK map approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 185–197). Singapore: Springer.
- Park, S., Jang, J.-Y., Chen, Y.-C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching? Evidence from an empirical study. *Research in Science Education*, 41(2), 245–260. <https://doi.org/10.1007/s11165-009-9163-8>.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park: SAGE.
- Pohlmann, M. (2019). Förderung ethischer Bewertungskompetenz: der Einfluss ausgewählter Lerngelegenheiten auf die inhaltliche Ausdifferenzierung und die Kohärenz der Komponenten des fachdidaktischen Wissens von Biologielehrkräften, Oldenburg. <https://d-nb.info/1182534066/34>. Zugegriffen: 18.03.2022.
- Reynolds, W. M., & Park, S. (2020). Examining the relationship between the educative teacher performance assessment and pre-service teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, 58, 721–748. <https://doi.org/10.1002/tea.21676>.
- Rothland, M. (2021). Anmerkungen zur Modellierung und Operationalisierung (allgemeindidaktischer) Unterrichtsplanungskompetenz. *Unterrichtswissenschaft*. <https://doi.org/10.1007/s42010-021-00111-0>.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum Qualitative Sozialforschung*. <https://doi.org/10.17169/fqs-15.1.2043>.
- Schröder, J., Riese, J., Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., et al. (2020). Die Messung der Fähigkeit zur Unterrichtsplanung im Fach Physik mit Hilfe eines standardisierten Performanztests. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 103–122. <https://doi.org/10.1007/s40573-020-00115-w>.
- Seel, A. (2011). Wie angehende Lehrer/innen das Planen lernen. In K.-H. Arnold, T. Bohl & K. Zierer (Hrsg.), *Jahrbuch für Allgemeine Didaktik 2011: Thementeil: Entwicklung und Weiterentwicklung allgemeindidaktischer Modelle der Unterrichtsplanung* (S. 31–45). Baltmannsweiler: Schneider Verlag Hohengehren.
- SenBJF (2017). *Handbuch Vorbereitungsdienst: Materialien für den reformierten Berliner Vorbereitungsdienst*. Berlin. [https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch\\_vorbereitungsdienst.pdf](https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch_vorbereitungsdienst.pdf). Zugegriffen: 18.03.2022.
- Shulman, L. S. (1986). Those who understand: a conception of teacher knowledge. *American Educator*, 10(1), 9–15.
- Stender, A. (2014). *Unterrichtsplanung: Vom Wissen zum Handeln: theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung*. Berlin: Logos.
- Strietholt, R., & Terhart, E. (2009). Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung. *Zeitschrift für Pädagogik*. <https://doi.org/10.3262/ZP0904622>.
- Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, 4(2), 99–110.
- VERBI Software (2019). *MAXQDA 2020 [computer software]*. Berlin: VERBI Software.
- Vogelsang, C., & Riese, J. (2017). Wann ist eine Unterrichtsplanung gut? – Planungsperformanz in Praxisratgebern zur Unterrichtsplanung. In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung*:

- ein in Vergessenheit geratener Kompetenzbereich?!: Status Quo und Perspektiven aus Sicht der empirischen Forschung (S. 47–61). Bad Heilbrunn: Julius Klinkhardt.
- Weingarten, J. (2019). *Wie planen angehende Lehrkräfte ihren Unterricht? Empirische Analysen zur kompetenzorientierten Gestaltung von Lernangeboten*. Münster: Waxmann.
- Weitzel, H., & Blank, R. (2020). Pedagogical Content Knowledge in peer dialogues between pre-service biology teachers in the planning of science lessons. Results of an intervention Study. *Journal of Science Teacher Education*, 31(1), 75–93. <https://doi.org/10.1080/1046560X.2019.1664874>.
- Werner, J., Wernke, S., & Zierer, K. (2017). Der Einfluss didaktischer Modelle auf die allgemeindidaktische Unterrichtsplanungskompetenz von Lehramtsstudierenden. In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?!: Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 104–120). Bad Heilbrunn: Julius Klinkhardt.
- Wernke, S., & Zierer, K. (2017). Die Unterrichtsplanung – Ein in Vergessenheit geratener Kompetenzbereich?! In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?!: Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 7–16). Bad Heilbrunn: Julius Klinkhardt.
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292–305.
- Whittington, K., Southerland, S. A., & Tekkumru-Kisa, M. (2021). Examining relevance in pre-service science teachers' lesson plans. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2021.2012850>.
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. Alexandria: Association for Supervision and Curriculum Development.
- Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Zaragoza, A., Seidel, T., & Hiebert, J. (2021). Exploring preservice teachers' abilities to connect professional knowledge with lesson planning and observation. *European Journal of Teacher Education*. <https://doi.org/10.1080/02619768.2021.1996558>.

## Beitrag 4

**Großmann, L.** & Krüger, D. (2022). Students' Conceptions as a Neglected Perspective in Trainee Teachers' Biology Lesson Plans. In K. Korfiatis & M. Grace (Hrsg.), *Current Research in Biology Education. Contributions from Biology Education Research* (S. 181-193). ERIDOB, Springer. [https://doi.org/10.1007/978-3-030-89480-1\\_14](https://doi.org/10.1007/978-3-030-89480-1_14)

Reproduced with permission of Springer.

# Chapter 14

## Students' Conceptions as a Neglected Perspective in Trainee Teachers' Biology Lesson Plans



Leroy Großmann and Dirk Krüger

### 14.1 Introduction

Following a constructivist view on learning processes, good teaching practice requires a thoroughly defined adjustment of instructional strategies to match the unique requirements of the group of students in question (Duit & Treagust, 2012). In that sense, a constructivist learning environment must fit specifically to the current individual needs of particular students. To satisfy those needs in instructional practice, teachers have to be aware of them by analyzing the group prior to teaching with regard to cognitive, motivational and social aspects (Hardy et al., 2019). This corresponds to the definition of pedagogical content knowledge (PCK) as the core-part of science teachers' expertise, i.e.

...the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes. (Gess-Newsome, 2015, p. 36).

Consequently, teachers are expected to explicitly consider the general learning context and characteristics of their students as well as topic-specific learning difficulties and misconceptions when planning a lesson.

Notwithstanding the vast amount of research in the field of teachers' expertise and the diverse approaches to investigate science teachers' PCK (Chan & Hume, 2019), there is a lack of research on what aspects of their students' understanding in science they take into consideration when planning a lesson in an authentic context (Weingarten, 2019). So far, research focusing science teachers' considerations and implementations of students' conceptions has assessed teachers' views basically in

---

L. Großmann (✉) · D. Krüger  
Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin, Germany  
e-mail: [leroy.grossmann@fu-berlin.de](mailto:leroy.grossmann@fu-berlin.de); [dirk.krueger@fu-berlin.de](mailto:dirk.krueger@fu-berlin.de)

© The Author(s), under exclusive license to Springer Nature  
Switzerland AG 2022

K. Korfiatis, M. Grace (eds.), *Current Research in Biology Education*,  
Contributions from Biology Education Research,  
[https://doi.org/10.1007/978-3-030-89480-1\\_14](https://doi.org/10.1007/978-3-030-89480-1_14)

181

interview-based research designs with rather small sample sizes (Lin, 2016; Lucero et al., 2017; Moodley & Gaigher, 2019). To the best of our knowledge, there is no exploration of authentic biology teachers' lesson plans and their rationales provided focusing on the challenges that occur when using students' understanding in science for instructional decisions. To close this gap, our study provides further representative insights into the PCK dimension *Students' Understanding in Science* and its relatedness to *Instructional Strategies* (Park & Oliver, 2008).

## 14.2 Theoretical Background

### 14.2.1 *Biology Teachers' Enacted Pedagogical Content Knowledge in Lesson Planning (ePCK<sub>p</sub>)*

Teachers have to access knowledge that helps them to plan and conduct lessons that foster students' competencies as effectively as possible. For that particular knowledge, Shulman (1986) introduced the notion of *pedagogical content knowledge* (PCK) as the knowledge base that distinguishes teachers from other professions. In spite of differences between various conceptualizations of PCK in detail (Neumann, Kind and Harms, 2019), the Refined Consensus Model reflects the current representation of PCK that is held in the science educational community (Carlson & Daehler, 2019). For the purpose of this paper, we explicitly focus on the enacted pedagogical content knowledge during planning (ePCK<sub>p</sub>), i.e.

...the knowledge in action generated during, and made visible in, science teachers' planning. (Alonzo et al., 2019, p. 274).

There is consensus that knowledge about students' needs and their prior understanding is a constitutive part of ePCK<sub>p</sub> (e.g. Gess-Newsome, 2015). For instance, in their Pentagon model, Park and Oliver (2008) stress the importance of interplay between the five PCK dimensions *Orientations to Teaching Science*, *Science Curriculum*, *Assessment of Science Learning*, *Students' Understanding in Science* and *Instructional Strategies*. It was shown that particularly knowledge of students' understanding (including motivation and interest, learning difficulties, need and students' conceptions), and knowledge of instructional strategies, were central in the teaching practice of (biology) teachers (Park & Chen, 2012; Reynolds & Park, 2020).

These findings support Shulman's (1986) theoretical assertion and are in line with later empirical findings (e.g. Aydin & Boz, 2013) that those components are most crucial for science teachers' PCK development. However, conflicting findings which showed hardly any influence of teachers' assertions about their students' understanding on the choice of instructional strategies (Walan et al., 2017), raise the question about how far science teachers actually consider students' understanding for instructional decisions. Research on teachers' planning skills has shown that particularly the consideration of students' cognitive abilities and the anticipation of learning difficulties is challenging for teachers (e.g. Borko & Livingston, 1989;

Westerman, 1991). Therefore, it is important to further investigate this very interplay in detail and based on authentic non-experimental enacted planning of in-service teachers (Weingarten, 2019).

### 14.2.2 Leveraging Students' Conceptions for Instruction

A variety of aspects (tasks, activities etc.) needs to be considered when planning a lesson that is intended to be tailored to students' understanding in science. In current PCK studies, aspects like motivation and interest are barely focused upon, whereas particularly students' conceptions are regarded as crucial for engaging students in science classes and designing appropriate learning processes (Duit et al. 2012; Larkin, 2012).

Yet, as Krell (2020) points out, the notion of "conception" itself is ambiguously defined. In a constructivist sense, it is defined as the internal representation of an entity shaped by external representations shaped by other people. This view leads directly to the conceptual change theory according to which students are confronted with a phenomenon, become dissatisfied with their prior conceptions, and are provided an alternative, plausible and fruitful conception that can permanently or temporarily replace or complement their conception of the world (Duit & Treagust, 2012).

Diagnosing students' conceptions would require teachers' willingness to deliberately include the learning context and base instruction on those conceptions. Hence, Lin (2016) states that teachers can plan their instruction more effectively if they gain insights into their students' minds. Therefore, teachers are expected not only to be aware of the nature of their students' conceptions (Lin, 2016), but also of their sources (Taber & Tan, 2011). Yet, this is often not the case (Lucero et al., 2017). In their analysis of novice science teachers' teaching practice, Windschitl, Thompson and Braaten (2011) distinguish three levels of working with students' ideas: (1) *monitoring, checking, reteaching ideas*, i.e. not considering or working with students' ideas; (2) *eliciting students' understanding*, i.e. considering students' ideas but not using them for instructional design; (3) *referencing students' ideas and adapting instruction*, i.e. shaping instruction based on particular students' ideas.

Beyond that, even if students' conceptions are considered by teachers when planning a lesson, the practical implementation of instructional strategies is often difficult for teachers. This might be because they do not apply useful assessment tools to elicit students' understanding (Morrison & Lederman, 2003), or because teachers tend to unilaterally focus on transmitting subject matter knowledge rather than working with students' conceptions (Moodley & Gaigher, 2019).

Additionally, experience-based concepts and academic concepts are regarded as distinct, and thus neither one of them is considered instead of using their synergies (Otero & Nathan, 2008). Moreover, teachers prefer to restate their own understanding of a scientific topic instead of representing the content for students' learning, even if they believe to have insights into their students' thinking (Halim & Meerah, 2002).



## 14.3 Research Questions

The present study contributes to research on in-service trainee teachers' PCK by gaining further insights into the planning processes of biology lessons.

Two research questions are addressed:

RQ 1: Which facets of knowledge about *Students' Understanding in Science* as part of their ePCK<sub>p</sub> are most taken into account by German trainee teachers in written biology lesson plans?

RQ 2: To what extent do German trainee teachers align their planned instructional strategies in biology lessons to their analysis of students' conceptions?

## 14.4 Research Design and Methods

### 14.4.1 Data Material and Collection

The sample consists of 107 biology lesson plans (for students in grades 5–13) that were obtained from the Ministry for Education, Youth and Family in a federal state of Germany in an anonymized form. They were written by German trainee teachers during their final exam in the second phase of teacher education. German teacher education is organized in two phases: the first pre-service phase consists of 3 years of Bachelor's study and 2 years of Master's study at university; the second phase consists of 1.5 years of in-service teacher training (Neumann et al., 2017). It is a complete sample since it includes all lesson plans of trainee teachers who finished their training in the years 2018/2019. Therefore, those lesson plans can be regarded as representative indicators of the state-of-the-art of biology lesson planning in this federal state. Since those lesson plans have been written in the course of a formal examination, they had to follow certain structural regulations. They contain a description of the learning objectives, an analysis of the subject matter, an analysis of the group of students, a description of the planned lesson including rationales for planning decisions (activities, tasks, methods, etc.) and materials (e.g. worksheets including solutions). There were no more concrete instructions for writing the lesson plan. Due to data protection we were not allowed to collect further data such as grade in the exam, gender, teaching experience, interviews and so forth.

### 14.4.2 Data Analysis

RQ 1: Biology teachers' ePCK<sub>p</sub>

The lesson plans were analyzed within the methodological framework of qualitative content analysis (Elo et al., 2014; Schreier, 2012). We applied a mainly

**Table 14.1** Categories of trainee teachers' analysis of *Students' Understanding in Science*

Category	Subcategories	Examples
1. Cognition	A. prior knowledge	<i>My students know the terms natural and sexual selection as well as the related driving factors. [...] They do not know about the interplay of those factors yet.</i> (BT43, p. 4)
	B. students' conceptions	<i>My students understand evolution basically as a teleological process that does not continue once the intended goal is achieved.</i> (BT43, p. 8)
	C. learning difficulties	<i>As it turned out repeatedly, my students struggle to distinguish the essential from the non-essential information in texts. The more abstract the topic, the more problems they have.</i> (BT87, p. 4)
2. Methodological knowledge		<i>Some of my students have already participated in panel discussions concerning bioethical topics like the use of glyphosate.</i> (BT35, p. 3)
3. Motivation/interest		<i>My students told me that they like experimenting in groups.</i> (BT97, p. 5)
4. Social aspects		<i>Some of my students lack the ability to collaborate, which can be the source of severe disturbances in cooperative learning.</i> (BT21, p. 4)

BT biology teacher

deductive approach derived from the above-mentioned Pentagon model of PCK (Park & Oliver, 2008), i.e. the categories 1B, 1C and 3 (Table 14.1) have been directly adopted from this model. Category 4 was included since it is an important aspect in adaptive teaching (Hardy et al., 2019). Two categories have been added inductively to fully cover the trainee teachers' considerations of their *Students' Understanding in Science*, i.e. to ensure empirical validity (Schreier, 2012): Prior Knowledge (1A) is often referred to in order to explain the selection of content, and Methodological Knowledge (2) describes students' knowledge about and experiences with certain activities and materials.

For coding, the software MAXQDA 20 was used. The first author coded 20% of the sample twice to ensure reliability (intrarater-agreement), all lesson plans were coded by the first author and a student assistant to ensure objectivity (inter-rater agreement). Cohen's Kappa indicated very high intra-rater agreement ( $M_k = 0.85$ ) and moderate to very high inter-rater agreement ( $M_k = 0.76$ ; Wirtz & Caspar, 2002).

RQ 2: Leveraging students' conceptions for instruction.

To gain further insight into how biology teachers align their instructional planning to their students' conceptions, two cases that include students' conceptions and address the same conception have been purposefully selected (Patton, 1990) to compare their ePCK<sub>p</sub> and deduce similarities and differences. Therefore, rationales for the use of instructional strategies referring to students' conceptions are analyzed.

## 14.5 Results

### 14.5.1 Science Teachers' ePCK<sub>p</sub> [RQ 1]

Table 14.2 illustrates the distribution of codings in the lesson plans that include information from the six dimensions described above.

Nearly all trainee teachers analyze their students' prior knowledge and approximately two thirds refer to methodological, motivational and social aspects. Strikingly, students' conceptions are barely considered.

On average, prior knowledge was addressed eight times per lesson plan whereas all the other categories were coded approximately twice per lesson plan (if they were considered at all). Further statistical analyses revealed no significant differences between school types, grades or topics.

### 14.5.2 Leveraging Students' Conceptions for Instruction [RQ 2]

Three of those six trainee teachers considering students' conceptions only speculated on their students' ideas, two used empirical findings from literature without concrete application to their own students and only one explicitly used a strategy to elicit students' thinking. However, neither types nor sources of students' conceptions are analyzed in those lesson plans. Two lesson plans in the sample focus on the same topic and refer to students' conceptions, specifically to the idea that ecosystems are static instead of dynamic. The following sections compare those two cases in more detail.

#### Case 1: The lake (Biology Teacher: BT87)

In this lesson, students are intended to learn how it is possible that there are constantly sufficient minerals in the epilimnetic zone of a lake even though they tend to sink down into the hypolimnetic zone. Right at the beginning, BT87 claims that

**Table 14.2** Quantity of codings per lesson plan addressing facets of *Students' Understanding in Science*

Category	Lesson plans referring to the facets at least once (N = 107)	Relative frequency
1.A. prior knowledge	106	99.1%
1.B. students' conceptions	6	6.5%
1.C. learning difficulties	23	21.5%
2. methodological knowledge	75	70.1%
3. motivation/interest	71	66.4%
4. social aspects	73	68.2%

solving ecological problems requires systems thinking skills, yet there is no analysis at all in how far students have already experienced systems thinking. Instead, the focus lies on what activities particular students are familiar with and on how motivating this is for them:

Leonhard is a friendly student who has sophisticated knowledge in biology [...]. However, he does not participate in lessons if writing is necessary. [...] He tends to interrupt constantly when he has to work individually. Therefore, I prefer group work since due to the interdependencies, he will feel more responsible. [...] I decided to let them produce a graphic representation which requires little writing effort. (BT87, p. 3)

It is claimed that instructional decisions like group work and avoidance of writing are made in order to specifically support certain students in their individual learning abilities. However, analyses of individual students' requirements are usually used to legitimise instructional decisions for the whole group of students rather than for individual students. Hence, the instructional design often only appears to be differentiated.

As far as the relevance of systems thinking and the consideration of students' conceptions are concerned, BT87 refers to the importance of analyzing cyclic cause-and-effect relationships as well as possibly occurring side effects. In this context, empirical research is consulted to point out that 16/17-year-old students perceive ecosystems as static systems:

At the beginning of the teaching unit, I analyzed my students' prior knowledge (Hammann & Asshoff, 2014, S.194) [...]. I diagnosed that my students have similar conceptions. Since ecosystems change even without a harmful human influence, it is important to address this topic in biology classes. (BT87, p. 3)

Two aspects are noteworthy here. First, an instructional decision is made based on the consideration and use of empirical findings. However, this is only used as a rationale for the choice of topic, not for the instruction itself. Second, neither concrete misconceptions are named nor any kind of differences within the group of students are mentioned. Hence, all instructional decisions to follow refer to the whole group instead of addressing the needs of particular subgroups of students.

The reasons for the structure of the lesson are only partly related to the extensive analysis of *Students' Understanding in Science*:

At the beginning, I present the question and ask students to hypothesize about it. The purpose is to reduce linguistic inhibition and activate all students in the class. (BT87, p. 5)

Interestingly, there is neither any reference to the importance of hypotheses as far as scientific reasoning competencies are concerned, nor are those hypotheses used in a diagnostic way to tailor the further learning process based on the variety of students' ideas.

To summarize, BT87 has analyzed many aspects of his/her students' understanding in science and provides a clear insight into the classroom situation. However, most of those aspects are not used in the planning of instruction. Particularly, the claimed consideration of students' conceptions is not reflected in the tasks, the activities or in the learning outcome.

*Case 2: Succession of plants in the Berlin Wall death strip* (Biology Teacher: BT60)

In this lesson, students are intended to learn about ecological succession of plants. The context chosen is the Berlin Wall death strip that has been colonized by many plant species since the German reunification in 1990. To justify the choice of topic, BT60 claims two relevant things right at the beginning:

In the last unit, students experienced how ecosystems work by the example of forests. Now, we will deal with changes in ecosystems over time. Even though students generally notice changes in their environment, natural ecosystems are regarded as relatively static. (BT60, pp. 3–4)

First, the decision for the topic is made based on the prior knowledge about ecosystems in general. Second, BT60 refers to students' conceptions but without any reference to literature or without having diagnosed particular students' conceptions about ecosystems. Instead, this rationale rather appears more like an intuitive speculation than a valid diagnosis. In the following, since a variety of students' products is intended, the prerequisite for that is explained:

The classroom climate is positive and supportive. Problems and questions [...] can be presented and discussed in a constructive manner. The students work in groups in a friendly, respectful [...] way. (BT60, p. 3)

Based on that analysis, BT60 plans an instructional setting in which groups of students have to arrange descriptions and pictures of ten plants in a certain order based on how students imagine the chronological order in which those plants colonized the Berlin Wall death strip. Up to three groups have to present their ideas, the students then have to compare the different results and discuss their findings.

Afterwards, the students read informative texts about ecological succession of plants and compare the plants in early and late stages of this process. Based on this newly acquired knowledge, the students then get the opportunity to re-check their order systems from the beginning, change them if necessary and reflect on the criteria they used when ordering. Two reasons are given for that procedure. First, BT60 can diagnose whether the learning objective has been achieved, and second, the students themselves get some feedback on their ideas as well.

To summarize, BT60 has analyzed some aspects of his/her students' understanding in science and designed an instructional setting that is evidently based on students' initial conceptions. However, it should be noted that the setting (ten plants only) is rather strict so that the variety of students' conceptions cannot be visualized fully.

## 14.6 Discussion

This study aimed to identify which aspects of *Students' Understanding in Science* as one component of ePCK<sub>p</sub> are considered by German trainee teachers in written biology lesson plans, and how far these analyses have been used to design the learning process, particularly focusing the consideration of students' conceptions.

### **14.6.1 *Biology Teachers' ePCK<sub>p</sub> [RQ 1]***

The results show that trainee teachers rarely describe learning difficulties and students' conceptions. This could be explained by the fact that more long-term effort and in-depth analysis is needed to diagnose students' thinking, to analyze a topic thoroughly and to find certain difficulties that would need to be considered in the instructional setting (Hammann & Asshoff, 2014). It is far more convenient to describe the knowledge gained in the previous lessons as well as general methodological and social aspects that could be referenced for each lesson in that class anyway. Hence, those aspects could be regarded as personal PCK (pPCK) elements, i.e.

...the "reservoir of knowledge and skills that the teacher can draw upon during the practice of teaching" (Carlson & Daehler, 2019, p. 85).

But the more specific assertions are part of the ePCK<sub>p</sub> (Alonzo et al., 2019). However, it should be considered that trainee teachers might know about students' conceptions in general (pPCK), but don't recognize and acknowledge them in the enacted planning practice (ePCK<sub>p</sub>).

Moreover, about one third of the sample neglected to consider motivational aspects. This is surprising since sufficient motivation is considered crucial for students' learning (Hidi & Harackiewicz, 2000). This raises the question how trainee teachers perceive the task of writing a biology lesson plan. In most cases, they seem to be regarded as a bureaucratic proposition that has to be filled out tidily, rather than focusing on the aspects really needed to satisfy the unique needs of the particular group of students concerning the particular topic. Although it was not observed how the lesson plans were translated into teaching practice, we argue that this analysis can contribute valid insights into biology teachers' ePCK<sub>p</sub> since those lesson plans give a representative insight into the state-of-the-art lesson planning in this federal state. Since the lesson plans were not written under scientific guidance but in the course of formal examination, we could not control whether other persons (colleagues, friends etc.) have proofread and made contributions to those lesson plans. Therefore, we might have gained only partial insights into the ePCK<sub>p</sub> of individual trainee teachers rather than into collective PCK (cPCK) within a community of experts (Carlson & Daehler, 2019).

### **14.6.2 *Use of Students' Conceptions in Instruction Planning [RQ 2]***

Considering that those biology lesson plans have been produced in the course of the highest formal exam at the end of German teacher education (Neumann et al., 2017), it is remarkable that only 6 out of 107 trainee teachers referred to students' conceptions at all. This supports other empirical findings concerning the

insensitivity of science teachers to students' conceptions (Lin, 2016; Lucero et al., 2017) and raises the question for further research regarding why biology teachers apparently do not see the necessity to include students' thinking into their lesson planning processes. This is particularly striking since there is plenty of literature on students' conceptions and how to elicit and use them within the framework of conceptual change in instruction available in Germany (e.g. Hammann & Asshoff, 2014; Kattmann, 2016). Moreover, in the few lesson plans that include students' conceptions, neither types nor sources of students' conceptions are analyzed in those lesson plans confirming the claimed insensitivity of science teachers to students' conceptions (Lin, 2016; van Driel et al., 2002).

The comparison of two lesson plans about ecosystem dynamics illustrates the difference between a closed scenario in which the emphasis is put on learning certain subject matter (case 1) on the one hand, and providing the opportunity for students to put forward their initial ideas and reflect on them after an instructional input (case 2) on the other hand. The first case is in line with Halim and Meerah (2002), since BT87 actually diagnosed his/her students' ideas in advance but planned the lesson with more focus on the subject matter than on the students' thinking (Moodley & Gaigher, 2019). This corresponds to the second level in terms of Windschitl, Thompson and Braaten (2011). The second case can be understood as a partially responsive setting that uses formative assessment and a reflection phase on the learning process to some degree (Otero & Nathan, 2008). Unlike Walan et al.'s findings (2017), BT60 in case 2 explicitly referred to aspects of students' understanding and used it for an instructional decision. This might correspond with the highest level of working with students' ideas (Windschitl et al., 2011). However, we regard this example as a rather superficial analysis lacking an in-depth view of conceptions (Morrison & Lederman, 2003) that would allow an instructional setting to be associated with conceptual change (Duit & Treagust, 2012). This does not occur in the whole sample. Yet, as Weingarten (2019) points out, the analysis of lesson plans written in the context of a formal exam must consider the possibility that the trainee teachers' decisions are at least somehow influenced by what the examiners desire in the sense of symbolic interactionism (Mead, 1934). Consequently, trainee teachers might tend to avoid any risks in their planning and rather prefer whole-class discussions or teacher-centered consolidation phases to convey the biological content.

Since there are only six lesson plans considering students' conceptions in the complete sample, the scope of this study is limited. The selection of the two cases aimed to exemplify contrasting approaches in educational practices when working with students' conceptions. Broader research is needed to further investigate biology teachers' challenges and needs when leveraging students' conceptions for instruction on a larger sample of lesson plans that include students' conceptions. Therefore, we plan to explicitly ask pre-service teachers and trainee teachers to plan a lesson and take students' conceptions into account. By doing so, a much larger sample can be analyzed in a comparative case study.

## 14.7 Conclusion

In summary, this study reveals two major findings:

1. Trainee biology teachers analyze student groups basically on a general level, i.e. by merely naming declarative prior knowledge, methodological knowledge or motivational and social aspects. In-depth learning difficulties and students' conceptions concerning the topics to be taught are mostly neglected.
2. If in-depth aspects like students' conceptions are considered, they are hardly explicitly used for planning instruction, but rather named as considerations of minor relevance.

Both are astonishing since the lesson plans were produced in the course of the highest final exam in German teacher education after 3 years of Bachelor's study, 2 years of Master's study (both pre-service) and 1.5 years of in-service teacher training (Neumann et al., 2017). Hence, we might expect trainee teachers to show sophisticated and highly interconnected ePCK<sub>p</sub>. However, unlike constructivist learning theory and conceptual change theory suggest (Duit et al., 2012; Larkin, 2012), trainee teachers at the end of their teacher training either do not even consider students' conceptions or struggle with designing instruction based on students' conceptions. These findings are consistent with results from earlier studies (Lucero, 2017; Morrison & Lederman, 2003; Moodley & Gaigher, 2019; Otero & Nathan, 2008).

It should be noted that in most lesson plans, trainee teachers used differentiated instruction and obviously tried to include their analysis of their students' needs into the instructional design, e.g. by providing bonus tasks for high performing students or further help for low performing students (Hardy et al., 2019). Hence, trainee teachers do take their students' abilities and needs into account, but they fail to fundamentally design instructional processes based on their students' conceptions.

The present study demonstrates that prospective biology teachers need more support in eliciting students' conceptions in an appropriate manner and using them explicitly in the process of designing instruction. Teacher training programmes need to actively and repeatedly emphasize the importance of students' conceptions in biology classes. Pre-service teachers and trainee teachers need help to bridge the gap between theoretical assertions (e.g. conceptual change theory; Duit et al., 2012) as part of their PCK and practical strategies to elicit and use students' conceptions in planning and teaching biology.

**Acknowledgments** The project *K2Teach* is part of the "*Qualitätsoffensive Lehrerbildung*", a joint initiative of the German Federal Government and the Länder, which aims to improve the quality of teacher training. The programme is funded by the Federal Ministry of Education and Research (grant number 01JA1802). The authors are responsible for the content of this publication.



## References

- Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 271–286). Springer.
- Aydin, S., & Boz, Y. (2013). The nature of integration among PCK components: A case study of two experienced chemistry teachers. *Chemistry Education Research and Practice*, 14(4), 615–624.
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473–498.
- Carlson, J., & Daehler, K. R. (2019). The Refined Consensus Model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–92). Springer.
- Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: literature review of how science teachers' Pedagogical Content Knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_1](https://doi.org/10.1007/978-981-13-5898-2_1)
- Duit, R., & Treagust, D. F. (2012). How can conceptual change contribute to theory and practice in science. In B. J. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 107–118). Springer.
- Duit, R., Gropengießer, H., Kattmann, U., Komorek, M., & Parchmann, I. (2012). The model of educational reconstruction – A framework for improving teaching and learning science. In D. Jorde & J. Dillon (Eds.), *Science education research and practice in Europe: Retrospective and prospective* (pp. 13–37). Sense Publishers.
- Elo, S., Kaariainen, M., Kanste, O., Polkki, T., Utriainen, K., & Kyngas, H. (2014). Qualitative content analysis. *SAGE Open*, 4, 1–10.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, P. J. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Halim, L., & Meerah, S. M. M. (2002). Science trainee teachers' pedagogical content knowledge and its influence on physics teaching. *Research in Science & Technological Education*, 20(2), 215–225.
- Hammann, M., & Asshoff, R. (2014). *Schülervorstellungen im Biologieunterricht. Ursachen für Lernschwierigkeiten* [Students' conceptions in biology classes. Causes of learning difficulties]. Stuttgart.
- Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online*, 11(2), 169–191.
- Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70(2), 151–179.
- Kattmann, U. (2016). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht; zusätzliche Stichwörter zum Download* [Better understanding students: Students' conceptions in biology classes]. Aulis Verlag.
- Krell, M. (2020). Vorstellung und Kompetenz [Conception and competence]. In B. Reinisch, K. Helbig, & D. Krüger (Eds.), *Biologiedidaktische Vorstellungsforschung: Zukunftsweisende Praxis* (pp. 69–82). Springer Spektrum.
- Larkin, D. (2012). Misconceptions about misconceptions: Preservice secondary science teachers' views on the value and role of student ideas. *Science Education*, 96(5), 927–959.
- Lin, J. W. (2016). Do skilled elementary teachers hold scientific conceptions and can they accurately predict the type and source of students' preconceptions of electric circuits? *International Journal of Science and Mathematics Education*, 14(2), 287–307.

- Lucero, M. M., Petrosino, A. J., & Delgado, C. (2017). Exploring the relationship between secondary science teachers' subject matter knowledge and knowledge of student conceptions while teaching evolution by natural selection. *Journal of Research in Science Teaching*, 54(2), 219–246.
- MAXQDA. (n.d.). (Version 20) [Computer software]. VERBI Software GmbH.
- Mead, G. H. (1934). *Mind, self and society* (Vol. 111). University of Chicago Press.
- Moodley, K., & Gaigher, E. (2019). Teaching electric circuits: Teachers' perceptions and learners' misconceptions. *Research in Science Education*, 49(1), 73–89.
- Morrison, J. A., & Lederman, N. G. (2003). Science teachers' diagnosis and understanding of students' preconceptions. *Science Education*, 87, 849–867.
- Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. Pedersen, T. Isozaki, & T. Hirano (Eds.), *Model science teacher preparation programs* (pp. 29–52) IAP.
- Neumann, K., Kind, V., & Harms, U. (2019). Probing the amalgam: the relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, 41(7), 847–861.
- Otero, V. K., & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. *Journal of Research in Science Teaching*, 45, 497–523.
- Park, S., & Chen, Y. C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941.
- Park, S., & Oliver, J. S. (2008). National board certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45(7), 812–834.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Sage.
- Reynolds, W. M., & Park, S. (2020). Examining the relationship between the Educative Teacher Performance Assessment and preservice teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*. <https://doi.org/10.1002/tea.21676>
- Schreier, M. (2012). *Qualitative content analysis in practice*. Sage.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Taber, K. S., & Tan, K. C. D. (2011). The insidious nature of 'hard-core' alternative conceptions: Implications for the constructivist research programme of patterns in high school students' and pre-service teachers' thinking about ionization energy. *International Journal of Science Education*, 33(2), 259–297.
- Van Driel, J. H., Jong, O. D., & Verloop, N. (2002). The development of preservice chemistry teachers' pedagogical content knowledge. *Science Education*, 86(4), 572–590.
- Walan, S., Nilsson, P., & Mc Ewen, B. (2017). Why inquiry? Primary teachers' objectives in choosing inquiry-and context-based instructional strategies to stimulate students' science learning. *Research in Science Education*, 47(5), 1055–1074.
- Weingarten, J. (2019). *Wie planen angehende Lehrkräfte ihren Unterricht? Empirische Analysen zur kompetenzorientierten Gestaltung von Lernangeboten* [How do prospective teachers plan their lessons? Empirical analyses on competency-oriented instructional design]. Waxmann Verlag.
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292–305.
- Windschitl, M., Thompson, J., & Braaten, M. (2011). Ambitious pedagogy by novice teachers: Who benefits from tool-supported collaborative inquiry into practice and why? *Teachers College Record*, 113(7), 1311–1360.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität* [Rater agreement and rater reliability]. Hogrefe.

## Beitrag 5

**Großmann, L.** & Krüger, D. (in Druck). Erkenntnisgewinnung (v)erklärt. Wie plant man hypothesengeleiteten Biologieunterricht? In: B. Reinisch, D. Krüger & D. Mahler (Hrsg.), *Biologiedidaktische Nature-of-Science-Forschung: Zukunftsweisende Praxis*. Berlin: Springer.

Das Werk wird unter CC-BY-4.0-Lizenz erscheinen (<https://creativecommons.org/licenses/by/4.0/deed.de>). Der hier veröffentlichte Beitrag entspricht einer früheren Fassung des Manuskripts, das für die Publikation im o. g. Sammelband noch inhaltlich und formal überarbeitet wurde.

# Erkenntnisgewinnung (v)erklärt

Wie plant man hypothesengeleiteten Biologieunterricht?

Leroy Großmann & Dirk Krüger

## Zusammenfassung

Die Hypothese bildet einen zentralen Bestandteil im Rahmen naturwissenschaftlicher Erkenntnisprozesse. Ihr kommt als kognitiv-epistemischer Aspekt eines elaborierten *nature of science*-Verständnisses eine große Bedeutung zu. Der Begriff ist jedoch doppeldeutig: So wird der Begriff *Hypothese* einerseits sachangemessen für deduktiv aus einer soliden Wissensbasis abgeleitete, begründete und empirisch prüfbare Voraussagen eines Untersuchungsergebnisses genutzt. Wir schlagen vor, den andererseits in der Schule verwendeten Begriff *Hypothese* für vorläufige, häufig vage Ideen, mit denen man sich ein Phänomen zu erklären versucht, durch *Erklärung* oder *Erklärungsversuch* zu ersetzen. Die Nutzung des Hypothesenbegriffs, einerseits mit dem Schwerpunkt im Kompetenzbereich Erkenntnisgewinnung über Hypothesen zu reflektieren, und andererseits mit Sachkompetenz eine fachangemessene Erklärung zu finden, wird in diesem Beitrag anhand einer exemplarischen Analyse zweier Staatsexamensentwürfe im Fach Biologie illustriert. Abschließend werden Perspektiven für die empirische fachdidaktische Forschung aufgezeigt.

## Abstract

The hypothesis is a central component of scientific inquiry. It is of great importance as a cognitive-epistemic aspect of an elaborated understanding of nature of science. However, the term is ambiguous: On the one hand, the term *hypothesis* is appropriately used for empirically testable predictions that are deductively derived from a solid knowledge base. On the other hand, the term is used in school for initial, often vague ideas with which one attempts to explain a phenomenon. In the latter case, we propose to replace the term *hypothesis* by *explanation* or *explanation attempt*. In this paper, we illustrate the use of *hypothesis* by an exemplary analysis of two lesson plans written by trainee biology teachers. Finally, perspectives for empirical educational research are pointed out.

# 1. Einführung

Möchte man im alltäglichen Sprachgebrauch ausdrücken, dass man annimmt, dass etwas so oder so sein könnte, wird dies oft als *hypothetisch* attribuiert (Kattmann 2015). Gemeint ist damit in der Regel eine vage Annahme darüber, wie ein Phänomen zu erklären sei. Eine solche „Hypothese“ könnte zutreffen, sie kann ebenso falsch sein. Im Gegensatz zur Alltagssprache versteht die Wissenschaft unter dem Begriff *Hypothese* (gr. *hypóthesis*, d. h. Unterstellung) eine begründete Vermutung über Unterschiede, Zusammenhänge oder Veränderungen über die Zeit zwischen mindestens zwei Sachverhalten oder Variablen, die gerichtet oder ungerichtet aus Theorien abgeleitet werden (Döring und Bortz 2016).

Innerhalb der Didaktik der Biologie wird folgendermaßen definiert: Hypothesen sind „*empirisch überprüfbar, grundsätzlich widerlegbar, eindeutig und widerspruchsfrei [und] werden stets durch Vorwissen bzw. theoretische Konstrukte (z.B. Theorien, Regeln, Modelle) begründet*“ (Krell und Krüger 2022). Die Polysemie des Begriffs stellt Schüler\*innen und Lehrende vor eine Herausforderung, schließlich zielt der Begriff *Hypothese* im Sinne der *Erkenntnisgewinnungskompetenz* (KMK 2020) darauf ab, theoriegeleitet Hypothesen aufstellen zu können, diese empirisch zu testen und damit zu widerlegen oder zu stützen. Damit Schüler\*innen diese Kompetenz entwickeln können, sollen Lehrende über das für die Vermittlung notwendige fachliche und fachdidaktische Professionswissen verfügen, um entsprechende wissenschaftspropädeutische Aspekte in ihrem Unterricht sachgerecht behandeln zu können. Das allerdings fällt sowohl angehenden (Gyllenpalm und Wickman 2011) als auch erfahrenen Lehrkräften (Capps und Crawford 2013) oft schwer. Wenn Schüler\*innen einen reflektierten Umgang mit Hypothesen im naturwissenschaftlichen Sinne erlernen sollen (KMK 2020), wäre es wichtig, dass Biologielehrkräfte und ihre Auszubildenden in der ersten und zweiten Phase selbst über ein elaboriertes Begriffsverständnis verfügen und den wissenschaftlichen *Hypothesen*-Begriff nicht mit dem unscharfen Alltagsbegriff vermischen. Allerdings findet sich die Gleichsetzung des Hypothesen-Begriffs mit der alltagssprachlichen Bedeutung sowohl in naturwissenschaftsdidaktischer Literatur (z. B. Gyllenpalm und Wickman 2011) als auch in Staatsexamensentwürfen von Biologie-Referendar\*innen. Für die Planung guten Biologieunterrichts im Sinne der Förderung von Kompetenzen im Bereich Erkenntnisgewinnung erscheint es daher notwendig, das begriffliche Instrumentarium im Umgang mit Hypothesen auszuscharfen. Wir schlagen vor, den Begriff *Hypothese* nur für begründete, empirisch prüfbare Voraussagen zu nutzen, die aus einer theoretischen Basis heraus deduziert wurden. Wenn jedoch in einem problemorientierten

Einstieg Ideen gesammelt werden, wie ein Phänomen zu erklären sei, sollten diese Ideen nicht *Hypothesen*, sondern *Erklärungen* oder *Erklärungsversuche* genannt werden. Die folgenden zwei Thesen werden im vorliegenden Beitrag diskutiert:

1. Der Begriff *Hypothese* wird im Biologieunterricht häufig missverständlich verwendet, denn statt als epistemisches Konzept im Sinne einer Voraussage eines Untersuchungsergebnisses (Krell und Krüger 2022) wird er oft alltagssprachlich im Sinne einer vagen Erklärung verwendet. Daher sollten die Begriffe *Hypothese* (als begründete Voraussage aus Theorien deduziert) und *Erklärung* (vorläufig und aus Vorwissen abduziert) voneinander unterschieden werden.
2. Es sollen explizit prozedurale und epistemische Aspekte der Begriffe *Hypothese* und *Erklärung* reflektiert werden, um Kompetenzen im Bereich *Erkenntnisgewinnung* zu fördern (*learning about science*, Hodson 2014).

Im Round Table zu *nature of science & Praxis* wurde kritisch darüber reflektiert, welche Gestaltungsmerkmale ein Unterricht zur Förderung eines Verständnisses über *nature of science* aufweisen muss und welche Konsequenzen die Berücksichtigung dieser Merkmale für die Gestaltung von Biologieunterricht hat. Der vorliegende Beitrag ist demnach innerhalb des *Family Resemblance Approaches* als einer Konzeptualisierung von *nature of science* (Erduran & Daegher, 2014) dem *scientific knowledge* (Reinisch & Fricke, 2022) zuzuordnen.

Zunächst werden die Bedeutungen der Begriffe *Hypothese* und *Erklärung* einander gegenübergestellt. Darauf aufbauend werden exemplarisch zwei geplante Biologiestunden skizziert, die im Rahmen einer Analyse von Staatsexamensentwürfen (Großmann und Krüger 2022) untersucht wurden. Hierin werden einerseits die Herausforderungen deutlich, die angehende Biologielehrkräfte noch am Ende ihrer Ausbildung in Bezug auf den Umgang mit dem Begriff *Hypothese* haben. Andererseits wird kontrastiert, wie im Sinne von *nature of science* mit Erklärungen bzw. Hypothesen im naturwissenschaftlichen Unterricht gearbeitet werden kann und warum Wissen über diesen kognitiv-epistemischen Aspekt von *nature of science* (Reinisch & Fricke, 2022) für den Biologieunterricht relevant ist.

## 2. Diskurs

## 2.1. Eine *Hypothese* ist eine *Hypothese* ist keine *Erklärung*

Die Hypothese bildet insofern ein zentrales Konzept innerhalb des Kompetenzbereichs Erkenntnisgewinnung (z. B. KMK 2020), als sie als Ausgangspunkt naturwissenschaftlicher Forschungsprozesse eine vorläufige Antwort auf eine Forschungsfrage darstellt, die es durch eine Untersuchung zu überprüfen gilt (Krell und Krüger 2022; Lübeck 2020). Hypothesen sollten logisch aus Theorie hergeleitet werden und präzise sowie begründet einen Zusammenhang zwischen Variablen beschreiben, der empirisch getestet werden kann (Döring und Bortz 2016).

Mit einer Hypothese wird also begründet vorausgesagt, welches Ergebnis bei einer naturwissenschaftlichen Untersuchung (z. B. einem Experiment) erwartet wird. Damit kommt ihr im Rahmen des in den naturwissenschaftlichen Fächern weit verbreiteten hypothetisch-deduktiven Erkenntniswegs eine zentrale Rolle zu.

Diese vermeintliche Klarheit muss mit Blick auf den Diskurs in den Naturwissenschaften jedoch infrage gestellt werden: Keineswegs wird der Begriff *Hypothese* einheitlich im Sinne einer deduktiven, begründeten und in die Zukunft gerichteten Voraussage verstanden. Da die Annahme, dass Erklärungen und Voraussagen logisch betrachtet strukturgleich sind, nicht haltbar ist (Kornmesser und Büttemeyer 2020, S. 139), wirkt die Verbindung von Erklärung und Voraussage unter dem *Hypothesen*-Begriff überdenkenswert. McComas (2020) zeigt auf, dass dem *Hypothesen*-Begriff keineswegs ein einheitliches Verständnis zugrunde liegt und der Begriff gleichermaßen für *predictions* (d. h. deduktive Voraussagen) sowie in den Wendungen *generalizing hypotheses* (d. h. Erklärungen, die zu naturwissenschaftlichen Gesetzen werden können) und *explanatory hypotheses* (d. h. Erklärungen, die zu Theorien werden können) genutzt wird. Der Begriff *Hypothese* umfasst demnach mindestens drei Definitionen: “*For that reason, the term ‘hypothesis’ probably should be abandoned and replaced or at least used with caution*” (McComas 2020, S. 45).

Diesem Plädoyer schließen wir uns an. Wir schlagen vor, den Begriff *Hypothese* nur dann anzuwenden, wenn Voraussagen (*predictions*) über Zusammenhänge, Unterschiede oder Veränderungen messbarer Variablen deduktiv aus theoretischen, gut begründeten Aussagegefügen abgeleitet werden und ein zeitlich in der Zukunft liegendes Ergebnis postuliert wird. Für Ideen, mit denen man versucht, sich ein Phänomen ad hoc zu erklären, schlagen wir vor, von *Erklärungsversuchen* zu sprechen. Diese Versuche können in Bezug auf theoretische Grundlagen oder durch Analogiebildung gefunden oder beim Fehlen von entsprechenden Erfahrungen kreativ erfunden werden (Schurz 2008). Was diese Erklärungsversuche im Vergleich zu den in die Zukunft weisenden Hypothesen unterscheidet, ist, dass Erklärungen

potentielle Ursachen für das Zustandekommen eines Phänomens liefern, die vor dem bereits existierenden Phänomen liegen und zeitlich in die Vergangenheit reichen. Eine solche Schlussweise nennt man Abduktion (vgl. Hammann et al. in diesem Buch). Mit dem Begriff *Erklärungsversuch* soll deutlich werden, dass die gesammelten Erklärungen potentielle Ursachen benennen, deren empirische Bestätigung, dann über Hypothesen, noch aussteht.

Wie im Folgenden zu zeigen sein wird, werden die verschiedenen Lesarten des *Hypothesen*-Begriffs in der Unterrichtspraxis nicht unterschieden. Lehrkräfte nutzen ihn eher im Sinne eines didaktischen Lehr-Lern-Arrangements in Form problemorientierten Unterrichts und weniger im Sinne einer naturwissenschaftlichen Denkweise (Gyllenpalm und Wickman 2011). So zeigen Capps und Crawford (2013) beispielsweise, dass selbst erfahrene Lehrkräfte in der Überzeugung unterrichten, Erkenntnisgewinnungskompetenzen zu fördern, ohne dies jedoch eigentlich zu tun. Zwar wird oft praktisch gearbeitet, allerdings wird in den untersuchten Stunden von 26 *science*-Lehrkräften weder explizit noch implizit ein Verständnis über den naturwissenschaftlichen Erkenntnisgewinnungsprozess vermittelt. Auch eine Differenzierung zwischen *Hypothesen* und *Erklärungen* bzw. *Erklärungsversuchen* spielt dabei keine Rolle.

Krüger und Upmeyer zu Belzen (2021) zeigen am Beispiel des Modellierens die Bedeutung des abduktiven Schließens, also des Entwickelns von Erklärungen im naturwissenschaftlichen Erkenntnisprozess. Auf der Basis der entwickelten Ideen zur Erklärung eines Phänomens wird demnach ein gedankliches Modell entwickelt, aus dem wiederum deduktiv Hypothesen abgeleitet werden, die dann mit empirischen Arbeitsweisen getestet werden können. Erklärungsversuche und Hypothesen stehen hier also in einem zeitlichen und logischen Verhältnis zueinander. Zumindest für die Lehrkräftebildung, möglicherweise aber auch für die Forschung, besteht hierin Potential der Erweiterung des *Scientific Discovery as Dual Search*-Modells (Klahr und Dunbar 1988), dessen Potentiale für die biologiedidaktische Forschung von Hammann (2007) eingehend dargelegt wurden. Demnach bewegt man sich beim Lösen eines naturwissenschaftlichen Problems in zwei Räumen, dem Hypothesenraum und dem Experimentierraum. Zunächst wird im Hypothesenraum nach Hypothesen gesucht, die entweder aus dem Vorwissen oder aus vorhergehenden Untersuchungen abgeleitet werden können. Diese Hypothesen werden dann im Experimentierraum getestet und anschließend evaluiert, was zu einer neuen Suche nach Hypothesen im Hypothesenraum führen kann. Hier scheint implizit ebenfalls die doppelte Bedeutung des Hypothesen-Begriffs Anwendung zu finden, denn Klahr und Dunbar (1988) gehen offenbar auch nicht ausschließlich von einwandfrei formulierten, deduktiv hergeleiteten und begründeten *Hypothesen* aus, die im Hypothesenraum entwickelt werden, sondern beziehen auch initiale Erklärungsversuche mit



ein. Hier könnte es nützlich sein, einen dritten Raum, den *Erklärungsraum*, zu proklamieren, der der Suche im Hypothesenraum vorangestellt ist und bezüglich seiner Lage vor den „Hypothesenraum“ verortet wird. Vollmeyer und Burns (1999) nennen einen solchen dritten Raum „Modellraum“, der der Bedeutung des abduktiven Schließens beim Entwickeln (kreativer) Ideen zur Erklärung eines Phänomens Rechnung trägt und der besser als das Zwei-Räume-Modell (Klahr und Dunbahr, 1988) erklären kann, warum jemand bestimmte Hypothesen entwickelt und testet. Diese Differenzierung wäre vor allem auch mit Blick auf die Lehrkräftebildung von Bedeutung, da sie dazu beitragen könnte, die Doppeldeutigkeit des *Hypothesen*-Begriffs zu explizieren und (angehende) Lehrkräfte dabei zu unterstützen, die dem Kompetenzbereich *Erkenntnisgewinnung* innewohnenden Herausforderungen (Krell und Krüger 2022) zu bewältigen. Hypothesen ergeben sich aus Erklärungsversuchen und sind nicht dasselbe – daher sollten diese Konzepte begrifflich konsequent voneinander unterschieden werden.

Professionstheoretisch gesprochen handelt es sich für Lehrkräfte hierbei um eine komplexe Aufgabe, da für die Gestaltung wissenschaftspropädeutischen Unterrichts im Sinne von *nature of science* neben dem notwendigen fachlichen und fachdidaktischen Wissen auch Überzeugungen und Einstellungen notwendig sind, um kompetenzorientierten, hypothesengeleiteten Biologieunterricht zu planen, der dann in der Tat Kompetenzen im Bereich Erkenntnisgewinnung fördert. Zu den kognitiven Herausforderungen, die in der Planung zu bewältigen sind, zählen unter anderem die Formulierung klarer Lernziele, die Gestaltung von Aufgabenstellungen sowie die Entwicklung einer kohärenten Phasenstruktur (König et al. 2021). Die Verknüpfung dieser planungsrelevanten Elemente bringt in Bezug auf die Planung hypothesengeleiteten Biologieunterrichts Konsequenzen mit sich, die im Folgenden illustriert werden.

## **2.2. Erkenntnisse gewinnen über Erkenntnisgewinnung**

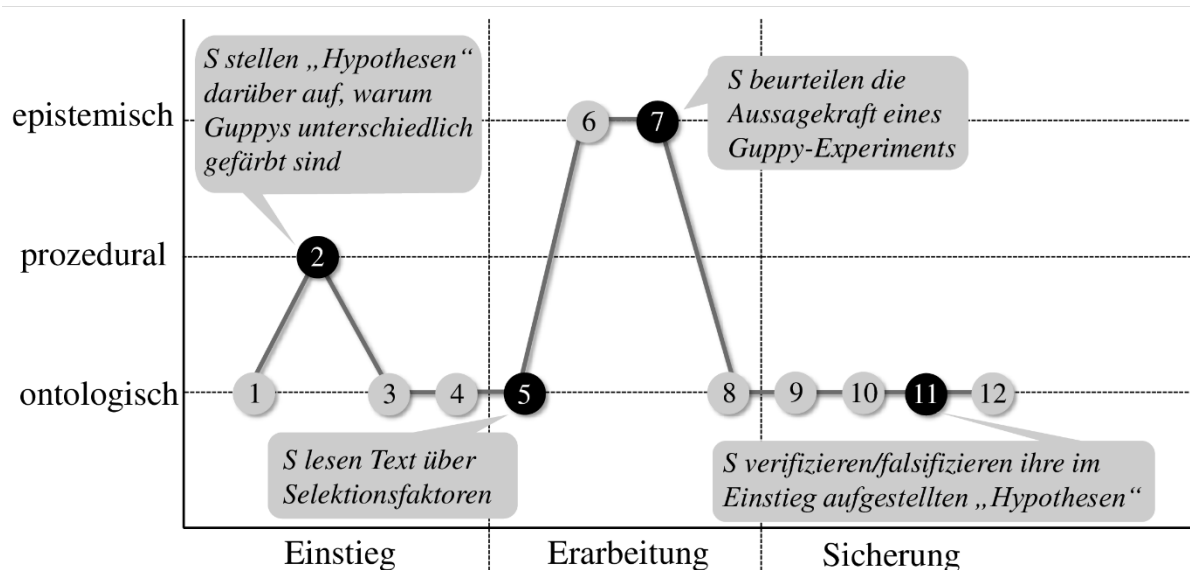
Um im Rahmen kompetenzorientierten Unterrichts eine Kompetenzentwicklung bei Schüler\*innen anzubahnen, sollten Lehrkräfte instruktionale Kohärenz herstellen (Neumann 2020), d. h. die Gestaltung des Lehr-Lern-Prozesses (z. B. Aufgabenstellungen, Phasenstruktur) auf die angestrebten Lernziele hin abstimmen. Diese Aufgabe ist kognitiv herausfordernd (König et al. 2021). Mit Blick auf den Kompetenzbereich *Erkenntnisgewinnung* (KMK 2020) ergibt sich die besondere Schwierigkeit, dass die Benennung des Kompetenzbereichs irritiert: *Erkenntnisgewinnung* meint nicht, mithilfe praktischer Tätigkeiten (z. B. Experimenten) neues Wissen über biologische Phänomene zu gewinnen, sondern Wissen über die

naturwissenschaftlichen Arbeits- und Denkweisen zu erwerben (Krell und Krüger 2022). Mit Hodson (2014) gesprochen meint Erkenntnisse gewinnen den Erwerb *ontologischen Wissens*, also biologische Sachkompetenz zu entwickeln (*learning science*). *Erkenntnisgewinnungskompetenz* entwickeln meint hingegen, einerseits *prozedurales Wissen* beim praktischen Anwenden naturwissenschaftlicher Arbeitsweisen, wie beim Experimentieren, umzusetzen (*doing science*). Andererseits umfasst Erkenntnisgewinnungskompetenz *epistemisches Wissen*, also Wissen über den naturwissenschaftlichen Erkenntnisprozess, wie zum Beispiel über die Kennzeichen einer wissenschaftlichen Hypothese in Abgrenzung zu (wissenschaftlichen) Erklärungen (*learning about science*), zu erwerben.

Die dargestellte Schwierigkeit zwischen Erkenntnisse gewinnen und Erkenntnisgewinnung zu unterscheiden scheint zu zwei gegensätzlichen Arten, Biologieunterricht zu strukturieren, zu führen. Für beide Arten wird im Folgenden jeweils ein Beispiel angeboten. Die Beispiele entstammen einer Analyse von Staatsexamensentwürfen im Fach Biologie und haben beide das Ziel, Erkenntnisgewinnungskompetenz zu fördern (Großmann und Krüger 2022). Ordnet man allen den Stundenplanungen zu entnehmenden einzelnen Unterrichtsschritten jeweils zu, ob sie ontologisches (d. h. Sachkompetenz) oder prozedurales bzw. epistemisches Wissen (d. h. Erkenntnisgewinnungskompetenz) adressieren, lässt sich veranschaulichen, in welchem Maße die Stunden instruktional kohärent die adressierten Kompetenzen fördern (Abb. 1; Abb. 2).

#### (1) *Erklärungen finden für fachliches Lernen*

Biologiestunden beginnen häufig damit, dass Schüler\*innen im Einstieg mit einem Phänomen konfrontiert werden und zu diesem Phänomen „*Hypothesen*“ formulieren sollen, die dann im Verlauf der Stunde überprüft und in der Sicherung bestätigt bzw. widerlegt werden (Abb. 1).



**Abb. 1.** Geplanter Verlauf einer Stunde zur Auswertung von Experimenten am Beispiel der Körperfärbung von Guppys. Von den 12 Unterrichtsschritten fokussieren nur die Schritte 2, 6 und 7 den adressierten Kompetenzbereich Erkenntnisgewinnung, die weiteren Aufgaben in der Erarbeitungsphase (Schritte 5 und 8) sowie der Einstieg (Schritte 1, 3 und 4) und die gesamte Sicherungsphase (Schritte 9-12) fokussieren den zu vermittelnden biologischen Fachinhalt, d. h. den Einfluss von natürlicher und sexueller Selektion auf die Färbung der Guppys (Sachkompetenz); S = Schüler\*innen.

In dieser Stunde wird das Ziel verfolgt, dass die Schüler\*innen ein Experiment zur Ausprägung der Schwanzflosse bei männlichen Guppys auswerten. Dazu wurde der Standard „Daten, Trends und Beziehungen interpretieren, diese erklären und weiterführende Schlussfolgerungen ableiten“ aus dem Berliner Rahmenlehrplan ausgewählt (SenBJF 2015, S. 19). Nachdem die Schüler\*innen im Einstieg zunächst mit einem Bild von zwei Guppys konfrontiert werden (Abb. 1, ①), sollen sie „Hypothesen“ aufstellen, warum die beiden Guppys unterschiedlich gefärbte Schwanzflossen haben (②). Als Begründung dafür wird der problemorientierte Unterricht genannt, demzufolge zunächst Problemfragen zu entwickeln und dann „Hypothesen“ zu formulieren seien. Über die Erwartung an diese „Hypothesen“ geben die drei antizipierten Antworten Aufschluss, die im Entwurf beschrieben werden:

*„Es handelt sich um den Sexualdimorphismus. Dabei liegen in Hinblick auf Gestalt, Färbung und Verhalten deutliche Unterschiede zwischen den Geschlechtern vor.“*

*„Es muss eine Mutation bei einem der Guppys vorliegen.“*

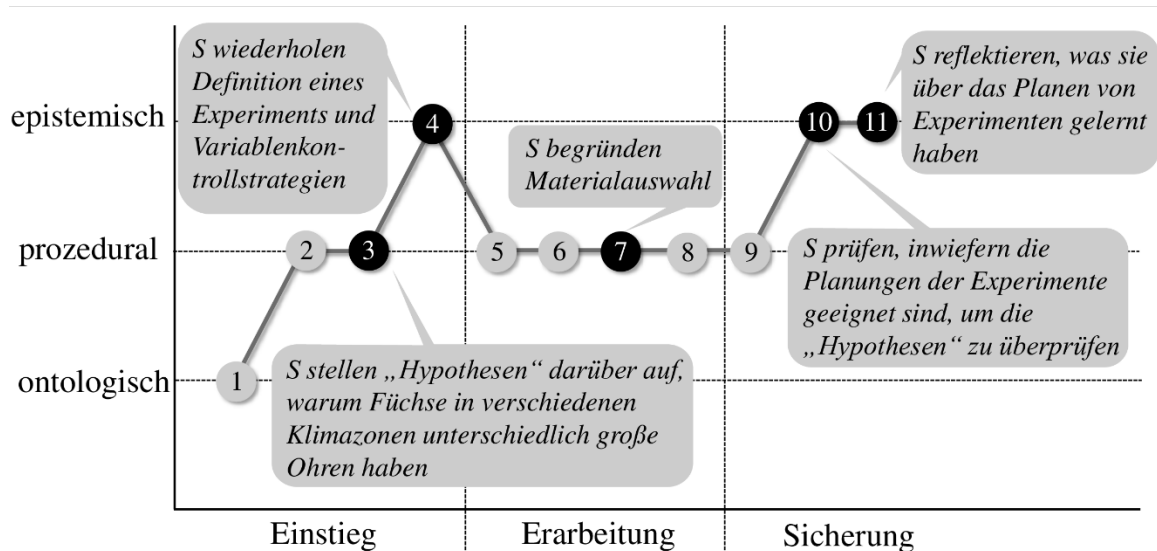
*„Die Guppys kommen an unterschiedlichen Orten vor.“*

Die Schüler\*innen sollen also mithilfe des bereits erarbeiteten biologischen Fachwissens über die Evolution nicht Hypothesen, sondern zeitlich vor dem Phänomen liegende Erklärungsversuche zum Phänomen formulieren. Diese Erklärungsversuche werden insofern geprüft, als die SuS in der Erarbeitungsphase zunächst einen Informationstext über Guppys und

verschiedene Selektionsfaktoren lesen sollen (5) und anschließend ein Experiment auswerten, in dem Guppys in zwei Teilgruppen getrennt gehalten werden – eine Gruppe mit, die andere als Kontrolle ohne einen Fressfeind,. Neben der inhaltlichen Auswertung des Experiments soll in einer weiteren Aufgabenstellung zudem die Aussagekraft des Untersuchungsplans und dessen Umsetzung beurteilt werden (7), so dass die Schüler\*innen sowohl mit der ontologischen als auch mit der epistemischen Ebene konfrontiert sind. Im letzten Schritt der Erarbeitung sollen die Schüler\*innen dann jedoch eine Kurzpräsentation vorbereiten, „indem Sie die Erkenntnisse aus der Auswertung der Versuchsergebnisse nutzen, um die Problemfrage zu beantworten“ (8). Damit wird der Unterrichtsgang wieder auf die ontologische Ebene zurückgeführt und schließt in sich logisch an den Einstieg an, denn nun werden die neu gewonnenen Erkenntnisse aus dem Arbeitsmaterial besprochen und geklärt, warum männliche Guppys unterschiedlich gefärbte Schwanzflossen haben (9, 10). Abschließend werden im Sinne einer didaktischen Klammer die im Einstieg formulierten Erklärungsversuche wieder aufgegriffen (11) und „reflektiert“. Es stellt sich die Frage, inwiefern die Reflexion über einen Vergleich der bereits gesicherten sachangemessenen Erklärung mit den im Einstieg gesammelten Erklärungsversuchen hinausgeht. Da die meisten Unterrichtsschritte in dieser Stunde der ontologischen Ebene zuzuordnen sind und dies auch Fokus der Sicherungsphase ist, ist davon auszugehen, dass die Schüler\*innen in dieser Stunde Sachkompetenzen und allenfalls Erkenntnisgewinnungskompetenzen implizit entwickelt haben. Wir vermuten, dass in der geplanten Stunde der angestrebte Kompetenzbereich hätte präzisiert werden können, wenn begrifflich konsequent statt von „Hypothesen aufstellen“ von „Erklärungen für ein fachliches Problem finden“ gesprochen und als Lernziel explizit die Förderung von Sachkompetenz benannt worden wäre. Somit würde vermieden werden, wissenschaftliche Hypothesen für vage, spontan geäußert Vermutungen zu halten.

## (2) *Hypothesen aufstellen für eine Untersuchung*

Sollen in einer Biologiestunde explizit Kompetenzen in der Erkenntnisgewinnung gefördert werden, sollten im Stundenverlauf die prozedurale oder epistemische Ebene konsequent erreicht werden (Abb. 2).



**Abb. 2.** Geplanter Verlauf einer Stunde zur Planung von Experimenten am Beispiel der Allen'schen Regel. Von den 11 Unterrichtsschritten fokussiert nur Schritt 1 als thematischer Aufhänger das biologische Fachwissen (Sachkompetenz). Darauf aufbauend wird Erkenntnisgewinnung adressiert, indem Hypothesen formuliert werden (Schritt 3), ein Experiment selbst geplant und aus bereitgestellten Materialien ausgewählt (Schritte 5-8) wird sowie letztlich reflektiert wird, inwiefern die geplanten Experimente tatsächlich zur Prüfung der Hypothesen geeignet sind (Schritte 10, 11); S = Schüler\*innen.

In dieser Stunde verfolgt die Lehrkraft das Ziel, dass die Schüler\*innen ein Experiment zur Überprüfung der Allen'schen Regel planen. Dazu wurde der Standard „exemplarische Experimente zur Überprüfung von Hypothesen planen“ aus dem Berliner Rahmenlehrplan ausgewählt (SenBJS 2006, S. 16).

Nachdem die Schüler\*innen im Einstieg mit einer Weltkarte konfrontiert werden (①), in der drei Fuchsarten mit verschieden großen Ohren gezeigt werden, werden die Schüler\*innen mit der Frage konfrontiert, warum beispielsweise der Wüstenfuchs viel größere Ohren hat als der Polarfuchs (③). Folgende „Hypothesen“ werden antizipiert:

*„Kleinere Ohren haben eine kleinere Oberfläche zur Wärmeabgabe, sodass sie in kälteren Gebieten einen Vorteil haben.“*

*„Wenn die Ohren größer sind, dann ist auch die Oberfläche zur Wärmeabgabe größer.“*

*Füchse, die in wärmeren Regionen leben, haben dadurch einen Vorteil.“*

Während die erste antizipierte „Hypothese“ eher einem *Erklärungsversuch* entspricht, in dem mithilfe von Vorwissen der Zusammenhang zwischen Oberfläche und Wärmeabgabe beschrieben wird, handelt es sich im zweiten Fall tatsächlich um eine deduktive Hypothese, für die eine „Wenn...dann“-Konstruktion gewählt und somit der Zusammenhang zwischen zwei Variablen präzise formuliert wird. Damit die Schüler\*innen vor dem Hintergrund ihrer in Gruppen formulierten Hypothesen entsprechende Experimente planen können, wird im Einstieg zunächst wiederholt, dass Experimente der Untersuchung eines kausalen

Zusammenhangs zwischen zwei Variablen dienen, dass man zu diesem Zweck einen Kontrollansatz benötigt und dass sich der Kontrollansatz vom Experimentalansatz nur in der Variation einer einzigen unabhängigen Variablen unterscheiden soll (4). Erweiternd hätte hier der Erklärungsversuch explizit von der Hypothese abgegrenzt werden können. Dieser Übergang auf die epistemische Ebene ist sinnvoll, weil die Schüler\*innen in der sich anschließenden Erarbeitungsphase eigenständig ein Experiment planen, mit dem ihre Hypothese untersuchbar wird. Dazu wählen sie aus bereitgestellten Materialien (z. B. Teelöffel, Esslöffel, warmen Kartoffeln) benötigte begründet aus (7). Die Schüler\*innen stellen Analogien zum Phänomen her (z. B. warme Kartoffel entspricht Fuchskörper, Löffel entsprechen den Ohren) und denken bewusst über die Gestaltung des Experiments nach. Die Experimentaldesigns (8) werden dann von zwei Gruppen präsentiert (9). Dazu wird ausgeführt:

*„Hierbei werden bewusst zwei Gruppen mit unterschiedlichen oder mit fehlerhaften Planungen ausgewählt, damit diese im Plenum ausgewertet und dezidiert besprochen werden können. Die restlichen [Schüler\*innen] erhalten einen Beurteilungsbogen, mit dem sie die Planungen einschätzen und reflektieren sollen“.*

Es wird deutlich, dass die Planung nicht auf die eine richtige Lösung abzielt. Stattdessen denken die Schüler\*innen auf einer epistemischen Ebene darüber nach, inwiefern die von den Gruppen vorgestellten Ansätze zur Untersuchung der im Einstieg formulierten Hypothesen geeignet sind (10). Der Beobachtungsauftrag durch den Beurteilungsbogen aktiviert kognitiv auch die Schüler\*innen, die ihre Planungen nicht vorstellen. Die Kompetenzfokussierung zeigt sich in einer abschließenden Selbstreflexion, in der die Schüler\*innen ihren eigenen Kompetenzzuwachs auf einer vierstufigen Likert-Skala einschätzen sollen (11), z. B. *„Ich habe meine eigene Planung des Experiments kritisch reflektiert“*. Dem Entwurf ist zu entnehmen, dass die auf der Grundlage des Feedbacks modifizierten Planungen dann in der folgenden Stunde experimentell umgesetzt werden und dort die Entwicklung von Sachkompetenz über die Allen'sche Regel im Vordergrund stehen wird.

### **3. Fazit und Ausblick**

Zusammenfassend zeigt sich, dass der *Hypothesen*-Begriff in der naturwissenschaftsdidaktischen Planungs- und Unterrichtspraxis angehender Biologielehrkräfte überstrapaziert wird. Es erscheint notwendig, die verschiedenen Lesarten des Begriffs klarer zu trennen. Wir schlagen vor, den Begriff *Hypothese* nur dann zu verwenden, wenn eine auf theoretischer Basis begründete, empirisch prüfbare Voraussage eines

in der Zukunft potentiell empirisch zu untersuchenden Zusammenhangs gemeint ist. Wenn die Schüler\*innen jedoch im Sinne eines problemorientierten Einstiegs Ideen sammeln, um Ursachen für ein Phänomen zu nennen, sollten diese Ideen nicht *Hypothesen*, sondern als potentielle und in der Vergangenheit liegende Ursachen für das Phänomen *Erklärungsversuche* genannt werden.

Dies könnte Anlass bieten, das Prinzip der Problemorientierung im Biologieunterricht insofern kritisch zu reflektieren, als ihm im Rahmen kompetenzorientierten Unterrichtens eine Schwierigkeit innewohnt: Wenn das Ziel verfolgt wird, ein Phänomen zu verstehen bzw. verständlich erklären zu können, wird Sachkompetenz (KMK 2020) adressiert. Oft jedoch formulieren (angehende) Lehrkräfte ein Lernziel, das auf Erkenntnisgewinnungskompetenz (KMK 2020) abzielt. Fokussiert man aber in der Sicherungsphase vor allem die fachliche Korrektheit der Erklärungen des biologischen Phänomens, liegt der Schwerpunkt auf dem Erwerb von Sachkompetenz. Erst ein explizites Nachdenken über „Basiskonzepte der Erkenntnisgewinnung“ (Lübeck 2020) bzw. über das „Methodenwissen“ der Biologie (Krell und Krüger 2022) fördert Erkenntnisgewinnungskompetenz. In solchen Stunden im Einstieg *Erklärungen* zu fordern, daraus *Hypothesen* abzuleiten und anschließend die Angemessenheit der Planung bzw. Durchführung einer Untersuchung zu reflektieren, wird einem problemorientierten Einstieg mit *Erklärungen* bzw. *Erklärungsversuchen* und *Hypothesen* gerecht. Die Gleichsetzung vager Ideensammlungen mit theoretisch begründeten Voraussagen unter dem Begriff „Hypothese“ erscheint uns hingegen problematisch.

Es sei mit Blick auf das erste Stundenbeispiel hervorgehoben, dass die Verfolgung des Standards des Berliner Rahmenlehrplans („*Daten, Trends und Beziehungen interpretieren, diese erklären und weiterführende Schlussfolgerungen ableiten*“) als fachliche Klärung interpretiert werden kann. Die Schwierigkeit besteht darin, dass die kurzen Standardformulierungen im Geiste des jeweiligen Kompetenzbereichs verstanden werden müssen.

Vor diesem Hintergrund sind die folgenden Fragen für die zukünftige biologiedidaktische Forschung von Interesse. Im Bereich der biologiedidaktischen...

... Kompetenzforschung wäre zu untersuchen, inwiefern sich die Auftrennung des *Hypothesen*-Begriffs in abduktive *Erklärungen* bzw. *Erklärungsversuche* und deduktive *Hypothesen* nicht nur beim Modellieren (Krüger und Upmeyer zu Belzen 2021) sondern auch bei den Arbeitsweisen Beobachten, Vergleichen und Ordnen sowie Experimentieren reflektieren lässt.

... Professionsforschung wäre zu untersuchen, welche Variablen der professionellen Kompetenz von Lehrkräften (z. B. Fachwissen, fachdidaktisches Wissen, epistemologische Überzeugungen) als Prädiktoren für ein wissenschaftlich adäquates Verständnis zur Unterscheidung von Erkenntnisgewinnungs- und Sachkompetenz oder spezieller von *Erklärungen* bzw. *Erklärungsversuchen* und *Hypothesen* gelten können.

## Literaturverzeichnis

- Capps, D.K., & Crawford, B.A. (2013). Inquiry-Based Instruction and Teaching About Nature of Science: Are They Happening? *Journal of Science Teacher Education*, 24(3), 497–526.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer.
- Erduran, S., & Dagher, Z. R. (2014). Reconceptualizing the nature of science for science education. Scientific knowledge, practices and other family categories. Springer.
- Großmann, L., & Krüger, D. (2022). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1).
- Gyllenpalm, J., & Wickman, P.-O. (2011). The Uses of the Term Hypothesis and the Inquiry Emphasis Conflation in Science Teacher Education. *International Journal of Science Education*, 33(14), 1993–2015.
- Hammann, M. (2007). Das Scientific Discovery as Dual Search-Modell. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (187-106). Berlin Heidelberg: Springer-Verlag Berlin Heidelberg.
- Hodson, D. (2014). Learning Science, Learning about Science, Doing Science: Different goals demand different learning methods. *International Journal of Science Education*, 36(15), 2534–2553.
- Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht*. Hallbergmoos: Aulis Verlag.
- Klahr, D., & Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. *Cognitive Science*, 12, 1–48.
- KMK (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife*. Carl Link.



- König, J., Krepf, M., Bremerich-Vos, A., & Buchholtz, C. (2021). Meeting Cognitive Demands of Lesson Planning: Introducing the CODE-PLAN Model to Describe and Analyze Teachers' Planning Competence. *The Teacher Educator*, 56(4), 466–487.
- Kornmesser, S. & Büttemeyer, W. (2020). Erklärungen und Theorien. In S. Kornmesser & W. Büttemeyer (Hrsg.), *Wissenschaftstheorie: Eine Einführung* (S. 135–150), Stuttgart: J.B. Metzler.
- Krell, M., & Krüger, D. (2022). Erkenntnisgewinnung im Biologieunterricht: Welches Methodenwissen soll vermittelt werden? *MNU Journal*(5), 376–382.
- Krüger, D., & Upmeyer zu Belzen, A. (2021). Kompetenzmodell der Modellierkompetenz – Die Rolle abduktiven Schließens beim Modellieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 27(1), 127–137.
- Lübeck, M. (2020). 'Basiskonzepte' der Erkenntnisgewinnung im Biologieunterricht: Ein Nachschlagewerk mit Aufgabenbeispielen. Münster: Waxmann Verlag GmbH.
- McComas (2020). Principal Elements of Nature of Science: Informing Science Teaching while Dispelling the Myths. In W. McComas (Hrsg.), *Nature of Science in Science Instruction: Rationales and Strategies* (S. 35–65). Cham: Springer International Publishing; Imprint Springer.
- Neumann, K. (2020). Die Bedeutung instruktionaler Kohärenz für eine systematische Kompetenzentwicklung. *Unterrichtswissenschaft*, 48(1), 1–10.
- Reinisch, B. & Fricke, K. (2022). Broadening a nature of science conceptualization: Using school biology textbooks to differentiate the family resemblance approach. *Science Education*, 106, 1375–1407.
- Schurz, G. (2008). Patterns of abduction. *Synthese*, 164, 201–234.
- SenBJF (2015). Rahmenlehrplan für die Jahrgangsstufen 7-10: Teil C (Biologie).  
[https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/rahmenlehrplaene/Rahmenlehrplanprojekt/amtliche\\_Fassung/Teil\\_C\\_Biologie\\_2015\\_11\\_10\\_WEB.pdf](https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/rahmenlehrplaene/Rahmenlehrplanprojekt/amtliche_Fassung/Teil_C_Biologie_2015_11_10_WEB.pdf).
- SenBJS (2006). Rahmenlehrplan für die gymnasiale Oberstufe: Biologie.
- Vollmeyer, R. & Burns, R. D. (1999). Problemlösen und Hypothesentesten. In H. Gruber, W. Mack & A. Ziegler (Hrsg.), *Wissen und Denken: Beiträge aus Problemlösepsychologie und Wissenspsychologie* (S. 101–118). Wiesbaden: Deutscher-Universitäts Verlag.

## Beitrag 6

**Großmann, L.** & Krüger, D. (2023). Identifying Performance Levels of Enacted Pedagogical Content Knowledge in Trainee Biology Teachers' Lesson Plans. In: G. S. Carvalho, A. S. Afonso & Z. Anastácio (Hrsg.), *Fostering Scientific Citizenship in an Uncertain World. Selected Papers from the ESERA 2021 Conference* (S. 95-116). Cham: Springer. [https://doi.org/10.1007/978-3-031-32225-9\\_7](https://doi.org/10.1007/978-3-031-32225-9_7)

Reproduced with permission of Springer.

# Chapter 7

## Identifying Performance Levels of Enacted Pedagogical Content Knowledge in Trainee Biology Teachers' Lesson Plans



Leroy Großmann and Dirk Krüger

### 7.1 Introduction

Lesson planning is a core part of professional teachers' expertise (Carlson & Daehler, 2019). Therefore, lesson planning competencies are intended to be fostered in many national teacher training programs in Europe (e.g., Estonia: Kutsekoda, 2013; France: MEN, 2013; Germany: KMK, 2019; Ireland: Céim, 2020; Norway: NRLU, 2016) and beyond (e.g., Australia: AITSL, 2018). Yet, in many of these educational policy documents, it is barely operationalised what good lesson planning actually is. Lesson planning competence is an integral part of teachers' professional skills. In order to learn to plan and to reflect on the quality of planning, trainee biology teachers (TBTs) are visited repeatedly by their teacher educators in the course of their teacher training to get feedback and reflect on their teaching practice. For that purpose, TBTs are intended to write lesson plans in which they describe and justify their planned instructional practice. For instance, written lesson plans play a major role in lessons that are demonstrated throughout teacher training and particularly in the final examination of German science teacher education (König et al., 2020; Neumann et al., 2017). Yet, despite this relevance in teacher education, there are no objective criteria to assess lesson plan quality (Schröder et al., 2020). Consequently, TBTs and teacher educators have claimed that quality criteria in teacher training are mostly inconsistent (Döbrich & Abs, 2008; Kärner et al., 2019; Strietholt & Terhart, 2009). Although such criteria are still missing, research has shown a variety of aspects that contribute to TBTs' lesson planning skills and that should be learnt during professional development programs. When planning a lesson, teachers are confronted with

---

L. Großmann (✉) · D. Krüger  
Freie Universität Berlin (Biology Education), Berlin, Germany  
e-mail: [leroy.grossmann@fu-berlin.de](mailto:leroy.grossmann@fu-berlin.de)

© The Author(s), under exclusive license to Springer Nature  
Switzerland AG 2023

G. S. Carvalho et al. (eds.), *Fostering Scientific Citizenship in an Uncertain World*, Contributions from Science Education Research 13,  
[https://doi.org/10.1007/978-3-031-32225-9\\_7](https://doi.org/10.1007/978-3-031-32225-9_7)

95

various cognitive demands. They need to take many aspects into account that have an impact on students' learning processes, for example the analysis of students' needs, the creation of tasks, or the development of assessment strategies to diagnose whether students' learning was successful (Karlström & Hamza, 2021; König et al., 2021). Hence, TBTs need to consider those different aspects and interconnect them in the process of lesson planning. For instance, TBTs need to make sure that the intended learning outcome matches students' current needs or that the level of difficulty of the tasks is aligned with the students' level of competence (Zaragoza et al., 2021).

The knowledge and skill that science teachers basically need to do so is *pedagogical content knowledge* (PCK: Kunter et al., 2013; Park & Oliver, 2008; Shulman, 1986). PCK is regarded as crucial in the processes of planning, teaching, and reflecting (Alonzo et al., 2019; Carlson & Daehler, 2019) since it helps to address and interconnect those different PCK components in a sophisticated way to design highly effective learning processes. The integration of these components requires teachers to continuously readjust their planning thoughts, finally resulting in a more coherent interconnection between these components (Park & Oliver, 2008). In general, it was shown that a high level of PCK interconnectedness is positively correlated to teacher performance assessment scores (Reynolds & Park, 2021). However, a recent qualitative analysis of PCK reflected in written lesson plans found no significant positive correlation between the overall lesson plan quality on the one hand and the amount of PCK connections identified by applying the PCK map approach (Park & Suh, 2019) on the other hand (Großmann & Krüger, 2022b).

Consequently, good lesson plans must be characterised by more than the mere quantity of co-occurring PCK components. Although various aspects of lesson planning are challenging for pre-service teachers (Jones et al., 2011; Rusznyak & Walton, 2011), there is no research on lesson planning as a situation-specific skill, yet (König et al., 2021). In order to address TBTs' demand for objective criteria for good lesson plans, we further investigated the PCK episodes that were identified in our previous study and customised Chan et al.'s (2019) generic PCK template for use in the context of lesson planning. By inductively identifying differences in the level of sophistication in those PCK episodes, a scoring rubric with transparent criteria (Brookhart, 2018) was developed (Großmann & Krüger, 2022a). In so doing, this research can contribute to helping TBTs plan biology lessons in an efficient manner.

In this paper, we will give insights into how the criteria of the scoring rubric have been developed by inductive identification of performance level descriptions.

## 7.2 Theoretical Background

### 7.2.1 *Enacted Pedagogical Content Knowledge During Lesson Planning*

When TBTs plan, teach and reflect on their lessons, they need conceptual and procedural knowledge to make informed decisions about the biological content, the intended learning outcome, activities etc. (John, 2006). Besides content knowledge (CK) and pedagogical knowledge (PK), teachers need first and foremost pedagogical content knowledge (PCK; Shulman, 1986), i.e., “the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes” (Gess-Newsome, 2015, p. 36). In the sense of the Refined Consensus Model (Carlson & Daehler, 2019), an individual teacher’s general knowledge can be called “personal PCK” (pPCK). The part of pPCK that is made use of in a concrete situation (e.g., the planning of a particular lesson) is called “enacted PCK” (ePCK). More precisely, the situation-specific knowledge used to plan a particular lesson can be called “ePCK<sub>p</sub>” (planning; Alonzo et al., 2019). In the process of lesson planning, TBTs transform their professional knowledge as well as their motivation and beliefs into a (mental) lesson plan; hence they use pPCK (Stender et al., 2017). Afterwards, TBTs modify the initial lesson plan by taking the intended learning outcome, students’ needs etc. into account; hence they use ePCK<sub>p</sub>.

The aspects TBTs need to consider in this process are shown in the pentagon model of PCK (Park & Oliver, 2008), which serves as the conceptual framework for this study (Fig. 7.1). According to that, PCK consists of five distinct components that influence each other. Particularly the connection between the *Knowledge of Students’ Understanding of Science* (KSU) and the *Knowledge of Instructional Strategies* (KISR) has proven crucial for planning and teaching (Förtsch et al., 2016; Reynolds & Park, 2021). However, the other components also play an important role in the context of lesson planning because teachers need knowledge about the competencies they need to foster (OTS), the biological topics that need to be taught (KSC), and strategies to assess whether learning processes were successful (KAs). These five components need to be considered and interconnected in the planning of every lesson and therefore form TBTs’ ePCK<sub>p</sub>. However, as König et al. (2021) point out, there is no research on lesson planning as a situation-specific skill yet.

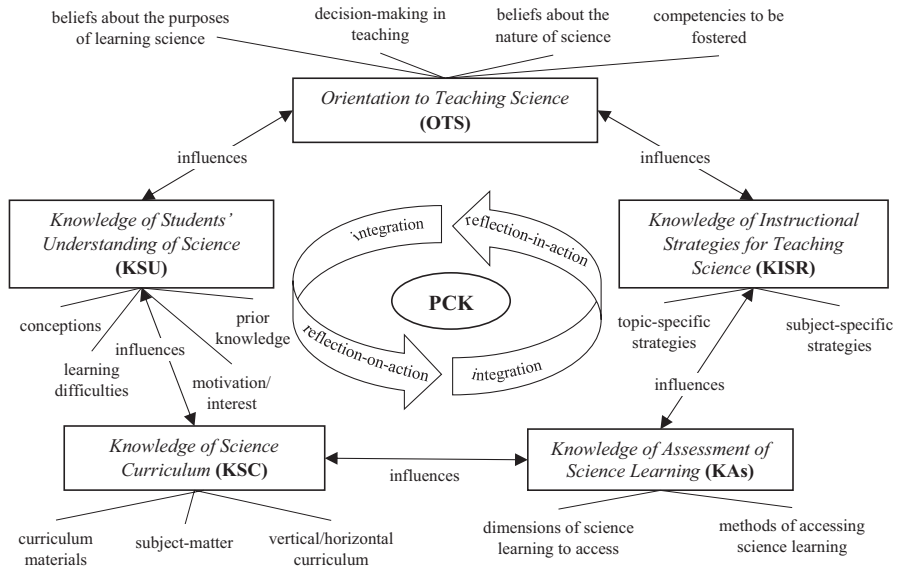


Fig. 7.1 Pentagon model of PCK. (Adapted from Park & Oliver, 2008; Park & Chen, 2012)

### 7.2.2 Lesson Planning

Empirical research on lesson planning has shown that unlike many theoretical models of lesson planning that suggest a linearity, teachers do not follow normative guidelines but rather undergo a repetitive, non-linear process (König et al., 2021). They develop ideas, choose activities or adapt tasks from curriculum materials without stating a clear intended learning outcome first. Novice teachers encounter various cognitive demands, difficulties and challenges in the process of lesson planning (Jones et al., 2011; König et al., 2021; Rusznyak & Walton, 2011; Westerman, 1991), such as formulating clear learning objectives or considering students' needs. In particular, it is challenging to rather make subject-specific planning decisions (i.e., PCK-based arguments) than just generic planning decisions (i.e., PK-based arguments; König et al., 2020). Specifically, unlike experts, novices rather argue on the level of surface structures (i.e., visible teaching methods such as activities, media etc.; Rusznyak & Walton, 2011; Westerman, 1991) than on the level of deep structures (i.e., invisible cognitive and learning processes of the students; Oser & Baeriswyl, 2001). As a consequence, there is no need for another prescriptive approach but for a descriptive attempt to help TBTs to meet the challenge of planning efficient lessons. As Zaragoza et al. (2021) suggest, teacher educators should help pre-service teachers connect their professional knowledge (ePCK<sub>p</sub>) to teaching practice (planning). Therefore, we use the CODE-PLAN model that describes six cognitive demands that teachers encounter in lesson planning (König et al., 2021) and align them to the ePCK<sub>p</sub> components (Table 7.1).

**Table 7.1** The six cognitive demands for lesson planning in the CODE-PLAN model (König et al., 2021) and the corresponding ePCK<sub>p</sub> components in the pentagon model of PCK (Park & Oliver, 2008)

Cognitive demand	ePCK <sub>p</sub>	Explanation
Unit contextualisation	OTS	Planning long-term learning processes, fostering competence repeatedly ( <i>f</i> )
Clarity of learning objectives	OTS	Choosing and specifying a standard from the curriculum as an intended learning outcome ( <i>i</i> )
Adaptation to student learning dispositions	KSU-OTS	Analysing the students' current skills, conceptions, and prior knowledge ( <i>i</i> )
Content transformation	KSC-KISR	Reconstructing biological content and making it accessible for students ( <i>i</i> )
Task creation	KISR-OTS	Designing challenging tasks that are aligned to the intended learning outcome ( <i>q</i> )
Phasing	KISR-OTS	Designing the phases of a lesson from start to end to structure the learning process ( <i>i</i> )

(*f*) frequency scale, (*i*) intensity scale, (*q*) quality scale (see Fig. 7.2)

The only ePCK<sub>p</sub> component that is not considered in the CODE-PLAN model is *Knowledge of Assessment of Science Learning* (KAs). This is important particularly in the U.S. for instance (Wiggins & McTighe, 2005) but is hardly found to be considered by pre-service teachers in empirical research (e.g., Chan & Yung, 2018; Großmann & Krüger, 2022b; Park & Chen, 2012; Weitzel & Blank, 2020). However, KAs might be necessary to assess whether or not students learned what was intended to be learnt (Karlström & Hamza, 2021).

As far as written lesson plans are concerned, TBTs are expected not only to consider those aspects when planning a lesson and describe them in a lesson plan. They are also expected to justify important planning decisions (Nilsson, 2009), which is often challenging for pre-service teachers (Zaragoza et al., 2021). More specifically, these five ePCK<sub>p</sub> components might encourage TBTs to refer to the level of deep structures in their planning and integrate these with levels of surface structures which is an important characteristic of lesson planning (Oser & Baeriswyl, 2001).

### 7.3 Research Questions

Since there is still a lack of objective, reliable and valid criteria to assess the quality of a lesson plan and to help TBTs plan efficient learning processes, we aim to develop such criteria in a qualitative way. In this study, we analyse TBTs' ePCK<sub>p</sub> identified in lesson plans, describe further cognitive demands that TBTs are confronted with, and inductively develop ePCK<sub>p</sub>-based performance level descriptions for the way TBTs deal with these demands. In so doing, it is possible to create a scoring rubric based on Chan et al.'s (2019) "Grand rubric for measuring science teachers' PCK" that can help TBTs, as well as teacher educators, to work with

transparent and objective criteria in the context of lesson planning. We focus on the following research questions:

RQ 1: What cognitive demands are TBTs confronted with when writing a lesson plan?

RQ 2: Which performance levels describe TBTs ePCK<sub>p</sub> when writing a lesson plan?

## 7.4 Methods

### 7.4.1 *Research Context*

In Germany, teacher education consists of two consecutive phases (Neumann et al., 2017): The first phase takes place at university and forms basically the theoretical part of teacher education. Pre-service teachers study two subjects (e.g., biology and history/politics/chemistry etc.) and acquire first a Bachelor's degree (after 3 years) and then a Master's degree (after another 2 years). The second phase is the induction phase, taking 1–2 years (depending on state regulations) and forms basically the practical part of teacher education. At this stage, pre-service teachers are called trainee teachers. They work up to 10 hours per week at a school and have to attend courses in pedagogy as well as in the two subjects they studied. Teacher educators are in-service teachers themselves and regularly visit the trainee teachers at their schools to give them feedback concerning their planning and teaching and to help them learn to reflect on their teaching practice. For these school visitations, trainee teachers are expected to write lesson plans in order to describe and justify their planning decisions. At the end of the second phase, trainee teachers have to take the Second State Examination: They have to teach two lessons (one per subject they studied) at their school. Therefore, they have to write a lesson plan, teach the lesson and reflect on it. The examination committee consists of up to five teacher educators and the school principal, who evaluate the performance in planning, teaching and reflecting. If trainee teachers pass this final examination, they receive a certificate that allows them to become full-time in-service teachers in Germany.

### 7.4.2 *Data Sources and Data Collection*

The lesson plans analysed here are derived from the Second State Examination in Germany. Since the lesson plans were written for the final exam, it is possible to investigate the current state of TBTs' lesson planning competence at the end of 6.5 years of teacher training. Since the aim is to describe performance levels, it can be assumed that we would find the highest performance levels rather in this sample than in those written by Bachelor's or Master's students.



For writing lesson plans, TBTs' followed the official guidelines (e.g., length, aspects to consider such as school context etc.) by the Ministry of Education (SenBJF, 2017). These lesson plans comprise approximately 10–15 pages, plus appendices (e.g., worksheets). There are several aspects that TBTs are expected to consider: They need to focus on a specific intended learning outcome based on the curricular standards; this learning outcome should be in line with the previous and the following lessons within the unit; they need to analyse in detail their students' current understanding in science (prior knowledge, level of competencies, students' conceptions) they need to analyse the biological subject-matter and describe a fruitful way to teach this topic. Based on that, TBTs create an instructional setting, i.e., a 45-minute lesson that aims to help students to achieve the intended learning outcome. This setting needs to be described, and relevant decisions (e.g., lesson structure, tasks) need to be justified. As an appendix, TBTs provide worksheets, PowerPoint slides and other additional resources relevant to the lesson.

The Ministry of Education provided 106 anonymised lesson plans, 46 were selected by *purposeful random sampling* (Patton, 1990) and were analysed in a previous study (Großmann & Krüger, 2022b). The authors of this paper are teacher educators associated with the first phase of teacher training, whereas the lesson plans were developed by TBTs at the end of the second phase of teacher training. Consequently, we had no influence on the data collection (i.e., the structure of the lesson plans, for example) but only received the lesson plans afterwards with no further information about the TBTs. On the one hand, this made it impossible to include further data sources (e.g., interviews); it is rather uncommon in PCK research to analyse lesson plans only (Chan & Hume, 2019). On the other hand, the purpose is to capture ePCK<sub>p</sub> as reflected in written lesson plans, whereas the construct captured in interviews would rather be pPCK (Carlson & Daehler, 2019; Alonzo et al., 2019). The analysis of authentic lesson plans that were not influenced by the research context allows an ecologically valid interpretation of the findings (Schröder et al., 2020), i.e., “research findings would generalise to settings typical of everyday life” (Wegener & Blankenship, 2007, p. 275).

### 7.4.3 Data Analysis

In order to capture TBTs' ePCK<sub>p</sub> in written lesson plans, we used the pentagon model of PCK (Park & Oliver, 2008) as the conceptual framework and applied the PCK map approach (Park & Suh, 2019). This method allows an in-depth qualitative analysis of ePCK<sub>p</sub> and has proven suitable for various qualitative studies in recent research in science education (Chan, 2022).

Two steps were carried out to analyse the data, (1) an in-depth analysis of ePCK<sub>p</sub> and (2) an enumerative analysis (Park & Suh, 2019). In the first step, a category system (Großmann & Krüger, 2020) was applied to identify individual ePCK<sub>p</sub> components in the lesson plans. In so doing, the first author and a trained student

assistant serving as second coder identified PCK episodes, i.e., sequences of varying length where at least two of the five PCK components (Park & Oliver, 2008) are combined. Each PCK episode was described in terms of which aspect of lesson planning was referred to.

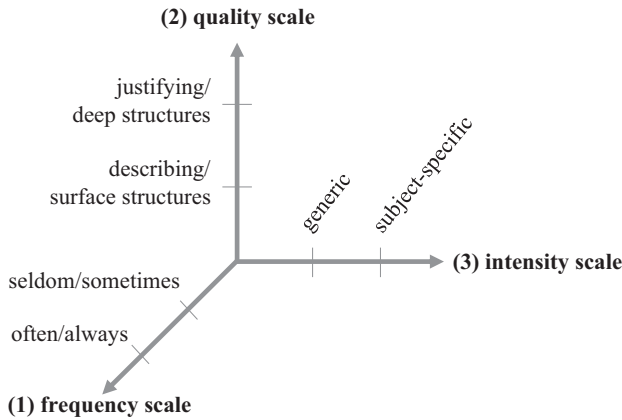
In the second step, all PCK episodes were summed up to create a PCK map that visualised the interconnectedness of PCK. This methodology follows the assumption that the more interconnections between PCK components are produced, the higher the teacher quality (Reynolds & Park, 2021). For the case of lesson planning, Großmann and Krüger (2022b) found no positive correlation between the overall lesson plan quality and the number of ePCK<sub>p</sub> connections. As a matter of fact, the pentagon model of PCK (Park & Oliver, 2008) suggests that all connections between the five PCK elements are equally important. However, there are various ways to interconnect PCK elements that can differ in their level of adequacy, preciseness, and purposefulness. Consequently, in this study, we carried out a cross-case analysis (Creswell, 2013) and compared the quality of all single ePCK<sub>p</sub> components and the quality of the PCK episodes for each cognitive demand that was found in the lesson plans. Thus, differences in level between those ePCK<sub>p</sub> components and the PCK episodes were identified.

#### 7.4.4 *Development of Scoring Rubric Criteria*

Finally, we customised Chan et al.'s (2019) "Grand rubric template for measuring science teachers' PCK" for the context of lesson planning: In addition to König et al. (2021), we identified further cognitive demands and aligned them to the ePCK<sub>p</sub> aspects (RQ1). These cognitive demands serve as criteria for the scoring rubric. Then, different performance levels for the ePCK<sub>p</sub> were described (RQ2). Hence, the rubric is analytic because each criterion is considered separately (Brookhart, 2018). To describe each criterion as distinctly and precisely as possible, verbal qualifiers had to be identified that are useful to illustrate the performance level differences (cf. Rohrmann, 2007): Rating scales can be designed by referring to, inter alia, frequency (*never – seldom – often/always*), quality (*neglected – described – justified*), and intensity (*neglected – generic/pedagogical reasons – subject-specific/educational reasons*). Depending on the cognitive demand, one of those dimensions was selected to describe the ePCK<sub>p</sub> performance levels.

### 7.5 Results

The purpose of this study was to identify further cognitive demands TBTs are confronted with when writing a lesson plan and develop criteria for a scoring rubric by describing different performance levels of TBTs' ePCK<sub>p</sub> when dealing with these demands. Results for both research questions will be reported separately.



**Fig. 7.2** A heuristic model for assessing lesson plan quality with performance level descriptions. Quality criteria can be assigned to three dimensions: (1) Frequency scale: Criteria show differences in how many times an aspect is considered or in how far it was considered seldom/sometimes or often/always; (2) Quality scale: Criteria show differences in whether planning decisions and the planned instruction are only described by taking surface structures into account or also justified with arguments on the level of deep structures; (3) Intensity scale: Criteria show differences in whether planning decisions are generic by basically using pedagogical knowledge (PK) or subject-specific by referring to educational arguments using PCK

### *RQ 1: Cognitive demands*

In addition to the six cognitive demands proposed by König et al. (2021) (Table 7.1), the cross-case analysis yielded another 18 cognitive demands (Table 7.2). Three of them describe quality aspects within one of the five ePCK<sub>p</sub> components, and the other 15 describe connections between two ePCK<sub>p</sub> components.

### *RQ 2: performance levels*

We identified three qualifier dimensions that have proven useful in establishing performance levels for the different criteria. Figure 7.2 shows the heuristic model for assessing lesson plan quality. Quality can be reflected in (1) how often an aspect is referred to (*frequency*), in (2) the manner that decisions are rather described on the level of surface structures (e.g., activities, media etc.) or justified with arguments on the level of deep structures (e.g., competence progression, students' cognitive processes) (*quality*), or (3) in the kind of arguments used to justify these decisions, i.e., generic reasons (e.g., students' participation) or subject-specific reasons (e.g., students' conceptions concerning a particular topic) (*intensity*).

Overall, two criteria address the frequency scale, two criteria address the quality scale, and the remaining 14 criteria address the intensity scale (Table 7.2). In this section, we will provide purposefully selected examples for all three dimensions from different lesson plans to illustrate the performance levels. For the frequency scale, we describe performance levels for the criterion “unit contextualisation” (Table 7.3); for the quality scale, we describe performance levels for the criterion

**Table 7.2** Extension and differentiation of the CODE-PLAN model (König et al., 2021) ordered by the five PCK components

Cognitive demand	ePCK <sub>p</sub>	Explanation
Presentation of the unit	OTS	Providing information not only about the topics to be taught in each lesson but also about the concrete intended learning outcomes ( <i>i</i> )
Evidence for students' learning	OTS-KISR	Stating clear evidence on how to assess whether students achieved the intended learning outcome ( <i>i</i> )
Analysis of topic	KSC	Analysing the topic, its inner perspectives, levels etc. as a basis for designing a way to teach it ( <i>i</i> )
Choice of topic	KSC-KSU	Explaining why the topic is relevant to the students ( <i>i</i> )
Technical terms	KSC-KISR	Choosing technical terms that are relevant for the lesson, neglecting technical terms that are confusing or irrelevant for achieving the learning objective ( <i>q</i> )
Competence progression (short-term)	KSU-OTS	Establishing a discrepancy between students' current skills and the intended learning outcome so that students learn something new ( <i>f</i> )
Students' conceptions	KSU-KSC	Referring to students' conceptions concerning the topic ( <i>i</i> )
Learning difficulties	KSU-KSC	Referring to learning difficulties (e.g., abstractness of the topic) that might arise ( <i>i</i> )
Methodical skills	KSU-KISR	Analysing students' methodical skills (e.g., group work) for methods and activities that are necessary for the lesson ( <i>i</i> )
Purpose of phases	KISR-OTS	Justifying the design of the phases by explaining why they are necessary for achieving the intended learning outcome ( <i>q</i> )
Activities	KISR-OTS	Describing and justifying why the activities are useful to achieve the intended learning outcome ( <i>i</i> )
Media	KISR-OTS	Choosing media and designing worksheets that are useful to achieve the intended learning outcome ( <i>i</i> )
Horizon of expectation	KISR-KSU	Providing a sample solution for the tasks that is appropriate for students' current knowledge and skill ( <i>i</i> )
Differentiation	KISR-KSU	Offering learning opportunities and additional tasks or further support for all weak learners ( <i>f</i> )
Anticipation of obstacles	KISR-OTS	Anticipating possible obstacles that might prevent students from achieving the intended learning outcome and describing counter-measures ( <i>i</i> )
Transparency of structure	KISR-OTS	Providing any kind overview for the students to tell them what they are going to learn in the lesson and to clarify the process ( <i>i</i> )
Transparency of criteria	KAS-OTS	Informing students about the teachers' expectations about what the students are intended to do ( <i>i</i> )
Feedback on learning	KAS	Giving students profound feedback on their performance and the progress they made in the lesson ( <i>i</i> )

Note: (*f*) frequency-, (*i*) intensity-, (*q*) quality-scale (Fig. 7.1)

**Table 7.3** Example of the criterion “competence progression” using the frequency scale

Examples	<i>Seldom</i>			<i>Often</i>		
	<i>Competence progression: planning experiments (TBT 35)</i>			<i>Competence progression: explaining effects of climate change (TBT 08)</i>		
	#	Topic	Learning objective: The students...	#	Topic	Learning objective: The students...
	1	<i>Effect of temperature on reactions</i>	X	1	<i>Carbon cycle and climate change</i>	X
	2	<i>Bergmann's rule</i>	X	2	<i>Forests and climate change</i>	<i>... explain the effects of climate change on forests.</i>
	3	<i>Bergmann's rule</i>	<i>... plan an experiment with beakers to test Bergmann's rule.</i>	3	<i>Forests</i>	X
	4	<i>Allen's rule</i>	X	4	<i>Forests as a carbon sink</i>	<i>... discuss the differences between jungles and commercial forests.</i>
	5	<i>Allen's rule</i>	<i>... plan an experiment about the ear size of foxes.</i>	5	<i>Sea versus lake</i>	[...]
	6	<i>Allen's rule</i>	X	6	<i>Coral reefs</i>	<i>... explain the origin of coral reefs.</i>
	7	<i>Adaptation to temperature</i>	X	7	<i>Coral bleaching</i>	<i>... describe the connection between temperature and CO<sub>2</sub></i>
	8	<i>Adaptation to temperature</i>	X	8	<i>Coral reefs and climate change</i>	<i>... explain the effect of climate change on coral reefs.</i>
	9	<i>Light as an abiotic factor</i>	X	9	<i>Climate change and biodiversity</i>	<i>... discuss the necessity to safe biodiversity</i>

(continued)

**Table 7.3** (continued)

	<i>Seldom</i>	<i>Often</i>
Performance description	There is a competence progression in the unit of several lessons, i.e., there is target competence addressed in increasing in complexity and abstractness of the topic and in students' autonomy in the working process. The target competence is considered in...	
	... some lessons of the unit.	...most lessons of the unit. Content-related and process-related aspects of competence are both fostered and conflated so that students have to apply their competencies in authentic problem-solving tasks.

Note. X: the learning objective is not aligned with the competence progression in the unit

**Table 7.4** Examples for the criterion “task creation” using the quality scale

	<i>Describing/ surface structures</i>	<i>Justifying/ deep structures</i>
Examples	<i>The pupils [...] are aware of the importance of CO<sub>2</sub> as a greenhouse gas. In the additional task, I want my students to use that knowledge and the information in the diagram to explain the consequences of the coral reefs dying from climate change [...]. (TBT 08)</i>	<i>In the first task, I want to activate my students' prior knowledge about muscle contraction. I want them to assign keyword-like descriptions [...] of the processes of muscle contraction to an illustration [...]. I decided not to provide full-sentence descriptions because [...] I want my students to write a forensic report in biological technical terminology independently [...]. The students will reflect on their writing process, which is the goal of the lesson. (TBT 09)</i>
Performance description	The tasks on the worksheet mostly address the target competence. The task design (e.g., level of difficulty, operators, order, linguistic complexity) is ...	
	... described.	... described and justified by considering the particular students' needs or the intended learning outcome.

“task creation” (Table 7.4); and for the intensity scale with the aspect “feedback on learning” (Table 7.5). The first two are proposed in the original CODE-PLAN model by König et al. (2021) (Table 7.1), and the third one was added inductively (Table 7.2).

As far as unit contextualisation is concerned in this sample ( $N = 46$ ), we found units that at least somehow pursued a certain goal but only in some lessons of the unit, so it does not form the main focus (Table 7.3, seldom;  $n = 16$ ). It would be more sophisticated to address the target competence repeatedly by providing different examples and connecting OTS and KSC (Table 7.3, repeatedly;  $n = 4$ ). The poor level that is not shown here would basically mean that the TBT only listed lesson contents without any target competence at all. This was found in the majority of lesson plans ( $n = 26$ ).

**Table 7.5** Example of the criterion “feedback on learning” using the intensity scale

	<i>Generic</i>	<i>Subject-specific</i>
Examples	<i>The students are intended to describe the movement of earthworms by referring to their anatomy. [...] If there is enough time left at the end of the lesson, one talented student presents his results [...]. The other students compare his solution with their own, discuss it and correct it if necessary. Together [with the teacher], they formulate an answer to the research question and copy it from the board. [...] (TBT 17)</i>	<i>The students are intended to reflect on ecological, economic and ethical issues concerning the invasion of racoons by [...] collecting arguments from the perspectives of different people, weighing the pros and cons and developing their own point of view. [...] At the end of the lesson, two students will present their points of view and discuss the persuasiveness of their arguments. All other students [...] fill out a feedback sheet for their two classmates to give feedback on their argumentation. (TBT 05)</i>
Performance description	In the consolidation stage, ... ... feedback refers to pedagogical aspects (e.g., discipline, participation). Students present their results.	... feedback refers directly to the intended learning outcome. Students present and discuss their results.

As far as tasks are concerned, we found evidence that TBTs describe what students are expected to do in certain tasks (Table 7.4, describing/ surface structures;  $n = 8$ ). It would be more sophisticated to also justify why the task is appropriate to foster students’ competencies (Table 7.4, justifying/ deep structures;  $n = 18$ ), i.e., by connecting KISR to OTS or KSU. The poor level that is not shown here would basically mean that the tasks are not described at all (except for showing them on the worksheets), which goes for nearly half the lesson plans ( $n = 20$ ).

As far as the assessment of science learning is concerned, we found lessons that focused on securing one correct solution without a fruitful discussion and without high participation of students (Table 7.5, generic;  $n = 11$ ). This generic procedure is applicable to all subjects. It would be more sophisticated to not only let students present their results but also to let them compare and discuss them, for example by having all students provide feedback on a prepared feedback sheet so that everybody is involved in the process (Table 7.5, subject-specific;  $n = 12$ ). More importantly, the feedback sheet covers items that are directly related to the intended learning outcome (i.e., in this example, the quality of arguments), hence a subject-specific teaching approach. Here, the TBT connected KAs and KISR or OTS. The poor level that is not shown here would basically mean that the TBT did not include any kind of feedback at all. This was found in half the lesson plans ( $n = 23$ ).

## 7.6 Discussion

### 7.6.1 Limitations

It is a challenging methodological endeavour to develop an objective, reliable, and valid instrument to measure a certain competence. Compared to classical knowledge or vignette tests, the development of a scoring rubric and the evaluation of the data is more time-consuming (Schröder et al., 2020). However, those tests would basically capture TBTs' pPCK, i.e., "reservoir of knowledge and skills that the teacher can draw upon during the practice of teaching" (Carlson & Daehler, 2019, p. 85). Instead, the criteria proposed in this study are rather intended to help TBTs in the process of planning a particular lesson for particular students, i.e., in the enactment of PCK (ePCK<sub>p</sub>) to overcome certain cognitive demands. Since TBTs often lack certain knowledge relevant to planning lessons or struggle to interconnect their knowledge to make sophisticated decisions (Zaragoza et al., 2021), qualitative criteria, as shown above, might help TBTs to take important cognitive demands into consideration.

Unlike traditional knowledge tests, the criteria proposed here are not solely intended to be used for grading purposes. Instead, the criteria can be used in formative assessment to support TBTs in planning by making expectations transparent (Brookhart, 2018; Panadero & Jonsson, 2013). Therefore, it is important to build the performance level descriptions based on observations made in authentic lesson plans. In doing so, the performance levels are not arbitrary constructions made by the authors (Schröder et al., 2020) but reflect the cognitive demands that TBTs usually are confronted with in lesson planning and authentic ways that TBTs use their ePCK<sub>p</sub> to overcome these demands.

Analytic rubrics used in higher education usually consist of less than ten criteria and show four or five different performance levels (Brookhart, 2018); the 24 criteria proposed here only show three. A higher degree of differentiation might confuse the target group, i.e., TBTs who are intended to use the final scoring rubric independently. Hence, we aimed to provide basically three clearly distinguishable levels. These levels are based on three dimensions that have proven suitable to describe the performances accurately. The frequency scale is rather easy to apply since it only demands counting the occurrences of certain aspects; many rubrics in higher education use this dimension (Brookhart, 2018; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). Since lesson plan quality does not basically depend on the amount of ePCK<sub>p</sub> components occurring but on their interconnectedness (Reynolds & Park, 2021), the quality and the intensity scale have been added. Thus, TBTs are instructed to consider two important aspects of lesson planning, i.e., to not only describe but also justify their decisions (quality scale; Nilsson, 2009; Zaragoza et al., 2021) and to plan instruction for a particular purpose to enhance students' learning (intensity scale; Gess-Newsome, 2015).



### 7.6.2 *TBTs' Cognitive Demands in Written Lesson Plans*

The cross-case analysis yielded 18 cognitive demands (Table 7.2) in addition to the six proposed by König et al. (2021). When transforming these cognitive demands into criteria for the scoring rubric, inductive analysis has suggested focusing on three basic dimensions of verbal qualifiers (frequency, quality, intensity; Rohrmann, 2007). Since TBTs often struggle to interconnect their ePCK<sub>p</sub> (Großmann & Krüger, 2022b) and since the lesson plan quality does not basically depend on the frequency of certain aspects, most criteria are aligned to the intensity scale. Hence, TBTs are expected to reflect whether the decisions they make in the complex process of lesson planning (John, 2006) are actually aligned with the intended learning outcome. For instance, two cognitive demands related to *Knowledge of Assessment Strategies* (Park & Oliver, 2008) have been added (Table 7.2), both triggering TBTs to think about the transparency of the learning process and formative feedback for the students. These aspects are often not considered by pre-service teachers (Chan & Yung, 2018; Großmann & Krüger, 2022b; Park & Chen, 2012; Weitzel & Blank, 2020).

### 7.6.3 *Frequency Scale: Competence Progression*

The clarity of learning objectives and unit contextualisation are important aspects of lesson planning (König et al., 2021). Any single lesson is expected to match an overarching learning objective for the unit so that students acquire new knowledge and develop competencies over time. Therefore, each lesson should build upon the previous one and connect to the following (Seidel et al., 2005). However, in this sample, we found that most units only listed lesson contents without any target competence as well as units that at least somehow pursued a certain goal but only in a few lessons of the unit. These findings support Weitzel and Blank's (2020) findings, which showed that even after an intervention, TBTs barely connect their instruction to the previous lessons. Therefore, we applied the frequency scale here to point out that within a unit, most lessons (i.e., more than half of them) should explicitly refer to a certain target competence to foster in the course of the unit (Table 7.3). For the highest level, we specified that content-related (e.g., ecogeographical rules) and process-related aspects (e.g., planning experiments) should be considered together in order to avoid any misunderstanding: For instance, it is not expected that in an ecology unit, TBTs pursue the same objective (e.g., planning an experiment) over and over again exclusively. Rather, a certain repetition in increasing complexity and difficulty is useful to foster students' competencies. This aspect can be fostered repeatedly in different experiments so that students both acquire content knowledge on ecogeographical concepts and learn to plan experiments by generating hypotheses as an aspect of scientific inquiry (Rönnebeck et al., 2016).

Hence, the two ePCK<sub>p</sub> components *Orientation to Teaching Science* (OTS) and *Knowledge of the Science Curriculum* (KSC) (Park & Oliver, 2008) have to be interconnected here to plan a unit properly.

#### 7.6.4 *Quality Scale: Instructional Tasks*

As König et al. (2021) point out, the creation of tasks is a crucial component of lesson planning. The intended learning outcome is more or less abstract, whereas the tasks are very concrete and need to perfectly match the intended learning outcome (Kang, 2017) as well as students' current needs and skills. Tasks should not exceed the students' skills, and they should not demand too little from them, either (Kang et al., 2016). TBTs are expected not only to provide any tasks in a lesson but to make an informed decision when choosing tasks that are cognitively demanding and help students develop profound science ideas. However, in this sample, we found that nearly half the TBTs did not refer to the tasks at all (except for showing them on the worksheets). Therefore, we applied the quality scale here to point out that important planning decisions should not only be described but also and foremost, be justified with valid reasons (Table 7.4). For the two highest levels, we specified that aspects like the level of difficulty of the tasks could be taken into consideration when describing the tasks (Maier et al., 2014). For the highest level, we specified that students' needs and prior knowledge, as well as the intended learning outcome, can be arguments for justification. A good task has to match both, and TBTs should be able to justify their choice of tasks by considering that. Hence, the three ePCK<sub>p</sub> components, *Knowledge of Instructional Strategies* (KISR), *Knowledge of Students' Understanding of Science* (KSU), and *Orientations to Teaching Science* (OTS) (Park & Oliver, 2008) have to be interconnected here to design reasonable tasks.

#### 7.6.5 *Intensity Scale: Assessment of Learning*

TBTs should not only plan a lesson, including activities, media etc. but also determine acceptable evidence to assess whether their students' learning was successful (Karlström & Hamza, 2021; Wiggins & McTighe, 2005). Particularly at the end of a lesson, TBTs tend to focus on securing lesson results or key terms, but they barely consider a reflection of the results or the learning process (Weitzel & Blank, 2020). This might be useful not only to diagnose the progress toward the goal the students already made (*feedback*) but also to identify the next steps the students need to undertake to make further progress (*feed forward*; Hattie & Timperley, 2007). However, in this sample, half of the TBTs did not plan to give any feedback related to the intended learning outcome (Jones et al., 2011) but either planned feedback concerning pedagogical aspects or no feedback at all. Therefore,

we applied the intensity scale here to point out that some kind of feedback should be planned, ideally related to the intended learning outcome (Table 7.5). Then, TBTs' students would get information on what they already do well and which skills they still need to improve. Hence, the three ePCK<sub>p</sub> components *Knowledge Assessment of Science Learning* (KAs), *Orientations to Teaching Science* (OTS), and *Knowledge of Instructional Strategies* (KISR) (Park & Oliver, 2008) have to be interconnected here to determine evidence for students' learning. Despite the relevance of assessment strategies, PCK research has repeatedly shown that particularly KAs is barely considered by pre-service teachers (e.g., Aydin & Boz, 2013; Park & Chen, 2012; Weitzel & Blank, 2020), which is in line with our findings.

## 7.7 Conclusion

Although lesson planning is a core part of teachers' expertise (Carlson & Daehler, 2019) and plays an important role in many teacher training programs, there are no objective, reliable and valid criteria to assess TBTs' lesson plan quality yet (Schröder et al., 2020). Since lesson plans are part of the final examination of teacher training in Germany, there is a need for such criteria to evaluate lesson plans and particularly to help TBTs to plan lessons in a sophisticated way (Döbrich & Abs, 2008; Kärner et al., 2019; Strietholt & Terhart, 2009). Therefore, an analytic scoring rubric has been developed that illustrates 24 criteria with two desirable performance levels each (Großmann & Krüger, 2022a). Particularly the highest level is important because it offers a transparent insight into what teacher educators expect an excellent lesson plan to contain. Thus, TBTs might be encouraged to consider aspects they would miss otherwise, such as students' conceptions (Großmann & Krüger, 2022c). In addition, as Chan et al. (2019) pointed out from a PCK perspective, it would be desirable if researchers would rather not conceptualise constructs like PCK individually, as this makes it difficult to compare findings from different researchers. Hence, our customisation of Chan et al.'s (2019) template might serve other researchers in the field of PCK as an example of how to develop a scoring rubric inductively on the one hand and might encourage researchers and teacher educators in the field of lesson planning to use or modify the criteria for use in research and teacher education.

## 7.8 Outlook

The next steps in the development of the scoring rubric include several procedures to test for reliability and validity (AERA, APA, & NCME, 2014; Brookhart, 2018; Jonsson & Svingby, 2007): Further lesson plans are being analysed, and inter- as well as interrater-reliability will be calculated to check whether the measurement is objective. Moreover, as sources of evidence for validity, we are comparing the

grades from the final exam (König et al., 2020; Neumann et al., 2017) to the scores achieved by application of the rubric (*validity evidence based on relations to other variables* and *validity evidence based on consequences of testing*). If the top graded TBTs also receive higher scores than the TBTs with poor grades, and if these TBTs score higher than Bachelor's and Master's students, this would support the assumption that the rubric is suitable to distinguish quality differences in TBTs' ePCK<sub>p</sub> in written lesson plans.

*Please contact the first author to receive the complete scoring rubric.*

**Acknowledgments** The project K2Teach (*Know how to teach*) is part of the “Qualitätsinitiative Lehrerbildung”, a joint initiative of the German Federal Government and the Länder, which aims to improve the quality of teacher training. The programme is funded by the Federal Ministry of Education and Research (grant number 01JA1802). The authors are responsible for the content of this publication.

## References

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Psychological Association/AERA.
- AITSL. (2018). Australian professional standards for teachers. Australian Institute for Teaching and School Leadership. Retrieved from <https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standards-for-teachers.pdf>.
- Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 271–286). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_12](https://doi.org/10.1007/978-981-13-5898-2_12)
- Aydin, S., & Boz, Y. (2013). The nature of integration among PCK components: A case study of two experienced chemistry teachers. *Chemistry Education Research and Practice*, 14(4), 615–624. <https://doi.org/10.1039/C3RP00095H>
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3(22), 1–12. <https://doi.org/10.3389/educ.2018.00022>
- Carlson, J., & Daehler, K. R. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–92). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_2](https://doi.org/10.1007/978-981-13-5898-2_2)
- Céim. (2020). *Standards for initial teacher education*. Retrieved from <https://www.teachingcouncil.ie/en/news-events/latest-news/ceim-standards-for-initial-teacher-education.pdf>
- Chan, K. K. H. (2022). A critical review of studies using the pedagogical content knowledge map approach. *International Journal of Science Education*, 1–27. <https://doi.org/10.1080/09500693.2022.2035011>
- Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_1](https://doi.org/10.1007/978-981-13-5898-2_1)
- Chan, K. K. H., & Yung, B. H. W. (2018). Developing pedagogical content knowledge for teaching a new topic: More than teaching experience and subject matter knowledge. *Research in Science Education*, 48(2), 233–265. <https://doi.org/10.1007/s11165-016-9567-1>

- Chan, K. K. H., Rollnick, M., & Gess-Newsome, J. (2019). A grand rubric for measuring science teachers' pedagogical content knowledge. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 253–271). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_11](https://doi.org/10.1007/978-981-13-5898-2_11)
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Sage.
- Döbrich, P., & Abs, H. J. (2008). Evaluation der zweiten Phase der Lehrerbildung. Pädagogische Entwicklungsbilanzen mit Studienseminaren in Hessen [Evaluation of the second phase of teacher training. Educational development in seminars in the federal state of Hesse]. *Schulverwaltung, Hessen, Rheinland-Pfalz*, 13(3), 70–73.
- Förtsch, C., Werner, S., von Kotzebue, L., & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. <https://doi.org/10.1080/09500693.2016.1257170>
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In A. Berry, P. J. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Großmann, L., & Krüger, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen [Development and application of a category system to analyze pre-service biology teachers' pedagogical content knowledge in lesson plans]. *Erkenntnisweg Biologiedidaktik*, 19, 21–39.
- Großmann, L., & Krüger, D. (2022a). Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen [Planning biology lessons successfully – scoring rubric for writing lesson plans]. *SEMINAR – Lehrerbildung Und Schule*, (1), 91–110.
- Großmann, L., & Krüger, D. (2022b). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? [What's the role of trainee biology teachers' pedagogical content knowledge for the quality of their written lesson plans?]. *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1). <https://doi.org/10.1007/s40573-022-00141-w>
- Großmann, L., & Krüger, D. (2022c). Students' conceptions as a neglected perspective in trainee teachers' biology lesson plans. In K. Korfiatis & M. Grace (Hrsg.), *Current research in biology education. Contributions from biology education research* (S. 181–193). ERIDOB, Springer. [https://doi.org/10.1007/978-3-030-89480-1\\_14](https://doi.org/10.1007/978-3-030-89480-1_14)
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- John, P. D. (2006). Lesson planning and the student teacher: Re-thinking the dominant model. *Journal of Curriculum Studies*, 38(4), 483–498. <https://doi.org/10.1080/00220270500363620>
- Jones, K. E., Jones, J. L., & Vermette, P. J. (2011). Six common lesson planning pitfalls: Recommendations for novice educators. *Education*, 131(4), 845–864.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kang, H. (2017). Preservice teachers' learning to plan intellectually challenging tasks. *Journal of Teacher Education*, 68(1), 55–68. <https://doi.org/10.1177/0022487116676313>
- Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching*, 53(9), 1316–1340. <https://doi.org/10.1002/tea.21329>
- Karlström, M., & Hamza, K. (2021). How do we teach planning to pre-service teachers – A tentative model. *Journal of Science Teacher Education*, 32(6), 1–22. <https://doi.org/10.1080/1046560X.2021.1875163>

- Kärner, T., Bonnes, C., & Schölzel, C. (2019). Bewertungstransparenz im Referendariat [transparency of evaluation criteria in the induction phase of teacher training]. *Zeitschrift für Pädagogik*, 65(3), 378–400.
- KMK. (2019). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019* [Standards for teacher education for all subjects. Resolution of the German ministers for Education from 16th December 2004 in its present form from 16th May 2019]. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf)
- König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I., & Glutsch, N. (2020). Pre-service teachers' generic and subject-specific lesson-planning skills: On learning adaptive teaching during initial teacher education. *European Journal of Teacher Education*, 43(2), 131–150. <https://doi.org/10.1080/02619768.2019.1679115>
- König, J., Krepf, M., Bremerich-Vos, A., & Buchholtz, C. (2021). Meeting cognitive demands of lesson planning: Introducing the CODE-PLAN model to describe and analyze teachers' planning competence. *The Teacher Educator*, 56(4), 466–487. <https://doi.org/10.1080/08878730.2021.1938324>
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *The Journal of Educational Psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Kutsekoda. (2013). *Professional Standard - Senior Teacher Level 7*. Retrieved from [https://www.kutsekoda.ee/wp-content/uploads/2019/ENG\\_veeb/Teacher-level-7\\_eng-13.pdf](https://www.kutsekoda.ee/wp-content/uploads/2019/ENG_veeb/Teacher-level-7_eng-13.pdf)
- Maier, U., Bohl, T., Drüke-Noe, C., Hoppe, H., Kleinknecht, M., & Metz, K. (2014). Das kognitive Anforderungsniveau von Aufgaben analysieren und modifizieren können: Eine wichtige Fähigkeit von Lehrkräften bei der Planung eines kompetenzorientierten Unterrichts [Analyzing and modifying the cognitive activation potential of tasks: An important skill of teachers when planning competence-oriented lessons]. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32(2), 340–358. Retrieved from [http://fox.leuphana.de/portal/de/publications/das-kognitive-anforderungsniveau-von-aufgaben-analysieren-und-modifizieren-koennen\(4c6b035f-3b67-4fe8-bb94-28692e82e98a\).html](http://fox.leuphana.de/portal/de/publications/das-kognitive-anforderungsniveau-von-aufgaben-analysieren-und-modifizieren-koennen(4c6b035f-3b67-4fe8-bb94-28692e82e98a).html)
- MEN. (2013). *Arrêté du 1er juillet 2013 relatif au référentiel des compétences professionnelles des métiers du professorat et de l'éducation* [Resolution from the 1 July 2013 concerning the professional standards of teachers]. Retrieved from <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027721614/>
- Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki, & T. Hirano (Eds.), *Model science teacher preparation programs: An international comparison of what works* (pp. 29–52). Information Age Publishing, Inc.
- Nilsson, P. (2009). From lesson plan to new comprehension: Exploring student teachers' pedagogical reasoning in learning about teaching. *European Journal of Teacher Education*, 32(3), 239–258. <https://doi.org/10.1080/02619760802553048>
- NRLU. (2016). *National guidelines for the primary and lower secondary teacher education programme for years 1–7*. Retrieved from [https://www.uhr.no/\\_f/p-1/i9667e583-aa3b-4f25-a8fe-64af8b199072/national\\_guidelines\\_for\\_the\\_primary\\_and\\_lower\\_secondary\\_teacher\\_education\\_programme\\_for\\_years\\_1\\_7.pdf](https://www.uhr.no/_f/p-1/i9667e583-aa3b-4f25-a8fe-64af8b199072/national_guidelines_for_the_primary_and_lower_secondary_teacher_education_programme_for_years_1_7.pdf)
- Oser, F. K., & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 1031–1065). American Educational Research Association.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>

- Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of Pedagogical Content Knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941. <https://doi.org/10.1002/tea.21022>
- Park, S., & Oliver, J. S. (2008). National Board Certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45(7), 812–834. <https://doi.org/10.1002/tea.20234>
- Park, S., & Suh, J. K. (2019). The PCK map approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 185–197). Springer.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed. [Nachdr.]). Sage.
- Reynolds, W. M., & Park, S. (2021). Examining the relationship between the educative teacher performance assessment and preservice teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, 58, 721–748. <https://doi.org/10.1002/tea.21676>
- Rohrmann, B. (2007). Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data. Project Report University of Melbourne/Australia. Retrieved from <http://www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf>.
- Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Rusznayk, L., & Walton, E. (2011). Lesson planning guidelines for student teachers: A scaffold for the development of pedagogical content knowledge. *Education as Change*, 15(2), 271–285. <https://doi.org/10.1080/16823206.2011.619141>
- Schröder, J., Riese, J., Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Kulgemeyer, C., Reinhold, P., & Schecker, H. (2020). Die Messung der Fähigkeit zur Unterrichtsplanung im Fach Physik mit Hilfe eines standardisierten Performanztests [Measuring the skills of lesson planning in the subject of physics using a standardized performance assessment]. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 103–122. <https://doi.org/10.1007/s40573-020-00115-w>
- Seidel, T., Rimmele, R., & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction*, 15(6), 539–556. <https://doi.org/10.1016/j.learninstruc.2005.08.004>
- SenBJF. (2017). *Handbuch Vorbereitungsdiens: Materialien für den reformierten Berliner Vorbereitungsdiens* [Manual for the induction phase of teacher training: Information and materials for the reformed induction phase in Berlin] (6th edition). Retrieved from [https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch\\_vorbereitungsdienst.pdf](https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch_vorbereitungsdienst.pdf)
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Stender, A., Brückmann, M., & Neumann, K. (2017). Transformation of topic-specific professional knowledge into personal pedagogical content knowledge through lesson planning. *International Journal of Science Education*, 39(12), 1690–1714. <https://doi.org/10.1080/009500693.2017.1351645>
- Strietholt, R., & Terhart, E. (2009). Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung [Evaluating trainee teachers. An explorative analysis of evaluation instruments in the second phase of teacher training]. *Zeitschrift Für Pädagogik*, 55, 622–645. Retrieved from <http://www.content-select.com/10.3262/ZP0904622>
- Wegener, D. T., & Blankenship, K. L. (2007). Evological validity. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 275–277). Sage Publications.
- Weitzel, H., & Blank, R. (2020). Pedagogical Content Knowledge in peer dialogues between pre-service biology teachers in the planning of science lessons. Results of an intervention

- study. *Journal of Science Teacher Education*, 31(1), 75–93. <https://doi.org/10.1080/1046560X.2019.1664874>
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292–305.
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Association for Supervision and Curriculum Development. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=133964>
- Zaragoza, A., Seidel, T., & Hiebert, J. (2021). Exploring preservice teachers' abilities to connect professional knowledge with lesson planning and observation. *European Journal of Teacher Education*, 1–20. <https://doi.org/10.1080/02619768.2021.1996558>



## Beitrag 7

**Großmann, L.** & Krüger, D. (2023). Assessing the Quality of Science Teachers' Lesson Plans: Evaluation and Application of a Novel Instrument. *Science Education*, 1-37. <https://doi.org/10.1002/sce.21832>

Das Werk ist unter CC-BY-4.0-Lizenz erschienen (<https://creativecommons.org/licenses/by/4.0/deed.de>). Der hier veröffentlichte Beitrag ist identisch mit der unter der o.g. DOI erreichbaren und von Wiley publizierten Fassung.

# Assessing the quality of science teachers' lesson plans: Evaluation and application of a novel instrument

Leroy Großmann  | Dirk Krüger 

Biology Education, Freie Universität Berlin, Berlin, Germany

## Correspondence

Leroy Großmann, Biology Education, Freie Universität Berlin, Schwendenerstraße 1, Berlin 14195, Germany  
Email: [leroy.grossmann@fu-berlin.de](mailto:leroy.grossmann@fu-berlin.de)

## Abstract

Lesson planning is a core part of teachers' professional competence. Written lesson plans play a significant role in science teacher education as a preparation for demonstration lessons during the final teacher certification exam. However, the few existing scoring rubrics on lesson plans are not particularly theoretically sound and are barely tested for the validity of score interpretations. In response to the demand for transparent and applicable criteria, we developed the *rubric to assess science lesson plans* (RALP) to assess science teachers' lesson plan quality. We employed a mixed-methods approach: First, we present multiple sources of validity evidence (based on *test content*, *internal structure*, *relations to other variables*, and *consequences of testing*) as mainly quantitative indicators for the quality of the RALP. Based upon that, we applied the RALP to lesson plans written by preservice and trainee science teachers ( $N = 100$ ) and provided a qualitative analysis of six cases to illustrate common patterns in these lesson plans. Results indicate that teacher educators consider the RALP criteria ( $N = 24$ ) relevant and objectively applicable. Correlation analyses of the scores and two teacher educators' holistic quality assessment of all lesson plans provide convincing evidence that the RALP can discriminate lesson plan quality

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Science Education* published by Wiley Periodicals LLC.

levels. Moreover, comparisons between preservice science teachers and trainee science teachers reveal that trainee teachers score significantly higher than preservice teachers, indicating that the RALP is sensitive to differences in teaching and planning experience. The application and in-depth analysis of three criteria of the RALP illustrate these differences in levels of planning quality. We discuss possible applications of the RALP in science teacher education and research in science teaching.

#### KEYWORDS

assessment, lesson plan, science teacher education, scoring rubric, validity

## 1 | INTRODUCTION

Lesson planning is a core part of teachers' professional competence (Carlson et al., 2019; Mutton et al., 2011; Zaragoza et al., 2021). Even though empirical evidence is still scarce, it is plausible to assume that high-quality planning correlates with high-quality classroom teaching. Lesson planning was shown to predict teaching effectiveness in terms of students' achievement gains (Darling-Hammond et al., 2013) and in terms of students' ratings of instructional quality (König et al., 2021). Consequently, lesson planning forms a constitutive part of teacher education, both in national professional standards (e.g., Australian Institute for Teaching and School Leadership, 2018; France: Ministère de l'Éducation Nationale, 2013; Germany: Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2019) and in official teacher accreditation requirements such as the American "educative Teacher Performance Assessment" (Sato, 2014). Thus, written lesson plans play a significant role in teacher training (König et al., 2020). In a lesson plan, teachers describe and justify a planned instructional setting. Prospective teachers write such lesson plans to get feedback for their planning, teaching, and reflection skills from their teacher educators who visit them at school and observe a planned lesson. Consequently, lesson planning, in general, and writing proper lesson plans, in particular, are key competencies prospective teachers are intended to acquire during teacher education (Morine-Dersheimer, 2011).

However, empirical research on lesson planning as a situation-specific skill of (prospective) science teachers is still scarce (Campbell et al., 2022; König et al., 2021). Krepf and König (2022, p. 14) argue that "there are hardly any scientifically proven criteria for the analysis and evaluation of written lesson plans, therefore [we] consider it an important task of empirical teacher education research to present suitable procedures for the assessment of written lesson plans." Specifically, Kang (2017) points out that "research on teachers' planning was often conducted in a controlled setting to uncover cognitive decisions made by teachers. Researchers tend to rely on either teacher self-reported data or data generated from think-aloud methods [...]. Little empirical evidence focuses on teachers' effective planning in the natural teaching environments" (Kang, 2017, p. 56). Scoring rubrics can be an evidentially helpful and effective instrument for assessing the quality of written plans. They are well-established in many research areas, such as education (Jonsson & Svingby, 2007) and higher education (Brookhart, 2018). Regarding research on lesson planning, there is a need for an instrument that was developed in an ecologically valid manner and enables researchers in science education to assess the quality of prospective science teachers' lesson plans. To close this gap, the *rubric to assess science lesson plans* (RALP) was developed within the framework of pedagogical content knowledge (PCK; Carlson et al., 2019; Park & Oliver, 2008). However, the RALP would only add a fruitful

methodology to teacher education and research in science education if it was carefully evaluated to ensure that the inferences made from the scoring results are valid. In recent years, there has been an increasing demand for empirical research that devotes more attention to the consideration of validity issues generally in research (AERA, APA, & NCME, 2014; Kane, 2013; Pellegrino et al., 2016) and particularly regarding scoring rubrics (Brookhart, 2018; Moskal & Leydens, 2000). However, this is not the case in many studies that present scoring rubrics in higher education (Reddy & Andrade, 2010), PCK research (Chan et al., 2019), or research on lesson planning. Hence, in this study, we aim to employ a mixed-methods approach (Creswell & Plano Clark, 2018) to evaluate the quality of the RALP concerning multiple sources of validity evidence for the interpretations of RALP scores. Specifically, we build a validity argument (Kane, 2013) referring to validity evidence based on *test content*, *internal structure*, *relations to other variables*, and *consequences of testing* (AERA, APA, & NCME, 2014). Based upon that, we will apply the RALP to original lesson plans and provide qualitative insights into prospective biology teachers' lesson planning competence.

## 2 | LESSON PLANNING AT THE HEART OF TEACHERS' PROFESSIONAL COMPETENCE

Nearly four decades have passed since Shulman (1986) introduced the differentiation of content knowledge (CK), pedagogical knowledge (PK), and, first and foremost, PCK as the unique component of teachers' professional knowledge. As far as lesson planning is concerned, only a few empirical studies explicitly locate themselves within Shulman's (1986) topology of teachers' professional knowledge or frameworks building upon that, for example, the *refined consensus model* (RCM) of PCK (Carlson et al., 2019). PCK is defined as the "knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes" (Gess-Newsome, 2015, p. 36). Carlson et al. (2019) recognize that PK, CK, knowledge of students, knowledge of curriculum, and knowledge of assessment are foundational to PCK. Without these knowledge bases, a teacher's PCK is limited. The RCM was a significant step forward in PCK research because it combines the static knowledge bases (i.e., knowing *that*) and the dynamic process of enacting the knowledge in the teaching cycle of planning, teaching, and reflecting (i.e., knowing *how*). In the current paper, we connect to this conceptualization of PCK and explicitly focus on it when enacted in the lesson planning process (i.e., enacted PCK during planning, ePCK<sub>p</sub>; Alonzo et al., 2019). Enacted PCK is part of an individual's personal PCK (pPCK). More specifically, we adapt the pentagon model of PCK (Park & Oliver, 2008) that describes the integration of some of the knowledge bases named in the outermost circle of the RCM. It consists of *orientations to teaching science* (OTS), *knowledge of the science curriculum* (KSC), *knowledge of students' understanding in science* (KSU), *knowledge of instructional strategies* (KISR), and *knowledge of assessment strategies* (KAs). Park and Oliver (2008) emphasize the importance of enacting PCK (i.e., planning, reflecting, and particularly teaching) and integrating the five components:

This integration is accomplished through the complementary and ongoing readjustment by both reflection-in-action and reflection-on-action, resulting in strengthened coherence among the components [...]. This model suggests that the development of one component within PCK will, in turn, influence the development of others, and ultimately enhance this holistic PCK. Because PCK, which comprises effective teaching, requires the integration of the components in highly complex ways, lack of coherence among components can be problematic in developing PCK and increased knowledge of a single component may not be sufficient to stimulate significant change in practice. (Park & Oliver, 2008, p. 814)

The interplay of reflection-in-action (i.e., teaching) and reflection-on-action (i.e., planning the lesson, and reflecting on it afterward) leads to an ongoing readjustment of the five PCK components. This process is



conceptualized as pedagogical reasoning (Park & Suh, 2019), which, in turn, forms the critical practice of the inner circle of the RCM (Carlson et al., 2019).

PCK is regarded to be the crucial knowledge base for teachers (Chan & Hume, 2019) due to two reasons: First, PCK is positively correlated with teaching quality in general (Park et al., 2011; Reynolds & Park, 2021), and with particular subspects such as cognitive activation (Keller et al., 2017) or student learning support (Kunter et al., 2013). Second, teachers' level of PCK seems to predict their students' achievement (e.g., Försch et al., 2016; Kunter et al., 2013; Mahler et al., 2017).<sup>1</sup>

"[P]lanning is essentially knowledge-based" (Mutton et al., 2011, p. 412). When planning a lesson, teachers transform their professional knowledge, motivation, and beliefs into a (mental) lesson plan (Stender et al., 2017). Afterward, they consider the intended learning outcome, students' needs, and other relevant aspects related to classroom teaching and thus, modify their initial lesson plan. Prospective teachers need help to use their professional knowledge to justify planning decisions (Zaragoza et al., 2021). They use PCK in appropriate breadth and depth or neglect important aspects (Koberstein-Schwarz & Meisert, 2022). König et al. (2020) showed that trainee teachers teaching the school subject German elaborate on their planning decisions in lesson plans significantly more often by use of generic (PK) instead of subject-specific (PCK) features. Vogelsang et al. (2022) found a significant positive correlation between preservice physics teachers; PCK and their lesson planning competence (measured using a standardized performance assessment instrument) before and after a 6-month practical semester. In line with that, Backfisch et al. (2020) showed that lesson plans written by trainee and in-service mathematics teachers are of higher instructional quality than those written by preservice teachers. The instructional quality of these planned lessons correlated moderately with mathematics teachers' PCK.

In response to the need for common ground in empirical research on lesson planning, König et al. (2021) developed a heuristic model of the cognitive demands of lesson planning (CODE-PLAN model). According to that, teachers are confronted with six significant demands in the lesson planning process: *content transformation*, *task creation*, *adaptation to student learning dispositions*, *clarity of learning objectives*, *unit contextualization*, and *phasing*. These demands cover many components of PCK, for instance, as described in the pentagon model (Park & Oliver, 2008) that strengthens the necessity to interconnect the different components of PCK in planning, teaching, and reflecting. Only KAs are missing in the CODE-PLAN model, which was developed by a team of German researchers, possibly because the assessment of learning is widespread in the American educational system (e.g., Wiggins & McTighe, 2005) but barely occurs in German theoretical literature on lesson planning and advice literature for teachers (Vogelsang & Riese, 2017).

Nevertheless, it is reasonable to assume that a high level of PCK facilitates meeting the six abovementioned cognitive demands (König et al., 2021) to plan effective lessons. Both research on lesson planning and PCK have shown that, unlike experts, novices struggle to interconnect their professional knowledge (Westerman, 1991) and hence have more difficulties planning effective lessons (Koberstein-Schwarz & Meisert, 2022). In addition, Weitzel and Blank (2020) found that preservice biology teachers often do not consider certain aspects of lesson plans (e.g., students' conceptions, assessment of science learning, and the structure of teacher-class dialogs). One recently suggested way to foster students' ability to apply professional knowledge in lesson planning is scaffolding, which draws on research on professional vision (Zaragoza et al., 2023). Similar to this approach, we suggest providing a scoring rubric as another possibility of helping prospective teachers to meet those cognitive demands in lesson planning and interconnect four of the five PCK components.

### 3 | ASSESSING THE QUALITY OF LESSON PLANS

#### 3.1 | The RALP

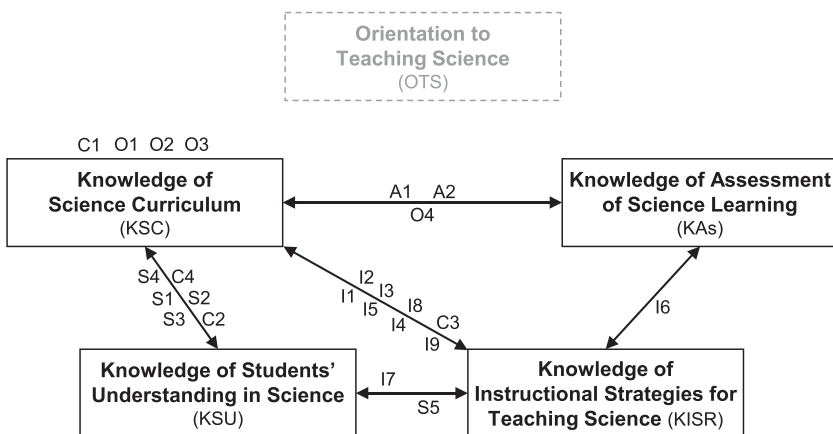
The RALP (Großmann & Krüger, 2022a, 2023) was developed based on the pentagon model of PCK (Park & Oliver, 2008) which served as a conceptual framework for our study. It illustrates the need to interconnect different

PCK components, which is necessary to overcome the challenges in the lesson planning process mentioned above. This study focuses on the four components that address knowledge (KSC, KAs, KISR, KSU). Although OTS is another crucial component of PCK as it shapes and is shaped by the remaining four components, we did not include it in the RALP development process. We aim to provide a scoring rubric that reflects theoretically and empirically based criteria to assess lesson plan quality rather than including science teachers' subjective beliefs about the purposes of learning science or the nature of science.

Since sophisticated PCK is expressed by the degree of interconnectedness of the PCK components (e.g., Chan & Hume, 2019) and since a high level of PCK correlates to teaching quality (Keller et al., 2017; Park et al., 2011; Reynolds & Park, 2021) and students' achievement (Förtsch et al., 2016; Kunter et al., 2013; Mahler et al., 2017), we developed criteria that explicitly target the interconnections between PCK components (Figure 1).

Twenty-four cognitive demands were identified based on König et al. (2021) that can be assigned to four of the five components of PCK. The 24 criteria were deduced from empirical and theoretical literature on teaching quality and from an extensive qualitative analysis of trainee science teachers' (TSTs') lesson plans (Großmann & Krüger, 2020, 2022c) using the PCK map approach (Park & Suh, 2019). Hence, the RALP reflects the realms of PCK needed in lesson planning: science teachers' pPCK and their ePCK<sub>p</sub> (Alonzo et al., 2019; Carlson et al., 2019). To give a concrete example of how the criteria were developed, criterion I4 (suitability of tasks) may provide an insight into the procedure: Creating tasks is a significant cognitive demand in lesson planning (König et al., 2021) and can be assigned to KISRs (Park & Oliver, 2008). One major challenge for prospective teachers is planning intellectually challenging tasks that directly relate to a specific intended learning outcome (Kang, 2017). Formulating intended learning outcomes to foster students' abilities is crucial. Criterion I4 explicitly builds a connection between these two aspects. Planning a lesson with tasks that do not relate to the intended learning outcome would correspond to level 0 in the RALP. Planning a task that addresses the intended learning outcome and analyzing the task in the lesson plan would correspond to level 1, and providing justifications for the design of the tasks (e.g., by relating to the students' current needs) would correspond to level 2. This procedure applies to all criteria assigned to the arrows in Figure 1.

To avoid the pitfalls of extant scoring rubrics in higher education, specific guidelines (Andrade, 2005; Brookhart, 2018; Dawson, 2017; Jonsson & Svingby, 2007), such as the use of descriptive language, have been



**FIGURE 1** Assignment of the 24 RALP criteria to the components and their interconnections in the pentagon model of PCK (Park & Oliver, 2008). The relative closeness of criteria on the interconnecting arrows indicates which of the two components the criterion focuses on most. For this study, the PCK component “Orientation to Teaching Science” was not considered and is therefore grayed out. PCK, pedagogical content knowledge; RALP, rubric to assess science lesson plans.



considered. The 24 RALP criteria were deduced from an extensive qualitative analysis of TSTs' lesson plans (Großmann & Krüger, 2022c) using the PCK map approach (Park & Suh, 2019). The comparison of lesson plans with highly and poorly interconnected PCK led to identifying three performance-level descriptions for all criteria. Connecting to König et al.'s (2021), six cognitive demands in lesson planning, we were able to add and specify another 18 (Großmann & Krüger, 2023), adding up to a total of 24 RALP criteria (Supporting Information: Appendix A). For instance, I4 refers to the tasks that students will be confronted with, which should be well-aligned with the intended learning outcome. Hence, I4 is positioned on the interconnection between KISR and OTS. Except for O1, O2, O3, C1, and A2, all criteria emphasize the need to interconnect PCK components (Figure 1). An extensive description of the development of the RALP is provided by Großmann and Krüger (2023).

### 3.2 | How the RALP can be beneficial for science teacher education

Rubrics can be used for research purposes on the one hand and for formative or summative assessment in teacher education on the other hand (Krebs et al., 2022; Panadero et al., 2018; Panadero & Jonsson, 2013). This would also account for lesson planning (Ozogul et al., 2008), which is a crucial aspect to teach in science teacher training (Drost & Levine, 2015; Karlström & Hamza, 2021). Since lesson planning is a cognitively demanding task (König et al., 2021) and rubrics have been shown to decrease the subjective cognitive load in self-assessment processes (Krebs et al., 2022), a scoring rubric on lesson planning is assumed to help prospective science teachers in better responding to the challenges of lesson planning. A distinction is commonly made between holistic and analytic scoring rubrics (Sadler, 2009; Tomas et al., 2019): Holistic rubrics consist of one single scale of combined criteria evaluated simultaneously, enabling an overall evaluation in a time-consuming way. In contrast, analytic scoring rubrics list a differentiated set of criteria and make performance expectations transparent. They are therefore regarded as “the gold standard of rubrics and are a good choice when [...] it is important to give students detailed feedback on their strengths and weaknesses” (Schreiber et al., 2012, p. 212). The RALP is an analytic scoring rubric displaying 24 criteria describing different progressing performance levels (Sadler, 2009; Tomas et al., 2019). In this way, the RALP is more helpful for teacher educators to give sophisticated and transparent feedback on written lesson plans than a holistic scoring rubric would.

An analytic scoring rubric is only appropriately used if all criteria match the intended purposes, for example, in terms of teachers' professional standards (Moskal & Leydens, 2000). However, as far as rubrics on science lesson planning are concerned, such objectives usually do not aim to cover lesson planning competence in general but planning scientific inquiry lessons in particular. To the best of our knowledge, only six studies published in peer-reviewed journals in science education over the last two decades use analytic scoring rubrics to analyze written lesson plans (Table 1).

In contrast to the RALP, Goldston et al. (2013) and von Kotzebue, (2022) aim at quantitative analyses by calculating values. Thus, their rubrics consist of levels without qualitative performance level descriptions that would make lesson plan quality transparent. Moreover, they explicitly focus on scientific inquiry or technology-enhanced biology teaching. Kademian and Davis' (2018) rubric assesses the extent to which preservice teachers plan to engage their students in science practices. Thus, their rubric is more generally applicable but still limited to engagement as a specific learning outcome. Only Jacobs et al. (2008) provide a generally helpful rubric beyond scientific inquiry or technological pedagogical content knowledge. Even though the authors provide a couple of arguments for validity, their rubric is only based on the authors' institutes' review protocol without explicitly using an underlying theory. Such a theoretical foundation might be necessary, though, to provide a set of criteria that does objectively cover an appropriate range of aspects that are assumed to be important in written lesson plans (e.g., the CODE-PLAN model; König et al., 2021). Neither Forbes and Davis (2010) nor Enugu and Hokayem (2017) nor Kademian and Davis (2018) nor von Kotzebue, (2022) refer to validity issues but exclusively to interrater reliability.



**TABLE 1** Overview of science education studies using an analytic scoring rubric in research on lesson planning.

		References				
Description		Forbes and Davis (2010)	Goldston et al. (2013)	Enugu and Hokayem (2017)	Kademian and Davis (2018)	von Kotzebue (2022)
Applicability	Generic	Scientific inquiry	scientific inquiry	Scientific inquiry	Engaging students in scientific practices	TPACK
Structure	20 criteria 3 weighted plds	5 criteria 4 plds	21 criteria 5 levels without plds	9 criteria 6 levels without plds	10 criteria 3 plds	5 criteria on instructional quality 3–4 levels without plds; 1 criterion on the implementation of digital technologies 4 levels without plds
Reliability	inter (%)	inter (%)	inter (%)	inter (%)	inter (%)	inter (k)
Validity	Correlation between rubric scores and scores from a lesson observation protocol and a questionnaire, respectively	-	Expert discussions, factor analysis	-	-	-
Qualitative insights into exemplary lesson plans	-	-	-	A good, a mediocre, and a poor example for two aspects of scientific inquiry each	Strengths and weaknesses in six participants' lesson plans	-

Abbreviations: inter (k), interrater agreement in terms of Cohen's k; inter (%), interrater agreement in terms of percentage; plds, performance level descriptions; TPACK, technological pedagogical content knowledge; -, indicates that this aspect is not considered in the referenced study.



Beyond science education, it seems to be common to focus on interrater reliability and at least validity evidence based on test content (e.g., expert discussions) in the development of scoring rubrics on lesson planning (see Backfisch et al., 2020; Ndiokubwayo et al., 2022; Ruys et al., 2012). For instance, Ndiokubwayo et al. (2022) draw the following conclusion regarding the validity of their interdisciplinary *lesson plan analysis protocol* (LPAP):

The study's design was motivated by the gap identified in the lack of tools to analyze pedagogical documents such as lesson plan [sic!] that reflect on the competency-based pedagogy. The draft protocol was developed first, together with its training manual [...]. The produced initial protocol was sent to different experts with considerable experience in the competence-based curriculum for validity purposes; this was done and yielded an improved version of the LPAP. After this stage, the reliability check process started; through this process, a very good LPAP was produced. The LPAP is a valid and reliable tool for teachers and educational evaluators. (p. 6)

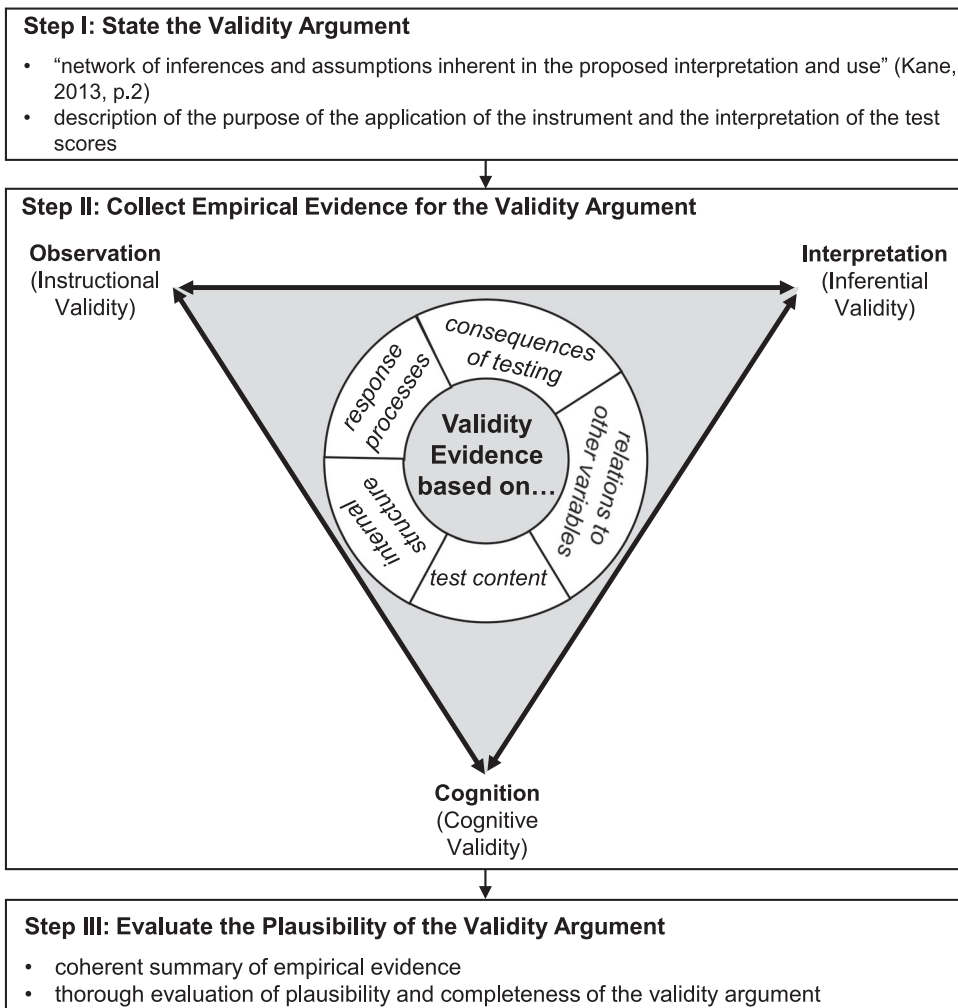
In this paper, we aim to build a validity argument for the RALP, and thereby, we illustrate that testing an instrument for validity is a challenging and complex endeavor (Kane, 2013); it is even understood as “the quest for the Holy Grail” (Scherer, 2017). Testing for validity needs to cover multiple aspects of validity (AERA, APA, & NCME, 2014) to build a sound and convincing validity argument and, thus, ensure that the scores can be interpreted validly.

### 3.3 | Toward a validity argument for the RALP

To ensure the quality of a scoring rubric, issues regarding objectivity, reliability, and validity should be considered in the rubric development process (Brookhart, 2018; Jonsson & Svingby, 2007; Moskal, 2000; Pellegrino et al., 2016). Mainly validity issues are often insufficiently addressed (Table 1; Reddy & Andrade, 2010). Validity “refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA, APA, & NCME, 2014, p. 11). It is a crucial prerequisite for the interpretation of empirical findings, yet it “is not a property of the test[,] [r]ather, it is a property of the proposed interpretations and uses of the test scores” (Kane, 2013, p. 3). Kane (1992) outlined the “argument-based approach to validity”: According to him, validation is characterized as a process of gaining empirical evidence and building an argument for the validity of the interpretations of the test scores. Three steps will be carried out in this study (Figure 2):

#### 3.3.1 | Step I: State the validity argument

First, it is necessary to make the validity argument or “interpretation/use argument”<sup>2</sup> (Kane, 2013, p. 2) transparent (Figure 2) and clearly state “why, how, by whom and in what contexts [the scoring rubric is intended to be] used” (Turley & Gallagher, 2008, p. 87): In the present study, we aim to ensure valid interpretations of the RALP scores. The RALP is intended to summatively assess the quality of preservice science teachers' (PSTs') and trainee science teachers' (TSTs') lesson plans (Supporting Information: Appendix; Großmann & Krüger, 2022a) within the framework of PCK (Carlson et al., 2019; Park & Oliver, 2008). More specifically, we intend for PSTs' teacher educators in the first and in the second phase of German teacher training to use the rubric as a framework for providing profound and differentiated feedback on their prospective science teachers' lesson plans. Consequently, PSTs' and TSTs' scores are interpreted as indicators of the lesson plan quality. For the sake of terminological clarity: In this study, we target prospective science teachers' ability to write proper lesson plans, which comprises both general lesson planning competence that is equally necessary for everyday lesson planning and the ability to justify their decisions.



**FIGURE 2** Building a validity argument in three steps (based on Kane, 1992, 2013). Assessment triangle was adapted from Pellegrino et al. (2016) and the allocation of the five sources of validity evidence from AERA, APA, & NCME (2014).

### 3.3.2 | Step II: Collect empirical evidence for the validity argument

Second, collecting empirical evidence that supports the validity argument is necessary (AERA, APA, & NCME, 2014; Kane, 2013). To build a sound, coherent, and comprehensive validity argument and cover a reasonable range of empirical evidence, we merged two frameworks of different grain-size that are concerned with validity issues: (1) Pellegrino et al.'s (2016) coarse-grained *Assessment Triangle* and (2) AERA, APA, & NCME, (2014) fine-grained “Standards for educational and psychological testing” (Figure 2).

- (1) Even though Pellegrino et al.'s (2016) framework explicitly refer to assessments close to classroom instruction, it appears to be eligible also in the context of lesson planning as a part of instruction in higher education (Drost & Levine, 2015; Karlström & Hamza, 2021). When applied in this context, the framework offers the advantage of strengthening the importance of an assessment instrument to (i) reflect the



knowledge domains to be assessed (*Cognition*; Figure 2). If the interpretation of the RALP scores reflects solely the ability to write a good lesson plan (and no other knowledge domains or competencies), the RALP would be characterized by a high degree of *cognitive validity*; (ii) be aligned to the professional standards in the curriculum and the teaching and learning before the assessment (*Observation*; Figure 2). If the RALP performance level descriptions match the intended learning outcomes regarding lesson planning in teacher training and cover the aspects of lesson planning that are part of instruction in teacher training, the RALP would be characterized by a high degree of *instructional validity*; (iii) yield reasonable conclusions about the observations made concerning reliability, appropriateness, and generalizability (*Interpretation*; Figure 2). If evidence can be collected for the RALP scores to provide reliable inferences about prospective teachers' competence to write a lesson plan employing statistical analyses, the RALP would be characterized by a high degree of *inferential validity*.

- (2) Within this *Assessment Triangle* (Pellegrino et al., 2016), different specific sources of valid evidence need to be taken into consideration (AERA, APA, & NCME, 2014):
- *Validity evidence based on test content* is needed to ensure that the RALP criteria are relevant and match teacher educators' expectations of good lesson plans (e.g., by expert ratings).
  - *Validity evidence based on response processes* is needed to ensure the target group can use the RALP (e.g., by documenting teacher educators' comprehension difficulties).
  - *Validity evidence based on internal structure* is needed to ensure that the RALP criteria show sufficient psychometric properties regarding the underlying conceptual frameworks of PCK and lesson planning (e.g., dimensionality, item discrimination).
  - *Validity evidence based on relations to other variables* is needed to ensure that the RALP scores correlate with scores from different methodological approaches that measure PCK or lesson planning competence. For instance, the experience might be a variable to test for, so comparing different samples that vary in their planning experience is necessary.
  - *Validity evidence based on the consequences of testing* is needed to ensure the interpretations of the RALP scores are appropriate for their intended use. For instance, high RALP scores are expected to reflect high lesson plan quality, and vice versa (i.e., "predictive validity" in outdated terminology of educational measurement), and PSTs/TSTs are expected to be able to improve their lesson plans after having received detailed formative feedback by their teacher educators. Moreover, high scores might be expected to correlate with the quality of teaching the lesson.

### 3.3.3 | Step III: Evaluate the plausibility of the validity argument

Finally, it is necessary to summarize the various pieces of empirical evidence and evaluate to what extent it is clear, coherent, and plausible and supports the validity argument. Kane (2013) points out that

the evaluation of evidence in the validity argument is not symmetric. To make a positive case for the proposed interpretations and uses of scores, the [validity argument] needs to provide adequate backing for all of the inferences [...] and to rule out challenges based on plausible alternative interpretations. However, a refutation of any core warrant can be decisive in undermining [a validity argument]. (p. 16)

The purpose of this study is to build a validity argument for the interpretations of the RALP as proposed by Kane (2013), whose argument-based approach demands researchers to lay out the validity argument and collect empirical evidence for it. Applied to the RALP, Table 2 shows multiple validity arguments for all three vertices of the

**TABLE 2** Validity arguments and related sources of validity evidence for the RALP.

Component of validity	Validity argument	Validity evidence ( <i>based on...</i> )
1. Cognition (cognitive validity)	<p>The RALP criteria reflect the expected cognitive construct (i.e., PCK needed for lesson planning) if ...</p> <ul style="list-style-type: none"> <li>- the criteria unpack relevant aspects of written lesson plans in science education.</li> <li>- the PCK components and their interconnections are aligned with relevant aspects of lesson planning.</li> </ul>	<p>1a. The RALP criteria build upon theoretical and empirical research on lesson planning in science education (<i>test content</i>).</p> <p>1b. Science teacher educators in the first phase of teacher education judge the theoretical and empirical foundation of the RALP criteria (<i>test content</i>).</p>
2. Observation (instructional validity)	<p>The observed RALP scores reflect the professional standards with regard to lesson planning if ...</p> <ul style="list-style-type: none"> <li>- Teacher educators consider the criteria meaningful and relevant for writing lesson plans in teacher training.</li> </ul> <p>The RALP criteria are suitable components of an analytic scoring rubric if ...</p> <ul style="list-style-type: none"> <li>- the criteria interconnect different PCK components.</li> <li>- the target group can meet the criteria.</li> <li>- the criteria help discriminate between high- and low-performing PSTs and TSTs, respectively.</li> </ul>	<p>2a. Science teacher educators in the second phase of teacher education judge that the RALP criteria are relevant in teacher training, that is, whether the rubric is fully aligned with the performance expectations at the end of teacher training (<i>test content</i>).</p> <p>2b. The manifest RALP criteria cannot be reduced to a small set of meaningful latent variables employing a dimensionality analysis (<i>internal structure</i>).</p> <p>2c. The item difficulty indicates that the RALP criteria are neither too difficult nor too easy to meet for PSTs and TSTs, respectively (<i>internal structure</i>).</p> <p>2d. The item discrimination indicates that the PSTs'/TSTs' overall RALP score correlates to their success on the single criteria (<i>internal structure</i>).</p>
3. Interpretation (inferential validity)	<p>The observed RALP scores account for science teachers' ability to plan lessons sophisticatedly if...</p> <ul style="list-style-type: none"> <li>- scores predict the lesson plan quality in terms of a holistic quality assessment by teacher educators.</li> <li>- the scores are sensitive to differences between PSTs and TSTs.</li> </ul>	<p>3a. RALP scores (0–48) and the holistic quality assessment by experienced teacher educators in terms of grades (1—very good, 2—good, 3—satisfactory, 4—sufficient, 5—poor) are negatively correlated (<i>consequences of testing</i>).</p> <p>3b. TSTs at the end of teacher training achieve higher RALP scores than PSTs (<i>relations to other variables</i>).</p>

Note: Procedure adapted from Zhai et al. (2021); components of validity derived from Pellegrino et al. (2016); sources of validity evidence derived from AERA et al. (2014).

Abbreviations: PSTs, preservice science teachers; RALP, rubric to assess science lesson plans; TSTs, trainee science teachers.

*Assessment Triangle* (Pellegrino et al., 2016) and the corresponding sources of validity evidence (AERA, APA, & NCME, 2014). It is important to note that “each type of evidence presented [above] is not required in all settings. Rather, support is needed for each proposition that underlies a proposed test interpretation for a specified use” (AERA, APA, & NCME, 2014, p. 14). In this study, we focus on some aspects relating to the relevance of the RALP criteria and their potential to assess lesson plan quality properly in a summative sense. By now, we focus on



ensuring that the RALP scores reflect lesson plan quality. We have not yet intended to investigate how far teacher educators can apply the RALP. This might be a subsequent step in the validation process. Hence, this paper does not provide information concerning *validity evidence based on response processes*.

## 4 | PROBLEM STATEMENT AND RESEARCH QUESTIONS

Although lesson plans are crucial, for example, in the German teacher training system (KMK, 2019; König et al., 2020; Neumann et al., 2017), a theoretically framed scoring rubric applicable to all kinds of science lessons and tested for validity (Table 1) comprising empirically tested criteria to evaluate the quality of written lesson plans (Krepf & König, 2022) are still missing. Such a rubric would offer a twofold advantage: First, it might facilitate comparability between teacher educators' assessment of lesson plan quality in teacher education and thus reduce the impact of teacher educators' subjective preferences. Second, it would allow researchers in science education to use an established and evaluated instrument to assess lesson plan quality in qualitative and quantitative studies.

This study pursues two objectives: First, we aim to evaluate the RALP quality by building a valid argument and collecting empirical evidence (Table 2). The following research questions will be addressed:

RQ 1.1: To what extent do science teacher educators consider the RALP criteria relevant for writing lesson plans? (*Validity evidence based on test content*).

RQ 1.2: To what extent does the scoring rubric allow for reliable measures?

RQ 1.2a: To what extent is the RALP scoring procedure stable over a 2-week interval when applying the scoring rubric to PSTs' and TSTs' lesson plans? (*Intrarater reliability*).

RQ 1.2b: To what extent do a science teacher educator and a trained student assistant reach an intersubjective agreement when applying the RALP to TSTs' lesson plans? (*Interrater reliability*).

RQ 1.3: To what extent do data reflect the intended interconnectedness of PCK components? (*Validity evidence based on internal structure*).

RQ 1.4: To what extent is the selection of criteria appropriate to assess lesson plan quality? (*Validity evidence based on internal structure*).

RQ 1.5: To what extent do the RALP scores correlate with science teacher educators' holistic quality assessment? (*Validity evidence based on the consequences of testing*).

RQ 1.6: To what extent do PSTs and TSTs achieve different scores? (*Validity evidence based on relations to other variables*).

Second, after ensuring that RALP scores can be interpreted validly, we aim to provide a qualitative insight into prospective science teachers' lesson planning competence:

RQ 2: What are the most significant differences between PSTs' and TSTs' lesson plans?

## 5 | METHODS

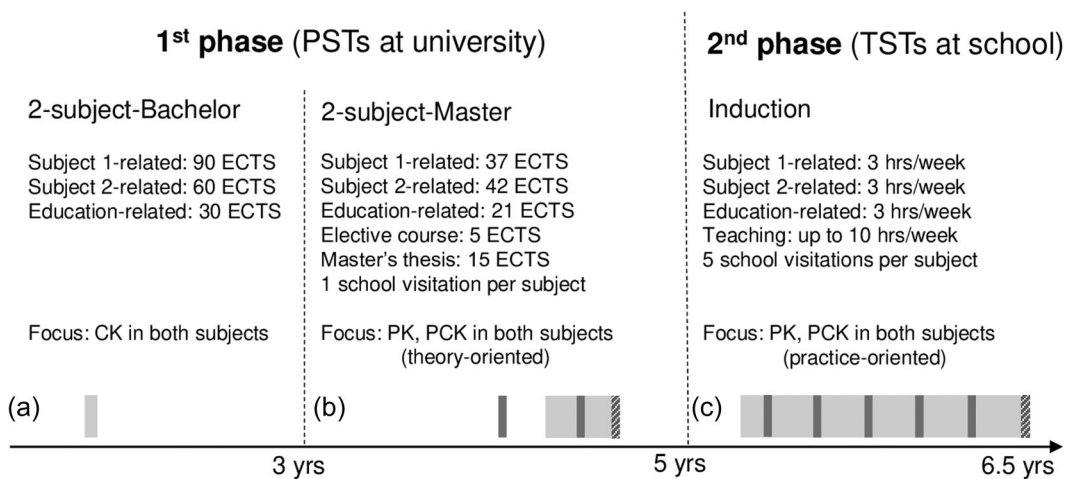
### 5.1 | Study context

German science teacher education is structured both concurrently and consecutively (Kotthoff & Terhart, 2013): It is concurrent because, from the very beginning, prospective teachers are confronted with all relevant elements (educational studies, CK and PCK in their two subjects, school practice, planning, teaching, reflecting). It is consecutive since teacher training consists of two distinct phases (Figure 3): The first phase is located at university (PSTs), is focused on theory, and lasts 5 years (Bachelor: 3 years, Master: 2 years). During the Master's studies, PSTs must complete a 5-month practical semester at school, where a science teacher educator visits them to give feedback on their planning, teaching, and reflection competence (Figure 3b). At the end of the practical semester,

and after being taught how to write lesson plans, PSTs write a lesson plan for a graded paper. The second phase is the induction phase, located at schools (*in-service trainee science teachers, TSTs*) accompanied by state-run seminars, and lasts another 1.5 years (Kotthoff & Terhart, 2013; Neumann et al., 2017). During the second phase, science teacher educators visit the TSTs at schools five times per subject to give feedback on their planning, teaching, and reflection competence (Figure 3c). As their final examination, TSTs must plan, teach, and reflect on one lesson each in both their subjects. In our case, their performance was only evaluated based on a lesson plan discussion, as the Covid-19 pandemic made it impossible to assess their teaching. An expert committee consisting of up to five science teacher educators working in the second phase of teacher education and the school principal carried out this evaluation. According to the standard certification requirements, prospective science teachers are intended to be able to plan lessons under consideration of different learning dispositions among students and teach effectively. More specifically, they are intended to acknowledge students' diversity, plan lessons accordingly, interconnect their CK, PK, and PCK to choose appropriate topics, media, and activities, and assess the quality of their teaching practice (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2019). If passed, TSTs receive a certificate, which qualifies them to work as full-time in-service teachers in Germany. Teacher education programs vary across the 16 federal states of Germany (Kotthoff & Terhart, 2013); the structure described here only holds for Berlin, the federal state where we conducted this study.

## 5.2 | Sample and data collection

For this study, we analyzed lesson plans ( $N = 100$ ) written by prospective biology teachers (PSTs, TSTs). This sample comprises lesson plans written by PSTs ( $n = 36$ ) who attended the two authors' courses at university. Lesson plans ( $n = 64$ ) written by TSTs due to their final examinations ("Second State Examination"; Kotthoff & Terhart, 2013) of



**FIGURE 3** Overview of the structure of teacher training in Germany. Light gray boxes indicate practical parts of teacher training, dark gray boxes indicate instances when PSTs/TSTs have to write a lesson plan, and striped dark gray boxes indicate samples analyzed in this study: (a) 4-week orientation internship in the first year of teacher training (no lesson plan); (b) 5-month practical semester in the fourth year of teacher training (three lesson plans); (c) 1.5-year induction phase at the end of teacher training (five lesson plans and another lesson plan in the course of the final examination). The authors of this paper work in the first phase at university and thus do not have any influence on the second phase. One ECTS credit equals 30 h of workload. CK, content knowledge; ECTS, European Credit Transfer and Accumulation System; hrs, hours; PCK, pedagogical content knowledge; PK, pedagogical knowledge; PSTs, preservice science teachers; TSTs, trainee science teachers; yrs, years.



German teacher training in the years 2019–2021 (Figure 3c, striped dark gray box) were provided anonymously by the Ministry of Education. Due to that, we could not collect demographic information about the participants or the schools they teach at. Since the RALP (Supporting Information: Appendix A) is intended to display performance level descriptions up to the highest level that can be expected for each criterion, this sample can be regarded as suitable: Lesson planning competence develops in the course of teacher training (Backfisch et al., 2020; König et al., 2021, 2022; Mutton et al., 2011; Vogelsang et al., 2022; Westerman, 1991). Moreover, the focus of teacher education shifts from acquiring CK during Bachelor studies (Figure 3a) to PK and PCK during Master studies (Figure 3b) and particularly during the practice-oriented induction phase (Figure 3c; Kotthoff & Terhart, 2013). Consequently, we do not assume Bachelor's or Master's students to be able to achieve the highest levels. Moreover, the authors did not influence the development of these lesson plans since the state, and not teacher educators at the university, run the induction phase. Therefore, the TSTs followed the official guidelines of the school administration (SenBJF, 2017), which is not associated with the authors' affiliations. Hence, data collection is characterized by a high degree of ecological validity, “the extent to which research findings would generalize to settings typical of everyday life” (Wegener & Blankenship, 2007, p. 275). For the sake of clarity: The term “lesson plan” used in this study does explicitly not refer to short, 1–2 pages long sketches of lessons (e.g., Morine-Dersheimer, 2011). Instead, a regular lesson plan in German teacher training covers approximately 10 pages (plus appendices).

To answer RQ 1.5, RQ 1.6, and RQ 2, a comparison to another sample is required to investigate whether the RALP is sensitive to the assumed differences in perspective science teachers' PCK and lesson planning competence (Figure 3). Therefore, lesson plans ( $n = 36$ ) were collected from PSTs' graded papers written after their practical semester in 2019–2021 (Figure 3b, striped dark gray box). For the data analysis, these lesson plans were anonymized. Both PSTs' and TSTs' lesson plans were written for graded exams. They prepared their lesson plans with no time constraints out of school and out of their seminars. Consequently, the lesson plans were not written under controlled conditions. We have no information on how long the PSTs and TSTs worked on their lesson plans, which resources they used, and whether they asked colleagues for feedback.

### 5.3 | Data analysis

Using Creswell and Plano Clark's (2018, pp. 77–84) denotation, our approach could be described as a QUAN → qual explanatory sequential design that aims to understand better the quantitative data used for evaluating validity. Consequently, we first analyzed mainly quantitative data collected by application of the RALP (RQ 1) and second shed light on selected relevant aspects in qualitative analysis (RQ 2). In detail, the data were analyzed as described below.

To ensure that the criteria are relevant for writing lesson plans (RQ 1.1; *Validity evidence based on test content*; AERA, APA, & NCME, 2014), we led expert discussions with the two authors' colleagues and afterward individual discussions with two professors of biology education, both working in different federal states of Germany. Since all federal states are individually responsible for their education policy in Germany, we thus aimed to increase the generalizability of the RALP scores. To ensure that teacher educators consider the criteria relevant, all 25 teacher educators working in the second phase in the same federal state as the authors of this paper (Berlin) were asked via e-mail to participate in a survey. 10 teacher educators agreed to participate. They were asked to rate the relevance of an initial set of 29 criteria on a 4-point Likert scale (1—*irrelevant*, 2—*barely relevant*, 3—*rather relevant*, 4—*very relevant*). In contrast to the teacher educators working in the first phase at university, the teacher educators working in the second phase are in-service teachers who run seminars for TSTs and form part of the examination board for the final examination of teacher training (Figure 3). For the rating, we only presented the highest performance level description of each criterion, as this is what TSTs are expected to achieve at the end of their teacher training.

As a prerequisite for further validity analysis, we needed to ensure that the scoring procedure was objective and reliable. To address RQ 1.2a and RQ 1.2b, we calculated weighted Cohen's  $\kappa$  as a measure of rater agreement (Landis & Koch, 1977), taking into account the degree of disagreement between two raters (Gwet, 2014). This paper will address two kinds of reliability: (a) For intrarater reliability (RQ 1.2a), the first author applied the RALP to 18/100 randomly selected lesson plans twice in a 2-week interval. Weighted Cohen's  $\kappa$  was calculated. The first author resolved disagreements between his two codings and reached a final version. (b) The interrater reliability (RQ 1.2b) of two different raters (the first author and a trained student assistant) was calculated as well by using weighted Cohen's  $\kappa$ . Again, disagreements in codings were discussed and resolved.

Afterward, we evaluated different sources of validity evidence (AERA, APA, & NCME, 2014) of the RALP: First, to investigate the empirical structure of the RALP (RQ 1.3; *validity evidence based on internal structure*; AERA, APA, & NCME, 2014), we performed an analysis within the framework of item response theory (IRT; Bond & Fox, 2001) using the software ACER Conquest (version 1.0.0.1; Wu et al., 2007). As an extension of the simple logistic model within IRT, we applied the partial credit model (Masters, 1982) to calculate item difficulties and personal abilities because the RALP criteria are polytomous (i.e., levels 0–2). As described above, the 24 criteria of the RALP mainly address interconnections between four of the five PCK components (Park & Oliver, 2008). As we differentiated between criteria that focus on the intended learning outcomes (O1–O4) and those that focus on the biological subject matter (C1–C4) within KSC (see Supporting Information: Appendix), we divided this component into two subcomponents for the statistical analysis (i.e., KSC—focus on intended learning outcomes, KSC—focus on biological content). Thus, we differentiated five components and neglected OTS. As we intend the criteria to reflect the interconnectedness, we would not assume to find five latent dimensions in the data but rather a one-dimensional, global latent dimension (i.e., the ability to write good lesson plans). Thus, data would reflect the intended interconnectedness of PCK components. In addition, two further models were specified and compared: a two-dimensional model separating all criteria with explicit reference to the intended learning outcome and those without, and a four-dimensional model that considers the relative closeness of criteria to each other, as shown in Figure 1. To decide which model fits best, we calculated the following values based on the analysis of  $N = 100$  lesson plans. Thereby, we followed an established procedure in science education research (e.g., Großschedl et al., 2014): The factor of final deviance indicates to what extent the collected data fit the underlying assumptions of each model. To test which model fits best, we calculated two descriptive estimates, namely Akaike (1981) information criterion (AIC) and the Bayesian information criterion (BIC; Schwarz, 1978). Considering the parsimony of the models, it is assumed for both AIC and BIC that the lower the value, the better the data fit the model (Wilson et al., 2008). To decide which model fits best to the data, we performed  $\chi^2$  tests. For that purpose, the difference between the deviances and the degrees of freedom as differences in the number of parameters were calculated, finally leading to an estimation of whether the two models are significantly different from each other (Bentler, 1990). Each model's reliabilities for each dimension regarding the expected a posteriori/plausible values will be reported.

Second, to gain further insight into the psychometric appropriateness of the criteria (RQ 1.4; *validity evidence based on internal structure*; AERA, APA, & NCME, 2014), the item discrimination index ( $a_i$ ) indicates how far the criteria can differentiate between persons with different abilities in terms of the correlation between the persons' scores on a particular criterion and their overall score. As a rule of thumb, items exceeding the threshold of 0.30 can be regarded as sufficiently discriminatory (Vaus, 2014). In addition, we will provide a Wright map (Boone et al., 2014) illustrating how far the PSTs' and TSTs' abilities match the difficulty of the criteria. It would be desirable if the average person's abilities were close to the average item difficulty.

Third, we aimed to investigate whether a high score reflects a high overall lesson plan quality (*validity evidence based on the consequences of testing*; AERA, APA, & NCME, 2014), that is, predictive validity (RQ 1.5). For that purpose, we calculated the Pearson correlation between the teacher educators' holistic quality assessments for the 36 PSTs' lesson plans and the teacher educator's holistic quality assessments for the 64 TSTs' lesson plans on the





one hand and the RALP scores on the other. The validity argument mentioned above (Table 2) refers to the purpose of the RALP as an instrument for teacher educators to assess prospective science teachers' lesson plan quality. Consequently, the RALP needs to match teacher educators' perspectives, and its scores need to reflect teacher educators' holistic quality assessment. As established in Germany, the following range of grades was used to assess lesson plan quality holistically: 1.0/1.3—very good; 1.7/2.0/2.3—good; 2.7/3.0/3.3—satisfactory; 3.7/4.0—sufficient; 4.3/4.7/5.0—poor. As mentioned above, the authors of this paperwork in the first phase of teacher education at the university, while the state runs the second phase (Figure 3). Consequently, the holistic quality assessments of PSTs' and TSTs' lesson plans differed slightly: PSTs' lesson plans ( $n = 36$ ) were assessed by six teacher educators working in the first phase at a German university (three professors, two postdoctoral researchers, one PhD student). One professor and the PhD student are the authors of this paper. To ensure an objective grading procedure, the six teacher educators used the same grading scheme with various criteria that PSTs were expected to consider in their lesson plans (Table 3). TSTs' lesson plans ( $n = 64$ ) were assessed by a retired biology teacher who served as a teacher educator in the second phase for 21 years (Figure 3). The teacher educator was instructed to grade the 64 lesson plans based on the holistic criteria he applied (Table 3) when he was still in service and part of official examination committees. In contrast to the PSTs, TSTs were not explicitly expected to consider students' conceptions. Still, they were instructed to provide a tabular unit and clarify what competencies they wanted to foster during this unit.

Two steps were taken to test the teacher educator's objectivity and trustworthiness. (1) We calculated the Pearson correlation between the teacher educator's grades and the original grades we obtained for no more than 17 of those 64 lesson plans provided by the Ministry of Education. (2) We hired a second teacher educator working in the second phase who graded a subsample of lesson plans as already described in Großmann and Krüger (2022c) and calculated the Pearson correlation coefficient again. Moreover, the teacher educator was asked to elaborate on all 64 lesson plans and give insights into the reasons for the grades he decided to give in a series of interviews. Both teacher educators were unfamiliar with the authors' affiliations and were unfamiliar with the RALP. Thus, we ensured that the holistic quality assessment could serve as a valid external criterion to evaluate the quality of the RALP.

PSTs have to meet the criteria established by their teacher educators at the university; TSTs have to meet the official guidelines for the second phase of teacher training and the final examination of teacher education (SenBJF, 2017).

**TABLE 3** Requirements of written lesson plans.

Requirements	PSTs	TSTs
Analysis of biological subject matter	x	x
Intended learning outcome	x	x
Unit contextualization		x
Students' level of competence (internal differentiation)	x	x
Students' conceptions of the subject matter	x	
Description and justification of planning decisions (e.g., tasks, activities)	x	x
Tabular schedule of the lesson	x	x
Appendices (e.g., worksheets)	x	x

*Note:* The requirements differ slightly between PSTs' lesson plans ( $n = 36$ ) and TSTs' lesson plans ( $n = 64$ ). PSTs have to meet the criteria established by their teacher educators at the university; TSTs have to meet the official guidelines for the second phase of teacher training and the final examination of teacher education (SenBJF, 2017).

Abbreviations: PSTs, preservice science teachers; TSTs, trainee science teachers.

Since holistic grading might always be subjective to a certain extent, we aimed to decrease the significance of these grades by dividing both subsamples into quartiles based on the holistic quality assessment: The top 25% ( $n_{\text{PSTs}} = 9$ ;  $n_{\text{TSTs}} = 16$ ), the middle 50% ( $n_{\text{PSTs}} = 18$ ;  $n_{\text{TSTs}} = 32$ ), and the bottom 25% ( $n_{\text{PSTs}} = 9$ ;  $n_{\text{TSTs}} = 16$ ) were compared concerning the RALP scores they achieved. Thus, we focus on the ranks within the sample rather than on the teacher educator's exact grades. Mann–Whitney  $U$  tests were calculated to determine whether these quartiles' differences are significant.

Fourth, to investigate the instructional sensitivity of the RALP (RQ 1.6; *validity evidence based on relations to other variables*; AERA, APA, & NCME, 2014), the Mann–Whitney  $U$  test was applied to compare whether the two subsamples of PSTs ( $n = 36$ ) and TSTs ( $n = 64$ ) differ significantly regarding their RALP scores.

After ensuring that the RALP scores allow for valid interpretations of scores, we intend to provide qualitative insights into prospective science teachers' lesson planning competence. More specifically, based on the assumption that lesson planning competence develops during teacher education (Backfisch et al., 2020; König et al., 2021, 2022; Mutton et al., 2011; Vogelsang et al., 2022; Westerman, 1991), we aim to illustrate in detail how PSTs and TSTs approach significant challenges in the process of lesson planning. First, we compared the PSTs' scores with the TSTs' scores, which were generally higher, and identified the three criteria showing the largest effect sizes. Second, we applied the *typical case sampling strategy* (Patton, 1990, pp. 173–174): We purposefully selected illustrative extracts from TPSTs' and PTSTs' lesson plans for each of the three criteria. With these cases, we were especially interested in differences in how PSTs and TSTs tend to plan a lesson and justify their decisions.

## 6 | RESULTS

### 6.1 | Relevance of criteria (RQ 1.1; test content)

The teacher educators' feedback provides evidence that the RALP criteria are well-aligned with professional standards and cover a range of relevant lesson-planning aspects. Since terminology is sometimes equivocal in German educational literature, some expressions were changed to be more generally comprehensible. In particular, feedback helped us describe sophisticated performances at the highest level.

Most criteria were considered rather relevant by the 10 teacher educators (Table 4). Two criteria were considered irrelevant (C2: use of references to educational literature,  $M = 1.45$ ; S3: consideration of students' conceptions,  $M = 1.55$ ). C2 was refined so that the focus now lies on the topic's relevance rather than how prospective teachers justify their choice; it no longer forms a single criterion but was added to the highest performance level description for some criteria. S3 was not deleted due to its relevance in research on science teacher education. More importantly, several teacher educators stressed that students' conceptions might be necessary for some lessons but not all. We do not consider this a convincing argument and thus decided to maintain S3 in the RALP. Furthermore, teacher educators suggested adding the selection of technical terms (C4) as an essential aspect of lesson planning, provided ideas for more precise formulations, and suggested further merges of criteria, which led to a reduction of 29 initial criteria to 24. These remaining criteria are shown in Supporting Information: Appendix A and were used for further data analysis.

### 6.2 | Stability and objectivity (RQ 1.2a, RQ 1.2b; intra- and interrater reliability)

On average, weighted Cohen's  $\kappa$  indicated “almost perfect” intra- ( $M = 0.93$ , range = 0.77–1.0) and interrater agreement ( $M = 0.88$ , range = 0.65–1.0; Landis & Koch, 1977, p. 165), indicating that the scoring procedure is both stable and intersubjectively comprehensible. In addition, a criteria-wise analysis of the reliability coefficients (Table 4) led to the identification of single criteria that were difficult to apply for the first coder (e.g., C1, I3, I7) or



**TABLE 4** Teacher educators' relevance assessment (mean and standard deviation) on a 4-point Likert scale, intrarater reliability, interrater reliability (weighted Cohen's  $\kappa$ ), and absolute frequency of the final codings, sorted by criterion.

RALP criteria		$M_{TE}$	$SD_{TE}$	$\kappa_{intra}$	$\kappa_{inter}$
O1	Intended learning outcome in the unit	2.18	1.00	0.89	0.89
O2	Progression throughout the unit	1.91	0.89	0.91	0.82
O3	Intended learning outcome in the lesson	2.91	0.30	0.89	0.58
O4	Indicators as evidence of the desired learning	2.64	0.49	0.91	0.72
C1	Analysis of biological content	2.00	0.60	0.70	0.76
C2	Choice of topic	1.45	0.92	0.91	0.82
C3	Educational reconstruction	1.91	1.00	0.92	1.00
C4	Selection of relevant technical terms	n/a	n/a	0.91	0.69
S1	Analysis of the level of competence	2.27	0.50	1.00	0.83
S2	Progression of competence development	2.82	0.30	1.00	0.67
S3	Students' conceptions	1.55	0.80	1.00	1.00
S4	Learning difficulties	2.64	0.46	1.00	0.88
S5	Methodical skills	2.64	0.64	1.00	0.83
I1	Structure of the development of competencies	2.70	0.46	0.88	0.57
I2	Lesson structure	n/a	n/a	1.00	0.80
I3	Suitability of the methods	2.90	0.30	0.20	0.79
I4	Suitability of tasks	2.50	0.50	1.00	0.81
I5	Suitability of materials	2.70	0.46	0.87	0.76
I6	Horizon of expectations	2.50	0.67	0.83	1.00
I7	Adaptive teaching	2.90	0.30	0.70	0.78
I8	Anticipated difficulties and alternatives	2.30	0.46	1.00	1.00
I9	Transparency of the learning process	2.80	0.40	1.00	1.00
A1	Transparency of performance expectations	2.56	0.68	0.91	0.71
A2	Products of students' learning	2.70	0.64	0.77	0.89

Note: C4 and I2 were added in response to the teacher educators' feedback, so no Likert scale results are reported. Abbreviations: n/a, not applicable; RALP, rubric to assess science lesson plans.

the second coder (e.g., O3, C4, I1). The formulations of these criteria were then further checked for imprecise or confusing expressions and clarified.

### 6.3 | Dimensionality (RQ 1.3; internal structure)

The decisive information-based criteria AIC and BIC are lower for the one-dimensional model than for the remaining three models (Table 5).  $\chi^2$  tests show that the one-dimensional model fits the data significantly better than the two-dimensional model ( $\chi^2 [2] = 7.72, p < 0.05$ ). In contrast, there is no significant difference to the

**TABLE 5** Fit statistics for the four PCM models.

Model	Parameters	Deviance	AIC	BIC	EAP/PV
One dimensional	25	4859.34	4909.34	4974.46	0.50
Two dimensional	27	4867.05	4921.05	4991.39	0.48/0.53
Four dimensional	34	4860.59	4928.59	5017.16	0.57/0.45/0.58/0.32
Five dimensional	39	4869.22	4947.22	5048.82	0.59/0.49/0.45/0.62/0.36

Abbreviations: AIC, Akaike (1981) information criterion; BIC, Bayesian information criterion (Schwarz, 1978); EAP/PV, expected a posteriori/plausible values; PCM, partial credit model.

four-dimensional ( $\chi^2 [9] = 1.25, p = 0.99$ ), and the five-dimensional model ( $\chi^2 [14] = 9.88, p = 0.77$ ). The psychometric analysis supports the assumption that a one-dimensional model fits the data best.

## 6.4 | Appropriateness of criteria (RQ 1.4; internal structure)

Overall, absolute codings for each criterion show that all criteria are achievable in both subsamples (Table 6). With the exceptions of unit contextualization (O1, O2) in PSTs' lesson plans and considering students' conceptions in TSTs' lesson plans (S3), levels 1 and 2 were achieved by many prospective science teachers in both subsamples.

All but four criteria exceed the threshold of 0.30 and can thus be regarded as sufficiently discriminatory (Vaus, 2014). The remaining criteria (O4, S3, S5, and I4) barely help discriminate PSTs' and TSTs' performances.

The Wright map (Figure 4) gives further insights into the relationship between the difficulty of the proposed criteria and the target group's abilities. Regarding the difficulty of the 24 criteria (right side), three observations are noteworthy: First, most criteria lie around the middle of the distribution. Hence, they can be assumed to be moderately difficult. Second, there is a gradient of difficulty. Hence, some criteria are rather difficult (e.g., I9, I4, O2), and some are relatively easy (e.g., I3, I2, O3). Third, no persons in this sample have a 50% solution probability for the two criteria, students' conceptions (S3) and horizon of expectations (I6). Hence, we conclude they are either too difficult or too easy.

## 6.5 | Holistic quality assessment (RQ 1.5; consequences of testing)

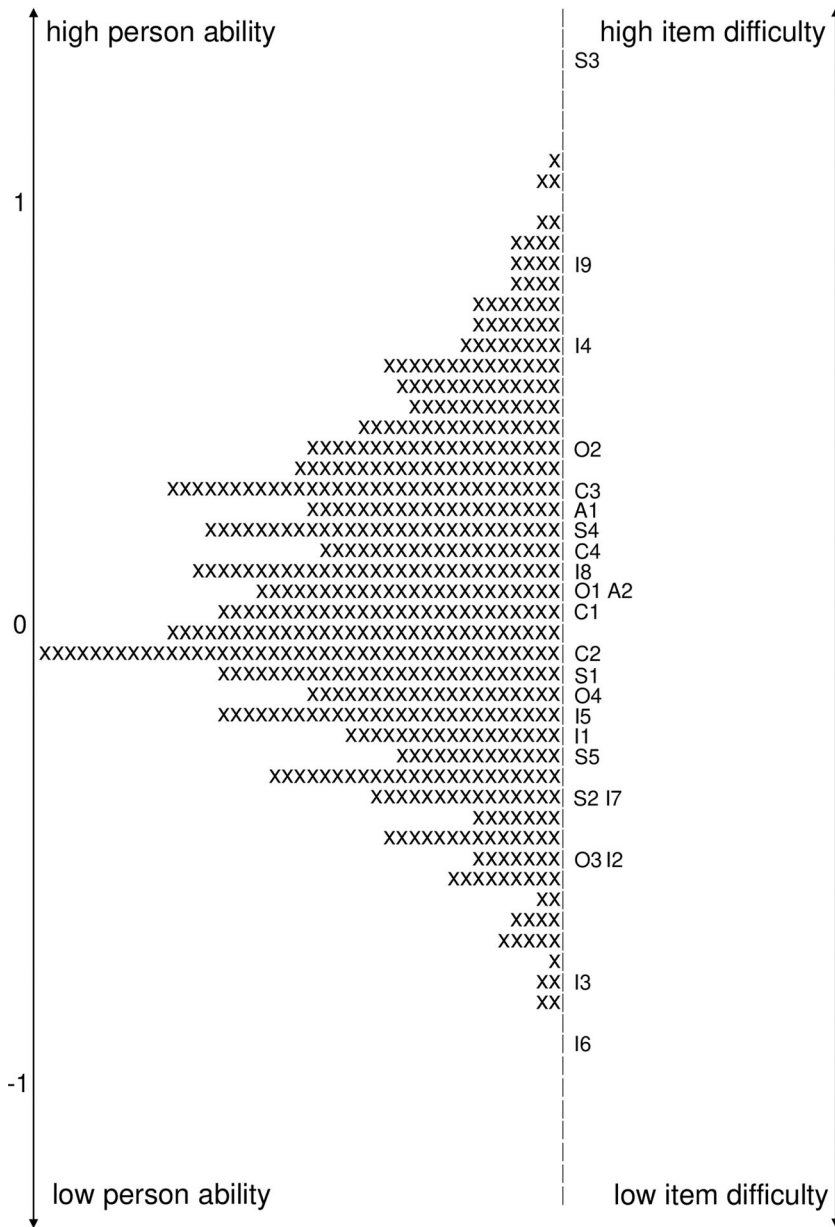
The teacher educators in the first phase and the external teacher educator in the second phase provided a holistic quality assessment in the form of grades for all lesson plans ( $N = 100$ ). A second teacher educator graded a subsample and provided a detailed qualitative review to ensure intersubjective comprehensibility and reduce the impact of the external teacher educator's subjective preferences and biases in grading. Their grades only differed slightly. We learned from their reviews that they appreciated and criticized the same aspects. Different grades resulted from the different weights they gave to certain aspects. Moreover, another subsample of 17 lesson plans written by TSTs and graded by the official examination committee correlates moderately with the teacher educator's grades ( $r = 0.43, p = 0.08$ ). Based on the  $p$  value, the null hypothesis (i.e., there is no correlation between the grades) cannot be rejected. It can be assumed that calculating the correlation for more lesson plans might result in a significant relationship. However, we did not obtain more than 17 lesson plans from the Ministry of Education to check that. Since the two teacher educators reached intersubjective comprehensibility, we interpret this as sufficient evidence for convergent validity, which allows us to trust the teacher educator's grades in terms of objectivity.

**TABLE 6** Absolute codings for PSTs ( $n = 36$ ) and TSTs ( $n = 64$ ) for levels 0–2 per criterion and item discrimination ( $a_i$ ).

RALP criteria		PSTs			TSTs			$a_i$
		0	1	2	0	1	2	
O1	Intended learning outcome in the unit	33	1	2	3	30	31	0.64
O2	Progression throughout the unit	34	2	0	9	36	19	0.68
O3	Intended learning outcome in the lesson	4	15	17	2	35	27	0.33
O4	Indicators as evidence of the desired learning	5	20	11	11	35	18	0.26
C1	Analysis of biological content	3	26	7	10	46	8	0.34
C2	Choice of topic	16	14	6	16	15	33	0.45
C3	Educational reconstruction	11	22	3	19	34	11	0.36
C4	Selection of relevant technical terms	13	18	5	6	52	6	0.38
S1	Analysis of the level of competence	13	14	9	16	18	30	0.46
S2	Progression of competence development	5	26	5	4	28	32	0.57
S3	Students' conceptions	16	12	8	60	3	1	-0.14
S4	Learning difficulties	16	17	3	9	44	11	0.54
S5	Methodical skills	2	22	12	10	31	23	0.21
I1	Structure of the development of competencies	11	17	8	1	40	23	0.43
I2	Lesson structure	2	20	14	2	34	28	0.45
I3	Suitability of the methods	2	20	14	0	25	39	0.48
I4	Suitability of tasks	14	14	8	38	18	8	0.22
I5	Suitability of materials	5	25	6	0	48	16	0.44
I6	Horizon of expectations	7	11	18	5	7	52	0.52
I7	Adaptive teaching	14	16	6	3	20	41	0.48
I8	Anticipated difficulties and alternatives	11	23	2	9	43	12	0.40
I9	Transparency of the learning process	30	4	2	22	38	4	0.43
A1	Transparency of performance expectations	18	15	3	9	46	9	0.38
A2	Products of students' learning	14	18	4	2	50	12	0.33

Abbreviations: PSTs, preservice science teachers; RALP, rubric to assess science lesson plans; TSTs, trainee science teachers.

Overall, there is a strong correlation between the RALP scores and the teacher educators' holistic quality assessment for the PSTs ( $r = -0.69$ ,  $p < .001$ ), as well as for the teacher educator's holistic quality assessment for the TSTs ( $r = -0.72$ ,  $p < 0.001$ ). Both relationships are negative because low values for the holistic quality assessment indicate high quality (grades 1.0–2.3), and higher values indicate low quality (grades 4.3–5.0). Hence, a high RALP score correlates to a low value for the holistic quality assessment and vice versa. To gain further insight into the ability of the RALP to detect quality differences among lesson plans, we divided both subsamples into the 25% of highest, the 50% of mediocre, and the 25% of poorest lesson plan quality based on the teacher educators' holistic quality assessment. These quality groups were plotted against the overall score reached by application of the RALP (Figure 5).



**FIGURE 4** Wright map for the analysis of lesson plans ( $N = 100$ ). On the left side, the person's abilities are shown. Each x equals 0.2 PSTs/TSTs. The closer an x is on the top, the more able the person is, and vice versa. On the right side, the item difficulties are shown. Criteria at the bottom represent those that are easier to achieve and vice versa. If an x occurs at the same position as a criterion, the person has a 50% probability of answering the item correctly. PSTs, preservice science teachers; TSTs, trainee science teachers.

Within the PST subsample, two groups of lesson plan quality differ significantly from each other: The top 25% ( $Mdn = 22$ ) scored significantly higher than the bottom 25% ( $Mdn = 15$ ) in the RALP ( $U = 5$ ,  $z = -3.15$ ,  $p < 0.001$ ,  $d = 2.28$ ; large effect), and the latter performed significantly worse than the middle 50% ( $Mdn = 21$ ;  $U = 6.5$ ,  $z = -3.851$ ,  $p < 0.001$ ,  $d = 2.18$ ; large effect).

Within the TST subsample, all three groups of lesson plan quality differ significantly in their RALP scores: The top 25% ( $Mdn = 33$ ) scored significantly higher than the middle 50% ( $Mdn = 28$ ) in the RALP ( $U = 89.5$ ,  $z = -3.65$ ,  $p < 0.001$ ,  $d = 1.24$ ; large effect), and the middle 50%, in turn, scored significantly higher than the bottom 25% ( $Mdn = 20.5$ ) in the RALP ( $U = 66.5$ ,  $z = -4.2$ ,  $p < 0.001$ ,  $d = 1.49$ ; large effect). As a logical consequence, the difference between the top 25% and the bottom 25% is highly significant ( $U = 2$ ,  $z = -4.76$ ,  $p < 0.001$ ,  $d = 3.09$ ; large effect).

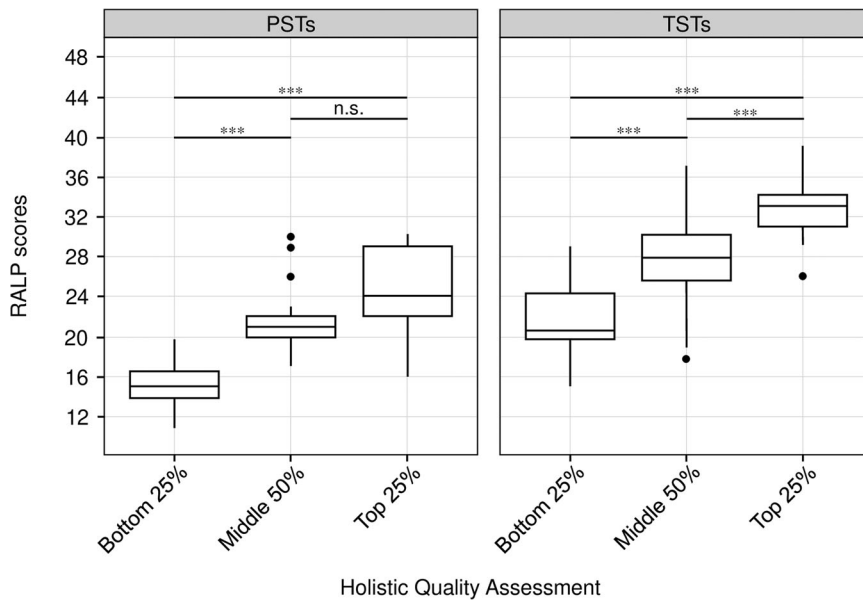
## 6.6 | Instructional sensitivity (RQ 1.6; relations to other variables)

Comparing both subsamples of PSTs ( $n = 36$ ) and TSTs ( $n = 64$ ), we find that TSTs ( $Mdn = 28$ ) score significantly higher in the RALP than the PSTs ( $Mdn = 20$ ;  $U = 421$ ,  $z = -5.26$ ,  $p < 0.001$ ,  $d = 1.23$ ; large effect). Descriptive statistics are shown in Table 7.

We find that scores range between 11 and 39, indicating that even low-performing participants reach 25% of the maximum score. In contrast, the highest-performing participant only reaches 81.25% of the maximum score.

## 6.7 | Exemplary cases (RQ 2)

Mann-Whitney  $U$  tests indicate that TSTs score significantly higher in 13 of the 24 criteria (O1, O2, C2, S2, S4, I1, I3, I6, I7, I8, I9, A1, and A2) than PSTs ( $0.41 < d < 2.01$ ), and PSTs score significantly higher only in S3 than TSTs. Among those criteria, the effect sizes for O1 ( $d = 2.01$ ), O2 ( $d = 1.85$ ), and S3 ( $d = 0.90$ ) can be explained with the



**FIGURE 5** Rubric to assess science lesson plans (RALP) scores of preservice science teachers (PSTs) and trainee science teachers (TSTs) per lesson plan quality group based on the holistic quality assessment. Sample size PSTs: bottom 25% and top 25%  $n = 9$ ; middle 50%  $n = 18$ ; sample size TSTs: bottom 25% and top 25%  $n = 16$ ; middle 50%  $n = 32$ . Possible RALP scores range between 0 and 48. Thick lines in boxes indicate medians and lower and upper box boundaries indicate 25th and 75th percentiles. Lower and upper vertical lines indicate the 10th and 90th percentiles. Dots indicate outliers outside the 10th and 90th percentiles. n.s., not significant; \*\*\* $p < 0.001$ .

different requirements for lesson plans within our sample (Table 3), so these criteria will not be considered. Among the remaining criteria, the three criteria with the largest effect sizes were *Choice of topic* (C2;  $d = 0.62$ ), *Progression of competence development* (S2;  $d = 0.66$ ), and *Adaptive teaching* (I7;  $d = 1.08$ ). To provide qualitative insights into PSTs' and TSTs' lesson planning competence and highlight significant differences in their lesson planning, we purposefully selected three PSTs who achieved level 1 and three TSTs who achieved level 2 for the abovementioned criteria. These cases can be regarded as representative examples, not holistically for all criteria on levels 1 or 2, respectively, but at least for the three selected criteria.

### 6.7.1 | Choice of topic

PST 22 planned a lesson about genetic fingerprints. Students are intended to describe how criminologists use DNA to determine a connection between biological evidence (e.g., hair, skin cells) and a suspect. To justify the relevance of this topic, PST 22 refers to its relevance in science, in society, and for students:

Over the last decades, molecular genetics has become an important scientific discipline. Researchers have developed methods in many fields, such as pharmacy, criminology, and genealogy. [...] Particularly, genetic fingerprints have raised the interest of many people beyond scientists, for example, regarding paternity tests and the search for missing relatives. [...] Such applications affect students' lives and the media they consume. Therefore, this topic is relevant for students. (pp. 9–11)

Even though PST 22 attempts to use a topic relevant to the real world and connect to students' interests, they seem to use these elaborations primarily to justify the topic rather than as a starting point for the instruction. In the planned lesson, PST 22 solely focuses on the relevance of genetic fingerprints in science without explicitly adapting to his/her students' funds of knowledge and the topic's relevance for his/her particular students.

In contrast, TST 43 planned a lesson about the neuronal and hormonal regulation of stress. The students are intended to create a diagram illustrating how the human body responds to stressful situations. To justify the relevance of this topic, TST 43 elaborates on their students' current situation at school in general and during the COVID-19 pandemic in particular:

School is not always fun for students. Particularly during the ongoing pandemic, students face extraordinary challenges. The school was a place to learn *and* interact socially with peers, make friends, and so on. [...] Many leisure activities (meeting in public spaces, sports, etc.) are still impossible apart from school. However, students still have to take exams and are still required to perform well, which I assume to be even more stressful than in regular times. Due to that, it is essential for me to deal with stress from a biological perspective and have my students reflect and discuss how they could cope with stress in their individual lives. (p. 6)

**TABLE 7** Descriptive statistics for both the RALP scoring and the holistic quality assessment.

	n	RALP scoring				Holistic quality assessment				
		Mean	SD	Min	Max	Very good	Good	Satisfactory	Sufficient	Poor
PSTs	36	20.5	5.0	11	30	10	15	7	3	1
TSTs	64	27.6	5.5	15	39	13	14	17	15	5

Note: The range of possible RALP scores is 0–48.

Abbreviations: Max, maximum value; Min, minimum value; PSTs, preservice science teachers; RALP, rubric to assess science lesson plans; SD, standard deviation; TSTs, trainee science teachers.





Based on this elaboration, TST 43 plans to start the lesson by pretending to take a test on the contents of the previous lesson that took place virtually. TST 43 plans to stop this simulated test situation after 3 min and ask students how they felt during this unannounced test. Based on their initial thoughts and after reading info texts on the physiological regulation of stress, students are intended to collect stressful situations in their personal lives, classify them, and develop ideas for reducing stress.

In summary, these two examples illustrate the typical differences in how prospective science teachers try to bridge the gap between the subject matter that needs to be taught on the one hand and their students' experiences on the other hand: While PST 22 justifies the choice of the topic with general arguments (i.e., the curriculum, relevance in science, society), TST 43 explicitly tries to connect to their students and attempts to make biology relevant to them. Most importantly, unlike PST 22, TST 43 explicitly connects to their students' experiences and tailors the lesson to their needs.

## 6.7.2 | Progression of competence development

PST 34 planned a lesson about the human circulatory system. In their analysis of their students' current level of competence, PST 34 claims:

All students can describe the structure and function of the heart as well as the difference between veins and arteries. [...] They can explain the connection between pulse and respiration. [...] Only some students can describe the circulatory system as a whole. [...] (p. 3)

Even though PST 34 explicitly states that all students already can describe and explain the structure and function of the circulatory system, the students are intended to read a text and solve tasks concerning the heart, veins, and arteries, which can be regarded as a repetitive exercise rather than a progression. Beyond that, the text contains information about systemic and pulmonary regulation interplay, which is new knowledge for most students. In the final task, students are provided a schematic illustration of the human body showing the heart, brain, arteries, and veins colored red and blue, respectively. Students are intended to assign given technical terms (e.g., aorta, pulmonary artery) to the illustration. It does not become clear how far this exceeds the students' current level of competence. By now, only some students seem able to describe the circulatory system as a whole. Learning that would require a more holistic approach, for example, tracing a blood cell from the heart through the body. However, this is not part of the planned lesson.

In contrast, TST 2 planned a lesson about sustainability by evaluating a set of options for action to reduce one's ecological footprint. The criteria for such an evaluation (e.g., economic costs, health, social justice) were developed in the previous lesson. To justify the need for this lesson, TST 2 analyzes their students' current level of competence:

The students have already practiced reflecting on and evaluating phenomena in biological/ecological contexts. [They] have applied this procedure to a discussion in society and politics about reducing carbon dioxide emissions. [They] have yet to reflect and analyze their behavior and impact on the environment. (p.3)

Based on this analysis, TST 2 has students calculate their ecological footprint and develop ideas for reducing their direct or indirect carbon dioxide emissions. Finally, students are intended to apply the criteria to discuss and evaluate the suitability of their ideas. Thus, they must reflect on their behavior critically, weigh the pros and cons of their personal lives and discuss concrete options for deciding to reduce their ecological footprint. This is a

progression since TST 2 stated that, so far, their students have reflected on the consequences of political or economic decisions on ecology, which is relatively abstract.

In summary, these two examples illustrate the typical differences in how prospective science teachers adapt to their students' current level of competencies: While PST 14 provides a mixture of repetitive and new elements without explicitly adapting to their students' current needs, TST 2 explicitly builds upon the criteria that were developed in the previous lessons and transferred the procedure of an ethical evaluation of political or economic behavior to each student's behavior. In doing so, TST 2 explicitly adapts to their students' current needs and helps them develop their competencies.

### 6.7.3 | Adaptive teaching

PST 9 planned a lesson about Mendel's Laws of Heredity by crossing a black and a brown dog, resulting in a uniform generation of dogs with black and brown fur. Most students can "develop crossing schemes for Mendel's Laws [...] and describe them with correct technical terms." However, PST 9 knows some students might even struggle with it. To help them out, they prepared a worksheet already containing the correct crossing scheme so that students "have more time to reread the text and deal with the hypotheses [that were collected at the beginning of the lesson]." In doing so, PST 9 provides additional help for low-performing or relatively slowly-working students. However, it might need to be reflected in how far this approach might go too far because providing solutions is more than helping students find their way to a solution.

In contrast, TST 30 planned a lesson on the symbiotic relationship between mycorrhizal fungi and plant roots by creating an infographic using the example of truffles. In their analysis of the students' current needs, TST 30 claims that the students form a highly heterogeneous group concerning their ability to illustrate textual or numerical information: While "some students still struggle to read tables and decide which type of diagram would be most suitable to visualize the data," the top-performing students can read and understand complex texts and discuss the pros and cons of various options for visualization. To adapt to these heterogeneous needs in the class, TST 30 uses two instructional approaches: First, low-performing students are provided aid cards to have students think about how to illustrate the information adequately. For instance, one of the pieces of advice is: "Recall our recent discussion on the exam question: Wouldn't it be necessary to use two y-axes in the diagram? Define which data are dependent and which are independent variables and assign them to the x- or y-axis." Second, the high-performing students are provided an additional task because they usually work more quickly than the rest of the class. They are requested to "include and explicitly name biotic (e.g., boars, pathogens) and abiotic factors (e.g., soil, the temperature during winter) that influence the growth of truffles" and thus illustrate the textual and numerical information rather than in the form of a concept map.

In summary, these two examples illustrate typical differences in how prospective science teachers adapt to their students' current level of competencies: While PST 9 at least offers additional help for low-performing students, TST 30 plans to provide tasks and help for both the low- and the high-performing students.

## 7 | DISCUSSION

There is a lack of high-quality instruments to assess science teachers' lesson planning competence, in general, and their ability to write proper lesson plans, in particular (Krepf & König, 2022). This study aimed to build an argument for the validity of the interpretations of the RALP scores by considering multiple sources of validity evidence (AERA, APA, & NCME, 2014; Brookhart, 2018; Jonsson & Svingby, 2007) and applying this novel instrument to a sample of  $N = 100$  prospective science teachers to gain further insights into the development of lesson planning competence. We regard the employed explanatory sequential mixed-methods design (QUAN → qual; Creswell & Plano



Clark, 2018) as one of the strengths of this study: On the one hand, the relatively large sample size allowed for a quantitative analysis that yielded significant findings concerning the validity of the RALP score interpretation. However, the quantitative analysis alone would not provide further insights into prospective science teachers' lesson planning. On the other hand, the qualitative analysis of three pairs of lesson plans showed challenges that PSTs and TSTs meet in the process of lesson planning on different levels of quality. However, the qualitative analysis alone would barely suffice to claim that the findings are generalizable, at least to a certain degree. The combination of the quantitative and the qualitative approach made it possible to apply the RALP validly and to purposefully select the three most fruitful criteria for comparing PSTs' and TSTs' lesson plans.

To the best of our knowledge, the RALP is currently the only available broadly applicable scoring rubric in science education that can be used to assess lesson planning competence. Unlike other existing rubrics in science education (Table 1: Enugu & Hokayem, 2017; Forbes & Davis, 2010; Goldston et al., 2013), the RALP can be applied not only to inquiry lessons but to all kinds of science lessons. Moreover, in contrast to Jacobs et al.'s (2008) rubric, the RALP is theoretically framed and connects to PCK-based research designs. Thus, even though we solely analyzed biology lesson plans, the RALP might also apply to chemistry or physics lesson plans. In the following section, we will first reflect on a couple of relevant limitations in this study and then discuss key characteristics of the RALP with regard to the abovementioned validity argument (Table 2). For the sake of comprehensibility, these characteristics will not be discussed in chronological order but sorted by the three vertices of Pellegrino et al.'s (2016) *Assessment Triangle* (Figure 2).

## 7.1 | Limitations

Overall, the RALP's application to  $N = 100$  lesson plans limited the possibility of statistical analyses. It was not manageable to obtain a more considerable amount of lesson plans written by PSTs from the university as, by now, the criteria are used in the authors' courses so that PSTs are trained to consider them. Consequently, results would be skewed if we had analyzed a sample trained to apply the RALP. Hence we analyzed lesson plans from 2019 to 2021, which were written under identical circumstances by PSTs who did not know the RALP yet. Most importantly, the present study is essentially qualitative and thus seems to have a large sample size compared to other qualitative studies analyzing written lesson plans (e.g.,  $N = 10$ , Chizhik & Chizhik, 2018;  $N = 22$ , Kademian & Davis, 2018;  $N = 10$ , Morine-Dersheimer, 1979;  $N = 18$ , Zaragoza et al., 2021).

Moreover, the analysis of these lesson plans also implicates several methodological challenges. For instance, we did not influence the criteria and formal regulations the TSTs considered when writing the lesson plans. Consequently, the word counts of TSTs' lesson plans ranged between approximately 1800 and 7800 (Großmann & Krüger, 2022c), affecting the probability that all 24 RALP criteria were equally considered. A more controlled setting where PSTs and TSTs write lesson plans might seem more appropriate. Thus, we would ensure we capture the prospective teachers' ePCK<sub>p</sub> or pPCK, respectively. Since the lesson plans analyzed in this study were part of formal examinations, PSTs and TSTs might have received feedback from colleagues or friends, leading us to capture collective PCK instead (Alonzo et al., 2019) rather than exclusively pPCK. However, writing a lesson plan is time-consuming and challenging to integrate into a typical university course. As described above, Kang (2017) suggested that researchers in the field of lesson planning should not rely on controlled research designs but consider teachers planning processes or products captured in natural environments. To fill this gap, we decided to focus on original lesson plans. We regard the ecological validity (Wegener & Blankenship, 2007) of our analysis as a crucial feature in the development and evaluation process of the RALP. Nevertheless, it might have been helpful to collect demographic information about the participants or information about the schools they were teaching at to get a complete impression of the circumstances under which the participants were planning and teaching.

Recently, there has been a shift from developing scoring rubrics for grading purposes (summative assessment) to supporting the target groups in their learning processes (formative assessment; Panadero & Jonsson, 2013;

Panadero et al., 2018). Thus, this study might be regarded as a shift back to a past time since we focused on the summative perspective. However, this study can also be regarded as the foundation for the further development of the RALP: As a first step, it needed to be investigated whether the RALP covers relevant aspects of lesson planning (RQ 1.1), whether it can distinguish quality levels (RQ 1.5), and whether it is instructional sensitive (RQ 1.6). Based upon that, further evidence should be collected, particularly with regard to formative and self-assessment (e.g., *validity evidence based on response processes*; AERA, APA, & NCME, 2014). As Brookhart (2018) suggests, future studies should consider how the target group uses a scoring rubric as a learning tool.

## 7.2 | Evaluation of the plausibility of the validity argument

To ensure that the scores of the RALP can be interpreted validly, empirical evidence is needed (AERA, APA, & NCME, 2014; Kane, 2013; Pellegrino et al., 2016), which is often not considered in research on scoring rubrics (Reddy & Andrade, 2010) and with regard to scoring rubrics on lesson planning (Table 1). As the underlying validity argument, we defined that the RALP scores are interpreted as indicators of lesson plan quality. Overall, the RALP covers essential aspects of lesson planning and is, first and foremost, able to discriminate between high- and low-quality lesson plans. Therefore, this section will take step III of building a validity argument (Figure 2) and discuss the empirical findings for all three vertices of the *Assessment Triangle* (Table 2; Pellegrino et al., 2016).

### 7.2.1 | Evidence for cognitive validity

Proceeding from the assumption that lesson planning requires professional knowledge in general (Zaragoza et al., 2021) and PCK in particular (Alonzo et al., 2019; Carlson et al., 2019; Großmann & Krüger, 2022c; Vogelsang et al., 2022), it appears reasonable to operationalize lesson planning competence in terms of PCK. Figure 1 shows that the 24 cognitive demands during lesson planning built upon König et al.'s (2021) CODE-PLAN model can be assigned to four of the five PCK components and their interconnections (Park & Oliver, 2008). In addition, as described by Großmann and Krüger (2022a), all criteria and performance levels build upon theoretical and empirical research. Thus, it can be assumed that the RALP sufficiently relates to the cognitive construct (*Validity evidence based on test content*; AERA, APA, & NCME, 2014). This was also confirmed by the expert discussions (RQ 1.1). Based on their feedback, we rechecked and improved the assignment of criteria to the pentagon model, the order of the criteria within the RALP, and the specificity and appropriateness of the performance level descriptions.

### 7.2.2 | Evidence for instructional validity

Among the 29 initial criteria rated by the teacher educators (RQ 1.1), only two were considered somewhat irrelevant: the foundation of justifications with theoretical and empirical literature and considering students' conceptions (Table 4). We accepted the suggestion to include justifications in the performance level descriptions of selected RALP criteria and not provide an individual criterion. Regarding students' conceptions, some teacher educators argued that it is unnecessary to consider them in each biology lesson, so they rated their relevance rather poorly. The neglect of students' conceptions in trainee biology teachers' lesson plans was recently described by Großmann and Krüger (2022b). However, students' conceptions play a significant role in science education (e.g., Lucero et al., 2017; Otero & Nathan, 2008) since they would allow educators to adapt their instruction to students' thinking in terms of the constructivist learning theory (Duit & Treagust, 2012). Due to that, and since the criteria are also meant to trigger prospective teachers to consider aspects they otherwise might miss during lesson planning, we decided to keep students' conceptions as a criterion in the RALP. Two teacher educators suggested selecting relevant technical terms as



another criterion. This is also an issue in science education research, showing that in-service biology teachers' extensive use of technical terms affects students' achievements and situational interests (Dorfner et al., 2020). We concluded that this aspect is relevant for prospective teachers and thus added it to the RALP.

Regarding the analysis of validity evidence based on the internal structure (RQ 1.3, RQ 1.4), the findings align with our assumptions. As expected, due to different requirements (Table 3), PSTs barely contextualized their lesson into a unit (O1, O2), whereas TSTs barely considered students' conceptions (S3) in their lesson plans (Table 6). More specifically, the intended learning outcomes often do not mediate prospective teachers' thinking about their students' needs (S1). Particularly, assessment at the end of a lesson is often not fully aligned with the intended learning outcome (A2), a frequently reported issue in lesson planning (e.g., Chizhik & Chizhik, 2018; Weitzel & Blank, 2020). Hence, the findings are also in line with other studies. Moreover, prospective teachers struggled to describe and justify the tasks they planned to give (I4). Among the many aspects to consider, planning tasks represent one of the essential cognitive demands (König et al., 2021). Tasks should be aligned with the intended learning outcome and match the students' current abilities to facilitate a meaningful learning process. Developing high-quality tasks is challenging for prospective teachers, mainly because it is also related to the phasing of a lesson (Kang et al., 2016). Thus, it is one of the critical decisions in the lesson planning process.

The psychometric analysis provides evidence for a one-dimensional model (Table 5). Thus the competence to write good lesson plans should not be separated into two or more latent variables. We interpret this as an indicator that the RALP criteria reflect the interconnectedness of PCK components (Figure 1) rather than distinct aspects. In contrast, König et al. (2021) suggested that lesson planning competence consists of six skills (content transformation, task creation, adaptation to student learning dispositions, clarity of learning objectives, unit contextualization, and phasing) that are needed to master six significant cognitive demands in lesson planning. Their confirmatory factor analysis showed a good model fit for a six-dimensional model and a poor model fit for a one-dimensional model. Thus, we would not argue that lesson planning competence generally cannot be specified in terms of latent variables. We only conclude that for our sample of  $N = 100$  prospective science teachers, the RALP reflects the intended unidimensionality instead of a reasonably assumed multidimensionality. Possibly, the ability to meet König et al.'s (2021) six cognitive demands might indeed form separable components of lesson planning competence. In contrast, the related PCK components investigated in this study develop their potential if they are interconnected, thus leading to a one-dimensional model. Based on the assumption that the underlying construct is unidimensional, Cronbach's  $\alpha$  (0.78) was calculated as a measure of internal consistency and is sufficiently high in this sample. Still, it should be noted that Cronbach's  $\alpha$  is affected by the number of criteria (Taber, 2018), so this value might result from many criteria ( $n = 24$ ) sharing some variance with others. In addition, the Wright map (Figure 4) suggests that prospective science teachers' ability to address the RALP criteria is normally distributed; most criteria match their ability. Hence the performance expectations were neither too demanding nor under-demanding for our sample. While the TSTs barely considered students' conceptions, and thus this criterion formed an outlier in the Wright map, the alignment of teachers' horizon of expectations to their students' current levels of competence (I6) might be considered a superfluous criterion.

However, since this analytic rubric is also intended to make expectations of good lesson plans transparent and to be used for formative feedback, we would keep these criteria. Most criteria exceed the threshold of 0.30 (Vaus, 2014) and can thus be regarded as sufficiently discriminatory. The remaining four criteria (O4, S3, S5, I4; Supporting Information: Appendix A) barely help to discriminate PSTs' and TSTs' performances (Table 6). Suppose the RALP is intended for summative assessment purposes. In that case, some highly discriminatory criteria might receive a higher weight in the scoring process, indicating that these criteria contribute more to a good lesson plan than others.

### 7.2.3 | Evidence for inferential validity

The RALP allows a stable (RQ 1.2a) and intersubjectively comprehensible (RQ 1.2b) scoring procedure. As stated above, we intend the RALP to be used by teacher educators for grading purposes and by TSTs as a learning and

self-assessment tool (Brookhart, 2018; Krebs et al., 2022; Panadero et al., 2018). For that purpose, we suggest providing at least one coded lesson plan in addition to the scoring rubric so that the target group gets an idea of how the rubric can be applied and which parts of a typical lesson they mainly need to focus on. Additional explanations and definitions might also be necessary to clarify what a good intended learning outcome (O3) is or when an activity can be regarded as appropriate (I3), which were two criteria with problematic  $\kappa$  values. Moreover, even though it is relatively uncommon to investigate intrarater agreement in research on scoring rubrics (Brookhart, 2018; Jonsson & Svingby, 2007), we argue that the high  $\kappa$  values are an essential prerequisite for the use of the RALP and should therefore be calculated when developing a scoring rubric.

Most importantly, we found evidence for the “predictive validity” of the RALP: The excellent correlations between the RALP scores and the teacher educator's holistic quality assessment and the large effect sizes between quality levels (RQ 1.5; *Validity evidence based on relations to other variables*; AERA, APA, & NCME, 2014) indicate that the RALP scores do indeed reflect lesson plan quality. The RALP can discriminate quality levels for the whole subsamples and between the subgroups of holistic quality assessment (Figure 5). In line with Tomas et al. (2019), we argue that analytic and holistic approaches should not be regarded as mutually exclusive but as complementary possibilities for substantiating interpretations made from analytic or holistic rubrics. However, if the RALP is intended for grading, it should be considered to add weights to the criteria, as executed by Jacobs et al. (2008). All 24 criteria contribute equally to the final score, which might skew the results. Probably, criteria referring to the intended learning outcome as the most crucial reference point for making planning decisions during lesson planning (Drost & Levine, 2015) might be more important than a justification of the choice of topic, so this might be taken into account in the scoring process by adding weights. Another option would be to use the relevance assessment (Table 4) to weigh the most relevant criteria (e.g., >2.70) three times, the criteria of medium relevance (e.g., 2.30–2.69) two times, and the minor relevant criteria (e.g., <2.30) one time. Hence, the scores 0, 1, and 2 for each criterion must be multiplied by 1, 2, or 3, respectively. Thus, one could ensure that the criteria found to be most relevant (e.g., formulating a precise intended learning outcome) contribute more to the overall score than criteria of minor relevance (e.g., considering students' conceptions). Another option would be that teacher educators who wish to use the RALP in their courses determine weights based on their individual preferences.

Moreover, the RALP consists of only three performance levels. Most published analytic scoring rubrics have four or five levels (Brookhart, 2018). In developing the RALP, we had to weigh the pros and cons of increasing or decreasing the number of performance levels. On the one hand, more performance levels would increase the instrument's sensitivity as it would detect quality differences on a more fine-grained level. While the holistic quality assessment is relatively fine-grained (1.0, 1.3, ..., 4.7, 5.0), the three performance levels only allow a coarse-grained analysis. On the other hand, the performance levels should be selective to ensure interrater reliability. Developing more than three distinct performance level descriptions was difficult for many criteria. Consequently, we accepted a lower sensitivity in favor of clarity and unambiguity and decided to provide three performance levels for each criterion.

Apart from the development process, even the top-rated TSTs who achieved the highest RALP scores could not even come close to the maximum score of 48, meaning that lesson plans can be of high quality even without considering every criterion. This might be another reasonable argument to add weight to the criteria and thus indicate that specific criteria contribute more to the overall lesson plan quality than others. However, we would not delete any criteria from the RALP since they were considered relevant by teacher educators (RQ 1.1). Moreover, neither the PSTs nor the TSTs in this sample explicitly practiced applying the 24 criteria because they had not existed yet when they wrote their lesson plans. Teacher educators that use the RALP in their courses should introduce the 24 criteria to ensure transparency. We expect that under such circumstances, the maximum score of 48 will be achievable by high-performing students.

Moreover, by our expectations, the more experienced TSTs scored significantly higher than the PSTs (RQ 1.6; Table 7; Figure 5). Even though we did not conduct a longitudinal but a cross-sectional study, our findings connect to previous research on the development of lesson planning competence during teacher training (Backfisch



et al., 2020; König et al., 2021, 2022; Mutton et al., 2011; Vogelsang et al., 2022; Westerman, 1991). TSTs gained much more field experience and professional knowledge during their induction phase than PSTs during their studies (Figure 3). Possibly, this facilitates the integration of different PK, CK, and PCK aspects during lesson planning (Stender et al., 2017; Zaragoza et al., 2021) and results in higher scores on the RALP. Another fruitful idea for teacher educators might be to distinguish *lesson plan analysis* and *lesson planning* and scaffold the application of professional knowledge to both practices in a long-term learning process (Zaragoza et al., 2023): Before PSTs engage in planning their lessons, teacher educators might provide a scaffold and have PSTs analyze simulated or authentic lesson plans. Thus, they would gradually be introduced to the complex nature of lesson planning. For the first step in Zaragoza et al.'s (2023) suggestion—the lesson plan analysis—the 24 RALP criteria might be used. After PSTs have learned to evaluate lesson plans, they start planning their own lessons and might use the RALP criteria as guidelines.

Finally, it should be noted that PSTs' and TSTs' lesson plans were written under different conditions: While it is only the third lesson plan that PSTs usually write in their teacher training, which is only one of many examinations that PSTs have to complete during Master's studies, TSTs have gained far more practice during their induction phase and can focus on the final examination which is of unique importance (Figure 3). Hence, as König et al. (2021) argue, expectations regarding lesson plan quality are much higher for TSTs' final examination. Therefore they might put more effort into it than PSTs, which might explain the large effect.

### 7.3 | Differences between PSTs' and TSTs' lesson plans

Our study adds to the research attempts in the field of lesson planning in science education mentioned above (Table 1) by applying the RALP to original lesson plans and providing qualitative insights into how prospective science teachers plan lessons and justify their decisions. The sample was large enough to claim that the following observations can be generalizable, at least for prospective biology teachers in Berlin. While novices need help to interconnect their professional knowledge in the process of lesson planning, expert teachers use their knowledge to explicitly adapt to their student's needs (Westerman, 1991). Our analysis sheds light on the differences between PSTs and TSTs.

In general, the more experienced TSTs focus clearly on their particular students and tailor the planned instruction and their justifications to their student's needs. The comparison between PST 22 and TST 43 illustrates a typical pattern found in many lesson plans: While TSTs use the analysis of the biological subject matter and their students' needs to plan a lesson, PSTs often elaborate in detail on rather irrelevant information. Since interconnecting these different knowledge bases with each other is challenging, particularly for novices (Westerman, 1991; Zaragoza et al., 2021), it seems to be necessary for teacher educators to provide support for lesson planning (Karlström & Hamza, 2021) and highlight for what purpose a lesson plan needs to be written. This might help PSTs to focus on essential aspects of teaching. Regarding the choice of topic, PST 22 justified the relevance with rather broad arguments, referring to society, for example. As König et al. (2020) point out, preservice teachers frequently use generic statements in lesson planning. Instead, it might be necessary to focus on subject-specific arguments to highlight how particular planning decisions help students learn science.

Regarding instructional planning, the significant difference between PSTs and TSTs in this sample is the degree to which the prospective science teachers adapt to their students' needs and plan lessons explicitly tailored to them (Westerman, 1991). Unlike experienced teachers, PSTs still lack mental plans (Borko & Livingston, 1989), which might explain their difficulties in building a lesson that connects to their students' current level of competencies. In their lesson plan, TST 2 connected to the previous lesson and made the transitions between the phases within the lesson transparent. The lesson is contextualized in a unit, resulting in a long-term learning process that entails a stepwise and progressive sequence of lessons to help students develop their competencies. The PSTs in this sample were not required to plan a whole unit (Table 3). Since unit contextualization is one of the significant cognitive

demands in lesson planning (König et al., 2021), teacher educators might reflect on whether teaching how to plan long-term learning processes needs to be considered early on in teacher education.

Regarding the adaptation to the students' needs, the RALP might help PSTs to broaden their perspective on their students: It is worth emphasizing that even the PSTs often consider how to help the low-performing students achieve the intended learning outcome, which can be regarded as an encouraging result. TST 30 adapts to the high-performing students' abilities, indicating a differentiated view of the class and the capacity to use instructional strategies to consider individual differences. There is no convincing reason why PSTs could not use such strategies for high-performing students. In that sense, the RALP might encourage PSTs to take low- and high-performing students into account, and thus, prospective science teachers might use it as a learning tool.

## 8 | CONCLUSION AND OUTLOOK

This study pursued two objectives: First, we aimed to evaluate the quality of the RALP considering multiple sources of validity evidence (AERA, APA, & NCME, 2014). Second, we aimed to apply the RALP to an ecologically valid sample of prospective science teachers to gain qualitative insights into their strengths and weaknesses. As suggested by Scherer (2017), referring to Pellegrino et al.'s (2016) *Assessment Triangle* (Figure 2) facilitates building a sound and persuasive validity argument since it automatically requires considering different participants (e.g., science teacher educators, PSTs, TSTs). Overall, various pieces of empirical evidence for validity were collected, and thus, a persuasive validity argument was built. In particular, the RALP covers relevant aspects of lesson planning and can discriminate lesson plan quality and can thus be regarded to be a valuable instrument for (1) teacher educators and (2) researchers: (1) Teacher educators can use the RALP for both grading and feedback purposes in teacher training. In doing so, teacher educators could make their expectations transparent and refer to an objective set of criteria instead of their own, maybe partially subjective criteria. Since the RALP is an analytic scoring rubric, feedback would be highly differentiated and help prospective science teachers to grasp the next step in their learning process. Therefore, we recommend not displaying the overall score but only focusing on single criteria (Panadero & Jonsson, 2020). (2) Researchers in science education might use the RALP in study designs that include the analysis of lesson plans, for example, in lesson studies (e.g., Juhler, 2018). Since the RALP can be used for both qualitative analyses based on the performance level descriptions and for quantitative analyses based on the overall score, we argue that this instrument might contribute to research in science education in various ways since it allows valid interpretations. The qualitative analysis of selected typical cases has shed light on significant differences between PSTs and TSTs, particularly regarding their ability to adapt to their particular students' needs. Presenting some of the RALP criteria to PSTs might help them consider aspects they might have missed otherwise. Moreover, the criteria might facilitate focusing on relevant aspects for instruction instead of detailed elaborations on aspects of minor relevance for the class. However, it should be noted that the RALP was developed in the context of German teacher education. As mentioned above, lesson plans consist of approximately 10 pages containing analyses, descriptions, and justifications. This concept of lesson plans differs from those common in other countries, where lesson plans consist of three pages at most. Consequently, if teacher educators or researchers intend to apply the RALP to their contexts, we recommend ensuring that the RALP criteria meet the local requirements. If PSTs are not required to justify their planning decisions, they cannot reach a high score. However, this would result from an inappropriate assessment procedure rather than the preservice teachers' failure. We suggest adjusting the lesson plan requirements to apply the RALP appropriately in these cases. The other option would be to modify the RALP according to the locally different lesson plan requirements. However, this might have undesired effects on certain validity aspects. For instance, deleting items might affect *validity evidence based on test content*, as critical, theoretically reasoned, and empirically tested criteria would be abandoned.





Since “the validation process never ends, as there is always additional information that can be gathered to more fully understand a test and the inferences that can be drawn from it” (AERA, APA, & NCME, 2014, pp. 21–22), we suggest investigating further aspects of validity that are of significant concern for the quality of the RALP. First, the assumption that the RALP applies to all sciences must be investigated. Moreover, bearing in mind that the RALP is not only intended to be used as a grading tool for summative assessment purposes but also for formative assessment (Panadero & Jonsson, 2020) and self-assessment purposes (Krebs et al., 2022), the following research desiderata (i.e., open questions for further research) particularly relate to the interpretation vertex of the *Assessment Triangle* (Figure 2; Pellegrino et al., 2016): If measuring prospective teachers' PCK (e.g., using the PCK-in biology inventory; Großschedl et al., 2019), is there evidence of convergent validity, indicating that the RALP does indeed measure PCK (*Validity evidence based on relations to other variables*)? How do prospective teachers engage with the RALP when applying it to their lesson plans? Can they improve a lesson plan after self-assessment (Krebs et al., 2022; Scherer, 2017; *Validity evidence based on response processes*)? If the RALP is applied by a teacher educator, by a peer student (peer assessment), or by students themselves (self-assessment) and the authors of the lesson plans are requested to improve their lesson plan afterward (Ozogul et al., 2008): Which of these three types of application helps prospective teachers to improve their lesson plans most? In an experimental-control-group-design, do prospective teachers who were introduced to the RALP write better lesson plans in terms of holistic quality assessment than those who do not know the RALP criteria beforehand (*Validity evidence based on the consequences of testing*)? Finally, we recommend shedding light on the underexplored relationship between lesson planning and instruction. In what way do the RALP scores predict the quality of classroom teaching?

## ACKNOWLEDGMENTS

The project K2Teach (Know how to teach) is part of the “Qualitätsoffensive Lehrerbildung,” a joint initiative of the German Federal Government and the Länder, which aims to improve the quality of teacher training. The programme is funded by the Federal Ministry of Education and Research (Grant Number 01JA1802). The authors are responsible for the content of this publication. Furthermore, the authors would like to thank the Berlin Senate Department for Education, Youth, and Family for providing lesson plans in an anonymized form. Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The lesson plans were anonymized, digitized, and analyzed exclusively to answer the scientific research questions. The trainee science teachers' lesson plans were provided, thankfully, by the Berlin Senate Department for Education, Youth, and Family under the condition that they are not published or disseminated. The copyright remains with the biology teachers who have written the lesson plans.

## ORCID

Leroy Großmann  <http://orcid.org/0000-0001-7635-1737>

Dirk Krüger  <http://orcid.org/0000-0003-0999-4382>

## ENDNOTES

- <sup>1</sup> It has not escaped our notice that there are persistent concerns in PCK research on the varying conceptualization of PCK, on measuring PCK, and on the comparability of research findings that might be affected by those different understandings and measures (e.g., Chan & Hume, 2019; Settlage, 2013). Consequently, caution is advised when correlating PCK to other variables, such as cognitive activation. Nevertheless, the relevance of PCK for the quality of teachers' practice is beyond doubt.

<sup>2</sup> Note that Kane (2013, p. 10) differentiated between the *interpretation/use argument*, “that specifies what is being claimed in the interpretation and use,” and the *validity argument*, “[...] which provides an evaluation of the proposed [interpretation/use argument].” Agreeing with Newton’s (2013) concluding remarks at the end of his critical analysis of the expediency of Kane’s (2013) terminological differentiation, we assume there is only one argument, the *validity argument*, which is the term we will adhere to in this paper.

## REFERENCES

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1), 3–14. [https://doi.org/10.1016/0304-4076\(81\)90071-3](https://doi.org/10.1016/0304-4076(81)90071-3)
- Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 271–286). Springer Singapore. [https://doi.org/10.1007/978-981-13-5898-2\\_12](https://doi.org/10.1007/978-981-13-5898-2_12)
- Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27–31.
- Australian Institute for Teaching and School Leadership. (2018). *Australian professional standards for teachers*. <https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standards-for-teachers.pdf>
- Backfisch, I., Lachner, A., Hische, C., Loose, F., & Scheiter, K. (2020). Professional knowledge or motivation? Investigating the role of teachers' expertise on the quality of technology-enhanced lesson plans. *Learning and Instruction*, 66, 101300. <https://doi.org/10.1016/j.learninstruc.2019.101300>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Wright maps: First steps. In W. J. Boone, J. R. Staver, & M. S. Yale (Eds.), *Rasch analysis in the human sciences* (pp. 111–136). Springer. [https://doi.org/10.1007/978-94-007-6857-4\\_6](https://doi.org/10.1007/978-94-007-6857-4_6)
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473–498.
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3(22), 1–12. <https://doi.org/10.3389/feduc.2018.00022>
- Campbell, T., Gray, R., Fazio, X., & van Driel, J. (2022). Research on secondary science teacher preparation. In J. A. Luft & M. G. Jones (Eds.), *Handbook of research on science teacher education* (pp. 97–118). Routledge. <https://doi.org/10.4324/9781003098478-10>
- Carlson, J., Daehler, K. R., Alonzo, A., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Chan, K. K. H., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., ... Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper & A. Borowski, (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–92). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_2](https://doi.org/10.1007/978-981-13-5898-2_2)
- Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer Singapore. [https://doi.org/10.1007/978-981-13-5898-2\\_1](https://doi.org/10.1007/978-981-13-5898-2_1)
- Chan, K. K. H., Rollnick, M., & Gess-Newsome, J. (2019). A grand rubric for measuring science teachers' pedagogical content knowledge. In A. Hume, R. Cooper & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 253–271). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_11](https://doi.org/10.1007/978-981-13-5898-2_11)
- Chizhik, E. W., & Chizhik, A. W. (2018). Using activity theory to examine how teachers' lesson plans meet students' learning needs. *The Teacher Educator*, 53(1), 67–85. <https://doi.org/10.1080/08878730.2017.1296913>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research (international student edition)* (3rd ed.). Sage.
- Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, 25(3), 179–204. <https://doi.org/10.1007/s11092-013-9163-0>
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>
- Dorfner, T., Förtsch, C., & Neuhaus, B. J. (2020). Use of technical terms in German biology lessons and its effects on students' conceptual learning. *Research in Science & Technological Education*, 38(2), 227–251. <https://doi.org/10.1080/02635143.2019.1609436>



- Drost, B. R., & Levine, A. C. (2015). An analysis of strategies for teaching standards-based lesson plan alignment to preservice teachers. *Journal of Education*, 195(2), 37–47.
- Duit, R., & Treagust, D. F. (2012). How can conceptual change contribute to theory and practice in science. In B. J. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 107–118). Springer.
- Enugu, R., & Hokayem, H. (2017). Challenges pre-service teachers face when implementing a 5E inquiry model of instruction. *European Journal of Science and Mathematics Education*, 5(2), 178–209.
- Forbes, C. T., & Davis, E. A. (2010). Curriculum design for inquiry: Preservice elementary teachers' mobilization and adaptation of science curriculum materials. *Journal of Research in Science Teaching*, 47(7), 820–839. <https://doi.org/10.1002/tea.20379>
- Förtsch, C., Werner, S., von Kotzebue, L., & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. <https://doi.org/10.1080/09500693.2016.1257170>
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, P. M. Friedrichsen & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Goldston, M. J., Dantzer, J., Day, J., & Webb, B. (2013). A psychometric approach to the development of a 5E lesson plan scoring instrument for inquiry-based teaching. *Journal of Science Teacher Education*, 24(3), 527–551. <https://doi.org/10.1007/s10972-012-9327-7>
- Großmann, L., & Krüger, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen [Development and Application of a category system for the analysis of pre-service biology teachers' pedagogical content knowledge in written lesson plans]. *Erkenntnisweg Biologiedidaktik*, 19, 21–39.
- Großmann, L., & Krüger, D. (2022a). Biologieunterricht erfolgreich planen—ein Kriterienraster zum Schreiben von Unterrichtsentwürfen [Planning biology lessons successfully—a scoring rubric for writing lesson plans]. *SEMINAR—Lehrerbildung und Schule*, 1, 91–110.
- Großmann, L., & Krüger, D. (2022b). Students' conceptions as a neglected perspective in trainee teachers' biology lesson plans. In K. Korfiatis & M. Grace (Eds.), *Contributions from biology education research. Current research in biology education* (pp. 181–193). Springer International Publishing. [https://doi.org/10.1007/978-3-030-89480-1\\_14](https://doi.org/10.1007/978-3-030-89480-1_14)
- Großmann, L., & Krüger, D. (2022c). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? [What's the role of trainee biology teachers' pedagogical content knowledge for the quality of their written lesson plans] *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1), 1–20. <https://doi.org/10.1007/s40573-022-00141-w>
- Großmann, L., & Krüger, D. (2023). Identifying performance levels of enacted pedagogical content knowledge in trainee biology teachers' lesson plans. In G. S. Carvalho, A. S. Afonso, & Z. Anastácio (Eds.), *Contributions from science education research. fostering scientific citizenship in an uncertain world: Selected papers from the ESERA 2021 conference* (pp. 95–116). Springer. [https://doi.org/10.1007/978-3-031-32225-9\\_7](https://doi.org/10.1007/978-3-031-32225-9_7)
- Großschedl, J., Mahler, D., Kleickmann, T., & Harms, U. (2014). Content-Related knowledge of biology teachers from secondary schools: Structure and learning opportunities. *International Journal of Science Education*, 36(14), 2335–2366. <https://doi.org/10.1080/09500693.2014.923949>
- Großschedl, J., Welter, V., & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439. <https://doi.org/10.1002/tea.21482>
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4th ed.). Advances Analytics LLC.
- Jacobs, C. L., Martin, S. N., & Otieno, T. C. (2008). A science lesson plan analysis instrument for formative and summative program evaluation of a teacher education program. *Science Education*, 92(6), 1096–1126. <https://doi.org/10.1002/sce.20277>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Juhler, M. V. (2018). Assessment of understanding: Student teachers' preparation, implementation and reflection of a lesson plan for science. *Research in Science Education*, 48(3), 515–532. <https://doi.org/10.1007/s11165-016-9574-2>
- Kademian, S. M., & Davis, E. A. (2018). Supporting beginning teacher planning of investigation-based science discussions. *Journal of Science Teacher Education*, 29(8), 712–740. <https://doi.org/10.1080/1046560X.2018.1504266>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>

- Kang, H. (2017). Preservice teachers' learning to plan intellectually challenging tasks. *Journal of Teacher Education*, 68(1), 55–68. <https://doi.org/10.1177/0022487116676313>
- Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching*, 53, 1316–1340.
- Karlström, M., & Hamza, K. (2021). How do we teach planning to pre-service teachers—A tentative model. *Journal of Science Teacher Education*, 32, 1–22. <https://doi.org/10.1080/1046560X.2021.1875163>
- Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, 5, 586–614.
- Koberstein-Schwarz, M., & Meisert, A. (2022). Pedagogical content knowledge in material-based lesson planning of preservice biology teachers. *Teaching and Teacher Education*, 116, 103745. <https://doi.org/10.1016/j.tate.2022.103745>
- König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I., & Glutsch, N. (2020). Pre-service teachers' generic and subject-specific lesson-planning skills: On learning adaptive teaching during initial teacher education. *European Journal of Teacher Education*, 43(2), 131–150. <https://doi.org/10.1080/02619768.2019.1679115>
- König, J., Cammann, F., Bremerich-Vos, A., & Buchholtz, C. (2022). Unterrichtsplanungskompetenz von (angehenden) Deutschlehrkräften der Sekundarstufe: Testkonstruktion und Validierung. *Zeitschrift Für Erziehungswissenschaft*, 25, 869–894. <https://doi.org/10.1007/s11618-022-01113-z>
- König, J., Krepf, M., Bremerich-Vos, A., & Buchholtz, C. (2021). Meeting cognitive demands of lesson planning: Introducing the CODE-PLAN model to describe and analyze teachers' planning competence. *The Teacher Educator*, 56(4), 466–487. <https://doi.org/10.1080/08878730.2021.1938324>
- Kotthoff, H. G., & Terhart, E. (2013). Teacher education in Germany: Traditional structure, strengths and weaknesses, current reforms. *Scuola Democratica*, 4(3), 1–9.
- von Kotzebue, L. (2022). Beliefs, self-reported or performance-assessed TPACK: What can predict the quality of technology-enhanced biology lesson plans? *Journal of Science Education and Technology*, 31(5), 570–582. <https://doi.org/10.1007/s10956-022-09974-z>
- Krebs, R., Rothstein, B., & Roelle, J. (2022). Rubrics enhance accuracy and reduce cognitive load in self-assessment. *Metacognition and Learning*, 17, 627–650. <https://doi.org/10.1007/s11409-022-09302-1>
- Krepf, M., & König, J. (2022). Structuring the lesson: An empirical investigation of pre-service teacher decision-making during the planning of a demonstration lesson. *Journal of Education for Teaching*, 1–16. <https://doi.org/10.1080/02607476.2022.2151877>
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lucero, M. M., Petrosino, A. J., & Delgado, C. (2017). Exploring the relationship between secondary science teachers' subject matter knowledge and knowledge of student conceptions while teaching evolution by natural selection. *Journal of Research in Science Teaching*, 54(2), 219–246. <https://doi.org/10.1002/tea.21344>
- Mahler, D., Großschedl, J., & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, 39(2), 213–237. <https://doi.org/10.1080/09500693.2016.1276641>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Ministère de l'Éducation Nationale. (2013). Arrêté du 1er juillet 2013 relatif au référentiel des compétences professionnelles des métiers du professorat et de l'éducation. <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027721614/>
- Morine-Dershimer, G. (1979). *Teacher plan and classroom reality: The South Bay Study, part IV* (Research Series No. 60). Institute for Research on Teaching College of Education Michigan State University.
- Morine-Dershimer, G. (2011). Instructional planning. In J. M. Cooper (Ed.), *Classroom teaching skills* (9th ed., pp. 45–81). Wadsworth CENGAGE Learning.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how?. *Practical Assessment, Research, and Evaluation*, 7, 3. <https://doi.org/10.7275/A5VQ-7Q66>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research, and Evaluation*, 7, 10. <https://doi.org/10.7275/q7rm-gg74>
- Mutton, T., Hagger, H., & Burn, K. (2011). Learning to plan, planning to learn: The developing expertise of beginning teachers. *Teachers and Teaching*, 17(4), 399–416. <https://doi.org/10.1080/13540602.2011.580516>



- Ndihokubwayo, K., Byukusenge, C., Byusa, E., Habiaryemye, H. T., Mboniyirivuze, A., & Mukagihana, J. (2022). Lesson plan analysis protocol (LPAP): A useful tool for researchers and educational evaluators. *Heliyon*, 8(1), e08730. <https://doi.org/10.1016/j.heliyon.2022.e08730>
- Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki, & T. Hirano (Eds.), *Model science teacher preparation programs: An international comparison of what works* (pp. 29–52). Information Age Publishing Inc.
- Newton, P. E. (2013). Two kinds of argument? *Journal of Educational Measurement*, 50(1), 105–109.
- Otero, V. K., & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. *Journal of Research in Science Teaching*, 45(4), 497–523. <https://doi.org/10.1002/tea.20229>
- Ozogul, G., Olina, Z., & Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development*, 56(2), 181–201. <https://doi.org/10.1007/s11423-006-9012-7>
- Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher*, 45(1), 13–31. <https://doi.org/10.1007/s13384-018-0258-y>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30, 100329. <https://doi.org/10.1016/j.edurev.2020.100329>
- Park, S., Jang, J. Y., Chen, Y. C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching? Evidence from an empirical study. *Research in Science Education*, 41(2), 245–260. <https://doi.org/10.1007/s11165-009-9163-8>
- Park, S., & Oliver, J. S. (2008). National Board Certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45(7), 812–834. <https://doi.org/10.1002/tea.20234>
- Park, S., & Suh, J. K. (2019). The PCK Map Approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 185–197). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_8](https://doi.org/10.1007/978-981-13-5898-2_8)
- Patton, M. Q. (1990). *Qualitative evaluation and research methods [Nachdr.]* (2nd ed.). Sage.
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, 51(1), 59–81. <https://doi.org/10.1080/00461520.2016.1145550>
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. <https://doi.org/10.1080/02602930902862859>
- Reynolds, W. M., & Park, S. (2021). Examining the relationship between the educative teacher performance assessment and preservice teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, 58, 721–748. <https://doi.org/10.1002/tea.21676>
- Ruys, I., Keer, H. V., & Aelterman, A. (2012). Examining pre-service teacher competence in lesson planning pertaining to collaborative learning. *Journal of Curriculum Studies*, 44(3), 349–379. <https://doi.org/10.1080/00220272.2012.675355>
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159–179. <https://doi.org/10.1080/02602930801956059>
- Sato, M. (2014). What is the underlying conception of teaching of the edTPA. *Journal of Teacher Education*, 65(5), 421–434. <https://doi.org/10.1177/0022487114542518>
- Scherer, R. (2017). The quest for the holy grail of validity in science assessments: A comment on Kampa and Köller (2016) "German National Proficiency Scales in Biology: Internal Structure, Relations to General Cognitive Abilities and Verbal Skills". *Science Education*, 101(5), 845–853. <https://doi.org/10.1002/sce.21278>
- Schreiber, L. M., Paul, G. D., & Shibley, L. R. (2012). The development and test of the public speaking competence rubric. *Communication Education*, 61(3), 205–233. <https://doi.org/10.1080/03634523.2012.670709>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- SenBJF. (2017). *Handbuch Vorbereitungsdiens: Materialien für den reformierten Berliner Vorbereitungsdiens* [Guide for teacher induction in Berlin(6th ed.)]. [https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch\\_vorbereitungsdienst.pdf](https://www.berlin.de/sen/bildung/fachkraefte/lehrausbildung/vorbereitungsdienst/handbuch_vorbereitungsdienst.pdf)
- Settlage, J. (2013). On acknowledging PCK's shortcomings. *Journal of Science Teacher Education*, 24(1), 1–12. <https://doi.org/10.1007/s10972-012-9332-x>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.

- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2019). Standards für die Lehrerbildung: Bildungswissenschaften [Standards for Teacher Education]. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf)
- Stender, A., Brückmann, M., & Neumann, K. (2017). Transformation of Topic-Specific professional knowledge into personal pedagogical content knowledge through lesson planning. *International Journal of Science Education*, 39(12), 1690–1714. <https://doi.org/10.1080/09500693.2017.1351645>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tomas, C., Whitt, E., Lavelle-Hill, R., & Severn, K. (2019). Modeling holistic marks with analytic rubrics. *Frontiers in Education*, 4, 89. <https://doi.org/10.3389/feduc.2019.00089>
- Turley, E. D., & Gallagher, C. W. (2008). On the “Uses” of rubrics: Reframing the Great Rubric Debate. *The English Journal*, 97(4), 87–92.
- de Vaus, D. A. (2014). *Surveys in social research*. Social research today (6th ed.). Routledge.
- Vogelsang, C., Kulgemeyer, C., & Riese, J. (2022). Learning to plan by learning to Reflect?—Exploring relations between professional knowledge, reflection skills, and planning skills of preservice physics teachers in a one-semester field experience. *Education Sciences*, 12(7), 479. <https://doi.org/10.3390/educsci12070479>
- Vogelsang, C., & Riese, J. (2017). Wann ist eine Unterrichtsplanung ‘gut’?—Planungsperformanz in Praxisratgebern zur Unterrichtsplanung [What makes good lesson planning? Planning performance in advice literature for lesson planning]. In S. Wernke & K. Zierer (Eds.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (pp. 47–61). Verlag Julius Klinkhardt.
- Wegener, D. T., & Blankenship, K. L. (2007). Ecological validity. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 275–277). Sage Publications.
- Weitzel, H., & Blank, R. (2020). Pedagogical content knowledge in peer dialogues between pre-service biology teachers in the planning of science lessons. Results of an intervention study. *Journal of Science Teacher Education*, 31(1), 75–93. <https://doi.org/10.1080/1046560X.2019.1664874>
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292–305.
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (expanded 2nd ed.). Association for Supervision and Curriculum Development. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=133964>
- Wilson, M., de Boeck, P., & Carstensen, C. (2008). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts: State of the art and future prospects* (pp. 91–120). Hogrefe & Huber.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest version 2: Generalised item response modelling software*. Australian Council for Educational.
- Zaragoza, A., Seidel, T., & Hiebert, J. (2021). Exploring preservice teachers' abilities to connect professional knowledge with lesson planning and observation. *European Journal of Teacher Education*, 1–20. <https://doi.org/10.1080/02619768.2021.1996558>
- Zaragoza, A., Seidel, T., & Santagata, R. (2023). Lesson analysis and plan template: Scaffolding preservice teachers' application of professional knowledge to lesson planning. *Journal of Curriculum Studies*, 55(2), 13–8152. <https://doi.org/10.1080/00220272.2023.2182650>
- Zhai, X., Krajcik, J., & Pellegrino, J. W. (2021). On the validity of machine learning-based next generation science assessments: A validity inferential network. *Journal of Science Education and Technology*, 30(2), 298–312. <https://doi.org/10.1007/s10956-020-09879-9>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Großmann, L., & Krüger, D. (2023). Assessing the quality of science teachers' lesson plans: Evaluation and application of a novel instrument. *Science Education*, 1–37. <https://doi.org/10.1002/sce.21832>

**Appendix: The Rubric to Assess Science Lesson Plans (RALP; Großmann & Krüger, 2022a)**

Criteria		Level 1 – Sufficient	Level 2 – Excellent
<b>O Knowledge of Science Curriculum (KSC) – Intended Learning Outcomes</b>			
O1	Intended learning outcome in the unit	The tabular learning unit contains a list of all lessons, in which the topic and the ... are mentioned. ... areas of competence to be developed ...	... operationalized and precisely formulated learning objectives based on curricular standards ...
O2	Progression throughout the unit	The learning unit has a clear focus on one competence that ... ... is progressively developed in some lessons in increasing complexity, abstraction, or independence.	... is developed progressively in increasing complexity, abstraction, or independence in most lessons. Content-related and process-related competencies are linked in a meaningful way; that is, in the lessons, connections are made to already acquired knowledge elements, and these are used in plausible, practical situations.
O3	Intended learning outcome in the lesson	The intended learning outcome is operationalized and is based on a curricular standard. The intended learning outcome ... ... describes the learning process rather than the learning goal.	... describes precisely what the students will be able to do better after the lesson than before.
O4	Indicators as evidence of the desired learning	The indicators for successful learning relate explicitly to the intended learning outcome and describe ... ... students' behavior.	... students' observable with suitable operators so that the achievement of the intended learning outcome can be demonstrated.
<b>C Knowledge of Science Curriculum (KSC) – Biological Content</b>			
C1	Analysis of biological content	In the analysis of the biological content, the topic is correctly ... ... analyzed from a mainly scientific perspective.	... presented as precisely as possible and as detailed as necessary from a mainly didactical perspective with the help of scientific and didactical literature, that is, analyzed in its structure (inner logic, levels, perspectives).
C2	Choice of topic	The choice of the topic to foster the competence described in the intended learning outcome is ... ... justified with pragmatic arguments (e.g., demands of the curriculum/internal school curriculum, shortage of time).	... with didactical relevance (e.g., relevance to students, relevance to society, relevance to science) and principles (e.g., problem orientation, action orientation).

C3	Educational reconstruction	... educationally reduced by excluding or simplifying selected aspects of the broader topic.	The topic is ... ... educationally reconstructed by didactically reducing selected aspects of the learning object given the learning group conditions and by describing and justifying a promising learning path.
C4	Selection of relevant technical terms	... to some relevant but also less relevant terms.	The technical terms to be used in the lesson are limited... ... to a few particularly relevant (subject-related) terms and the selection will be justified.
<b>S Knowledge of Students' Understanding in Science (KSU)</b>			
S1	Analysis of the level of competence	The students' competence level is analyzed in relation to the intended learning outcome for prototypical students on at least two levels by describing ... ... generally related aspects in which the students have already or not yet developed competencies.	... specific aspects related explicitly to the intended learning outcome in which students have already or not yet developed competencies.
S2	Progression of competence development	Between the students' current level of competence and the intended learning outcome for the lesson, a progression is ... ... partially recognizable because the intended learning outcome goes beyond the students' already developed competencies in a few aspects.	... recognizable because the intended learning outcome goes beyond the already developed competencies of the students in significant aspects.
S3	Students' conceptions	Topic-specific students' conceptions are ... ... described.	... explicitly described for the particular students. Students' conceptions are considered in the instructional design of the lesson.
S4	Learning difficulties	Learning difficulties are described that address... ... general educational problems (e.g., language barriers).	... topic-specific problems (e.g., the abstractness of the topic) are considered in the instructional design of the lesson.
S5	Methodical skills	The students' methodological skills necessary to achieve the intended learning outcome are ... ... analyzed.	... analyzed, and the description is limited to aspects necessary to achieve the learning objective.
<b>I Knowledge of Instructional Strategies for Teaching Science (KISR)</b>			
I1	Structure of the development of competencies	The phases are built upon each other and lead to a clear goal, whereby the selected competency focus is ... ... shown in the tasks of the elaboration phase and as a focus in the final phase of the lesson.	... is shown evidently in all phases and phase transitions.



I2	Lesson structure	The lesson structure, that is, the order of the phases, is...	
		... described chronologically by presenting selected instructional steps of the various phases in their design and planned progression.	... justified in chronological order by explaining the contribution of selected instructional steps of the various phases to achieving the intended learning outcome.
I3	Suitability of the methods	The selected methods/activities are suitable ...	
		... in some cases because they enable content- or process-related achievement of the intended learning outcome.	... because they enable content- or process-related achievement of the intended learning outcome.
I4	Suitability of tasks	The tasks mainly address the intended learning outcome. The design of the tasks in the elaboration phase is ... with regard to selected didactic aspects (e.g., degree of difficulty, openness, linguistic complexity).	
		... analyzed ...	... analyzed and justified with regard to the students' levels of competence or the competence area to be promoted ...
I5	Suitability of materials	The media (e.g., worksheets, films, help cards) address the targeted growth of competencies ...	
		... in some cases.	... for the major part. The design of the media is analyzed regarding selected didactical aspects (e.g., primary/secondary experience, abstractness, complexity) and justified regarding the learning group preconditions or the competence area to be promoted.
I6	Horizon of expectations	A Horizon of expectations is ...	
		... provided and correct.	... provided, correct, and appropriate to the students' current levels of competence.
I7	Adaptive teaching	Measures of adaptive teaching are intended to help ... achieve the intended learning outcome.	
		... individual students (disabilities, linguistic barriers) or the low-performing students...	... both low- and high-performing students...
I8	Anticipated difficulties and alternatives	Anticipated difficulties in the learning process are...	
		... mainly pragmatic (e.g., technical aspects) or general-pedagogical (e.g., disruptions).	... subject-related or subject-didactical.  In addition, alternative courses of action are developed to ensure that the desired increase in competence can be achieved.
I9	Transparency of the learning process	The structure of the lesson is made transparent to the students using structuring support which ...	
		... shows the course of the lesson in the sense of a program flow concerning organizational details (e.g., time).	... clearly shows the course of the lesson concerning the competencies to be developed.

A Knowledge of <u>A</u> ssessment of Science Learning				
A1	Transparency performance expectations	<p>Performance expectations in the form of criteria for the intended learning outcome (content, form, and scope) are provided to the students ...</p> <table border="1"> <tr> <td>... partially by listing some relevant criteria to which the students can refer.</td> <td>... fully transparent by stating relevant, specific criteria to which students can refer.</td> </tr> </table>	... partially by listing some relevant criteria to which the students can refer.	... fully transparent by stating relevant, specific criteria to which students can refer.
... partially by listing some relevant criteria to which the students can refer.	... fully transparent by stating relevant, specific criteria to which students can refer.			
A2	Products of students' learning	<p>In the final phase of the lesson, the products of students' learning are...</p> <table border="1"> <tr> <td>... presented. Students are partially active by presenting their results or giving feedback on other students' findings. The teacher gives feedback on a general-pedagogical level (e.g., cooperation).</td> <td>... presented and discussed. Students are mainly active, and the teacher gives feedback on their competence growth.</td> </tr> </table>	... presented. Students are partially active by presenting their results or giving feedback on other students' findings. The teacher gives feedback on a general-pedagogical level (e.g., cooperation).	... presented and discussed. Students are mainly active, and the teacher gives feedback on their competence growth.
... presented. Students are partially active by presenting their results or giving feedback on other students' findings. The teacher gives feedback on a general-pedagogical level (e.g., cooperation).	... presented and discussed. Students are mainly active, and the teacher gives feedback on their competence growth.			

*Note.* Level 0 (insufficient) is not displayed here since it would indicate the absence of a particular aspect for most criteria. Since this information has no added value for PSTs and TSTs, we display the two desirable levels 1 and 2.

## Beitrag 8

**Großmann, L.** & Krüger, D. (2022). Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen, *SEMINAR – Lehrerbildung und Schule*, 39(1), 91-110.

Dieser Beitrag ist Teil folgender Publikation: <https://dx.doi.org/10.3278/SEM2201W>

## Beitrag 9

**Großmann, L.**, Lotz, A., Mulke, S. & Krüger, D. (2022). Probleme beim Planen?! Acht Ratschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie, *MNU-Journal*, 75 (05), S. 368-376.



# Probleme beim Planen?

## Acht Vorschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie

LEROY GROßMANN – ALEXANDER LOTZ – STEFAN MULKE – DIRK KRÜGER

Die Unterrichtsplanung gehört zu den zentralen Aufgaben von Lehrkräften. In der ersten und insbesondere in der zweiten Phase der Lehrkräftebildung werden ausführliche schriftliche Unterrichtsentwürfe verlangt, in denen Planungsentscheidungen beschrieben und begründet werden müssen. Im Beitrag werden Einblicke in das Schreiben von Unterrichtsentwürfen angehender Biologie-Lehrkräfte gegeben und Hilfen für häufig auftretende Probleme angeboten. Dies geschieht auf der Basis von Erwartungen erfahrener Biologie-Fachseminarleiter der zweiten Phase sowie aus der Ausbildungsperspektive der ersten Phase.

### 1 Einleitung

Während die Unterrichtsplanung gemeinhin eine Kernaufgabe von Lehrkräften darstellt (KMK, 2019), wird das Verfassen schriftlicher Unterrichtsentwürfe von angehenden Lehrkräften oftmals als große und durchaus auch überflüssige Last empfunden. Für sie stehen häufig Aufwand und Nutzen in keinem angemessenen Verhältnis. Dieser Eindruck wird auch dadurch erzeugt, dass nur etwa die Hälfte der in der Lehrkräftebildung tätigen Seminarleitungen ihre Bewertungskriterien transparent macht (STRIETHOLT & TERHART, 2009). Zudem werden entsprechende Bewertungskriterien von zwei Dritteln der Referendar/innen und sogar von einem Drittel der Seminarleitungen als uneinheitlich empfunden (DÖBRICH & ABS, 2008), was möglicherweise zusätzlich zu Unzufriedenheit, Irritationen und infolgedessen zu Problemen beim Verfassen eines Unterrichtsentwurfs beiträgt. Derartige Probleme wurden auch im Rahmen einer

Analyse von Staatsexamensentwürfen (GROßMANN & KRÜGER, 2020; 2021) identifiziert.

Im Folgenden werden Vorschläge gemacht, die angehenden Lehrkräften in der Unterrichtsplanung im Allgemeinen und beim Schreiben ausführlicher Unterrichtsentwürfe im Besonderen helfen sollen.

### 2 Unterrichtsplanung als fachdidaktisches Problemlösen

Das Schreiben eines Unterrichtsentwurfs stellt ein Problem dar, das es zu lösen gilt. Problemlösen meint die Fähigkeit, von einem gegebenen Ausgangszustand aus mithilfe geeigneter Operationen zu einem gewünschten Zielzustand zu gelangen (FUNKE, 2003). Konkret für das Planen von Unterricht bedeutet das, im Entwurf die getroffenen Planungsentscheidungen im

Rahmen des Unterrichtsentwurfs zu legitimieren, die den Schüler/innen ausgehend vom aktuellen Kompetenzstand mithilfe einer geeigneten Lehr-Lern-Struktur eine Kompetenzförderung ermöglichen (Abb.1).

Als zentrale Kategorie für die Einschätzung der Angemessenheit einer Handlung wird hier deren Funktionalität verwendet. Eine Handlung ist dann funktional, wenn sie einen Beitrag zum Erreichen des angestrebten Zielzustandes leistet. In einem Unterrichtsentwurf werden gleich zwei Zielzustände angestrebt (Abb.1):

1. Auf der Ebene des Lehr-Lern-Prozesses werden im Planungsprozess Planungsentscheidungen getroffen, die den Schüler/innen Kompetenzzuwachs ermöglichen. Die Lehr-Lern-Struktur (Phasen, Methoden, Aufgaben etc.) soll funktional sein, um die ausgewählte Kompetenz der spezifischen Lerngruppe am ausgewählten Thema zu fördern (*Problem 1. Art*).
2. Auf der Ebene der Kommunikationssituation zwischen Autor/in des Entwurfs als Planungsprodukt und Rezipient/innen (Seminarleitung, Schulleitung, Kolleg/innen) gilt es, die Planungsentscheidungen nachvollziehbar darzustellen und vor allem zu legitimieren (*Problem 2. Art*). Die Darstellung ist dann funktional, wenn sie sich auf das Notwendige beschränkt.

Sämtliche Ausbildungsinstitutionen der ersten und zweiten Phase der Lehrkräftebildung verlangen bestimmte Formate der Unterrichtsplanung mit zum Teil dezidierten Vorgaben, welche Aspekte in den jeweiligen Kapiteln behandelt werden müssen. Gleiches gilt für die Vielzahl der allgemeindidaktischen Planungsratgeber (vgl. für eine Übersicht VOGELSANG & RIESE, 2017). Auf diese Weise wird suggeriert, dass der schriftliche Unterrichtsentwurf in seiner linearen Abfolge von Kapiteln bloß fleißig mit Inhalten gefüllt werden muss. Der schriftliche Unterrichtsentwurf ist jedoch kein Formular, sondern ein Instrument, in dem das gedankliche Probehandeln in der Unterrichtsplanung schriftlich fixiert und somit den Teilnehmenden im Rahmen von Hospitationen zugänglich gemacht wird.

Als eine der größten Herausforderungen angehender Lehrkräfte beim Verfassen schriftlicher Unterrichtsplanungen hat GASSMANN (2013) die möglichst vollständige Vernetzung der verschiedenen planungsrelevanten Aspekte identifiziert, d.h. der analysierten Vorbedingungen (curriculare Standards, Sachstruktur des Themas, Kompetenzstand der Lerngruppe und der daraus entwickelten Lehr-Lern-Struktur (Abb. 1). Hierin zeichnet sich zwangsläufig ein Dilemma ab: Einerseits erfordert die Planung von Lehr-Lern-Prozessen ein zirkuläres Abstimmen von Planungsentscheidungen, andererseits sollen diese in das etablierte Format eines schriftlichen Unterrichtsentwurfs mit seiner linearen Kapitelfolge eingefügt werden. Das Schreiben des Entwurfs selbst regt also gar nicht dazu an, kapitelübergreifend zu denken und beispielsweise in der Gestaltung der Sicherungsphase auf einen Aspekt der Lerngruppenanalyse zurückzugreifen, der einige Seiten zuvor erwähnt wurde. Daraus resultiert eine Reihe von Problemen beim Schreiben eines Unterrichtsentwurfs, die in der empirischen Forschung bislang noch nicht hinreichend untersucht worden sind (WEINGARTEN, 2019) und für die wir im Folgenden Vorschläge machen.

### 3 Vernetzung der Planungsentscheidungen (Probleme 1. Art)

Die ersten sechs Probleme betreffen einerseits die Wahrnehmung und Interpretation unterrichtsrelevanter Voraussetzungen, d.h. der curricularen Standards (Probleme 1, 2 und 3), der Sachstruktur des Themas (Problem 4), der Lerngruppenvoraussetzungen (Problem 5) sowie die funktionalen Phasengestaltung (Problem 6) als Aspekt kreierenden Planungshandelns (Abb. 1).

#### 3.1 Problem 1: Es fehlt ein langfristiges Konzept zur Kompetenzförderung.

Kompetenzentwicklung ist ein langfristiger Prozess, der einen systematischen Wissenserwerb (inhaltsbezogen: Fachwissen) und Fortschritte im Können (prozessbezogen: Erkenntnisgewinnung, Kommunikation, Bewertung) umfasst (LERSCH, 2010). In

vielen Entwürfen wird jedoch ausschließlich die Einzelstunde fokussiert, ohne den angestrebten Kompetenzzuwachs in ein langfristiges Konzept zur Kompetenzförderung einzubetten (WEINGARTEN, 2019). Dies zeigt sich darin, dass die zumeist tabellarische Planung der Unterrichtssequenz entweder fehlt oder sich auf eine bloße Aneinanderreihung von Stundenthemen beschränkt (Abb. 2). Dadurch wird es erschwert, den längerfristigen Lernprozess der Schüler/innen über eine Beschreibung einer Abfolge der bereits behandelten und der noch zu behandelnden Unterrichtsthemen hinaus zu analysieren und zu reflektieren. Hilfreich ist es hingegen, einen Kompetenzschwerpunkt festzulegen, der im Laufe einer Unterrichtssequenz von etwa acht bis zwölf Stunden gefördert werden soll. Dabei ist

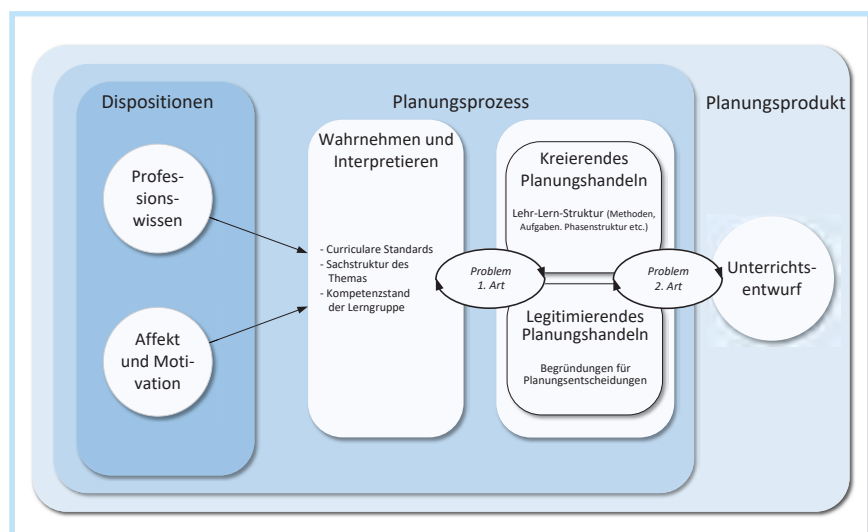


Abb. 1. Kompetenzmodell schriftlicher Planungskompetenz (nach BLÖMEKE, GUSTAFSSON & SHAVELSON, 2015, KÖNIG et al. 2017)

Std.	Thema	Kompetenzförderung <i>Die Schüler/innen ...</i>
1	Schülervorstellungen zur Pflanzenernährung	<i>... erklären, wie sich aus einem 1 g schweren Samen einen mehrere Tonnen schwerer Baum entwickelt, indem sie ihre Vorstellungen in Gruppen diskutieren und den Begriff „Fotosynthese“ definieren. (Fachwissen)</i>
2	Pflanzenwachstum I: Planung des Experiments	<i>... planen ein Experiment, um zu überprüfen, ob Pflanzen Erde zum Wachsen brauchen, indem sie Wachstum als abhängige, Erde als unabhängige sowie Licht, Sand und Wasser als zu kontrollierende Variablen definieren. (Erkenntnisgewinnung) Wann werten sie das aus?</i>
3	Pflanzenwachstum II: Übung zur Variablenkontrolle	<i>... wenden die Begriffe abhängige, unabhängige und zu kontrollierende Variable an, indem sie sie in verschiedene Versuchsdesigns mit verschiedenen Hypothesen zum Pflanzenwachstum zuordnen und begründen, welcher Versuchsansatz die Kontrolle darstellt. (Erkenntnisgewinnung)</i>
4	Sauerstoffnachweis I: Planung des Versuchs	<i>... planen einen Versuch, um die Frage zu beantworten, welches Gas bei der Fotosynthese entsteht, indem sie mithilfe bereit gestellter Materialien einen Versuchsaufbau entwickeln. (Erkenntnisgewinnung)</i>
5	Sauerstoffnachweis II: Durchführung und Auswertung	<i>... führen den Sauerstoffnachweis durch und werten ihn aus, indem sie CO<sub>2</sub> als Ausgangsstoff und O<sub>2</sub> als Produkt der Fotosynthese identifizieren und anhand eines Fachtextes erklären, dass der Kohlenstoff aus dem CO<sub>2</sub> in Glucose (C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>) überführt wird. (Fachwissen)</i>
6	Stärkenachweis	<i>... weisen mithilfe Lugolscher Lösung Stärke in Geranienblättern nach und erklären, warum Pflanzen als Produkt der Fotosynthese Stärke und nicht Glucose speichern. (Fachwissen)</i>
7	Fotosynthese-Grundversuch I: Planung des Experiments	<i>... planen ein Experiment mit Elodea, Indigocarmin und Natriumdithionit, bei dem in mehreren Versuchsansätzen einzelne Variablen (Licht, Temperatur, CO<sub>2</sub>-Gehalt) variiert werden, indem sie für jeden Versuchsansatz die unabhängige Variable bestimmen und jeweils erklären, welcher Versuchsansatz als Kontrolle für welchen anderen Versuchsansatz gelten kann. (Erkenntnisgewinnung)</i>
8+9	Fotosynthese-Grundversuch II: Durchführung und Auswertung	<i>... werten den Fotosynthese-Grundversuch aus, indem sie erklären, dass ein hoher Carbonat-Gehalt im Wasser (CO<sub>2</sub>) und Licht die notwendigen Voraussetzungen für die Fotosynthese bei Elodea sind und diese von der Wassertemperatur beeinflusst wird. (Fachwissen)</i>
10	Synthese I: Fotosynthese-Gleichung	<i>... erklären anhand der Ergebnisse des Sauerstoff- und des Stärkenachweises, dass sich Pflanzenwachstum auf die aufgenommenen CO<sub>2</sub>- und H<sub>2</sub>O-Moleküle zurückführen lässt, indem sie die Reaktionsgleichung der Fotosynthese analysieren und die Kohlenstoff- und Sauerstoffatome in einer Skizze verfolgen. (Fachwissen)</i>
11	Synthese II: Übung zur Fotosynthese-Gleichung	<i>... sagen begründet vorher, dass bei der „Fotosynthese“ von Schwefelwasserstoff-Bakterien neben C<sub>6</sub>H<sub>12</sub>O<sub>6</sub> und H<sub>2</sub>O elementarer Schwefel entsteht, indem sie in Gruppen eine Summengleichung entwickeln und die Kohlenstoff- und Schwefelatome in einer Skizze verfolgen. (Fachwissen)</i>
12	Zusammenfassung	<i>... fassen in Form einer Concept-Map zusammen, was das Experimentieren als Arbeitsweise auszeichnet und welche Aspekte (Variablen, Kontrollansätze) in der Planung von Experimenten beachtet werden müssen. (Erkenntnisgewinnung)</i>

Abb. 2. Sequenzplanung zum Thema Fotosynthese in der Sekundarstufe I. Eine Sequenzplanung ist mehr als eine Auflistung der Themen und gelegentlich der Kompetenzbereiche, die jeweils fokussiert werden sollen. Die Schüler/innen werden erst dann schrittweise Kompetenzzuwachs erfahren, wenn in der Sequenz ein klarer Schwerpunkt vorliegt (hier: Erkenntnisgewinnung – Experimente planen) und dieser in mehreren Stunden wiederholt aufgegriffen und progressiv gefördert wird. Die Progression besteht in diesem Fall darin, dass die Experimente zunehmend komplexer werden (Std. 2, 4, 7), da die Anzahl der zu berücksichtigenden Variablen zunimmt. Dies wird dann transparent, wenn man für jede Stunde neben Thema und Kompetenzbereich auch das Lernziel angibt (kursiv).

wichtig, dass dieser Kompetenzschwerpunkt schrittweise und in zunehmender Abstraktion, Eigenständigkeit und Komplexität gefördert wird. Eine Sequenz sollte also eine Progression aufweisen und die zu fördernden Kompetenzen jeweils mit geeig-

neten und relevanten Inhalten verknüpfen (LERSCH, 2010; MEISERT, 2016). Alternativ ist es auch möglich, ein langfristiges Konzept zur Kompetenzförderung nicht anhand einer mehrwöchigen Sequenz, sondern über das gesamte Schuljahr oder

eine Doppeljahrgangsstufe hinweg zu entwickeln. Neben der tabellarischen Sequenzplanung wäre dann wichtig, in wenigen Sätzen zu erläutern, inwiefern sich die geplante Einzelstunde in ein derartig langfristig angelegtes Konzept einfügt, d.h. welche Schritte auf dem Weg zum ausgewählten Standard die Lernenden bereits absolviert haben und welcher Fortschritt in der geplanten Stunde angestrebt wird.

#### Problem gelöst?

- Haben Sie zusätzlich zu einer Auflistung der Themen und ggf. der Kompetenzbereiche für jede Stunde ein Lernziel formuliert?
- Haben Sie einen eindeutigen inhalts- bzw. prozess-bezogenen Kompetenzschwerpunkt gewählt, den Sie im Laufe der Sequenz vorrangig fördern wollen?
- Bedienen Sie diesen Kompetenzschwerpunkt wiederholt im Laufe der Sequenz (oder alternativ im Laufe des Schuljahres bzw. der Doppeljahrgangsstufe)?
- Liegt Ihrer Sequenz eine Progression zugrunde, so dass die Lernenden zunehmend stärker gefordert werden und somit der ausgewählte Schwerpunkt schrittweise gefördert wird?

Kasten 1. Leitfragen zur Sequenzplanung (Problem 1)

### 3.2 Problem 2: Das Lernziel wird unpräzise formuliert.

Viele angehende Lehrkräfte versuchen, einen Standard in einer Unterrichtsstunde zu „erfüllen“, als wäre dieser Standard dann abschließend bearbeitet und als könne man zu einem anderen übergehen. Dadurch wird die Komplexität, die in einer Standardformulierung des Lehrplans angelegt ist, zu sehr reduziert, weil die Schritte auf dem Weg zu diesem Standard nicht analysiert werden. Auffällig ist, dass angehende Lehrkräfte häufig mehr als einen Standard auswählen, was jedoch naturgemäß zu einer unklaren Schwerpunktsetzung und somit zu einer fehlen-

den Fokussierung führen kann (Abb. 3, links). Folglich wird Unterricht selten langfristig progressiv geplant (Problem 1). Damit Lehr-Lern-Prozesse zielführend auf die Kompetenzentwicklung von Schüler/innen ausgerichtet sind, sollte ein Standard des Lehrplans ausgewählt und für das jeweilige Unterrichtsthema und die geplante methodisch-didaktische Umsetzung konkretisiert werden (*Lernziel*). Es wird empfohlen, im Rahmen einer sogenannten *Standardexegese* zu analysieren, welche vielfältigen und z.T. komplexen Teiloperationen Schüler/innen bewältigen müssen, um das in einem Standard formulierte Ziel zu erreichen. Darauf aufbauend wird dann ein Lernziel formuliert, das die ursprüngliche Formulierung des Standards im Lehrplan (z.B. den Operator) wieder aufgreift und für das konkrete Unterrichtsvorhaben präzisiert (WITTMER, KRATSCHMAYR & VOSS, 2020) (Abb. 3, rechts).

#### Problem gelöst?

- Haben Sie sich auf ein einziges Lernziel beschränkt?
- Greifen Sie in Ihrem Lernziel das Verb aus dem Standard des Lehrplans wieder auf?
- Beschreiben Sie überwiegend das Lernziel und weniger den Lernprozess?
- Sind die ausgewählten Methoden geeignet, um das Lernziel zu erreichen?

Kasten 2. Leitfragen zur Formulierung des Lernziels (Problem 2)

### 3.3 Problem 3: Die Indikatoren sind unpräzise und geben bloß den Stundenverlauf wieder.

Die in dem konkretisierten Standard beschriebene Erwartung an das Lernergebnis stellt den konkreten Beitrag der Unterrichtsstunde für die längerfristige Kompetenzentwicklung dar. Auf der Basis des Lernziels werden Indikatoren formuliert, anhand derer das Erreichen der angestrebten Lernergebnisse erkennbar werden soll (WIGGINS & MCTIGHE, 2005). Diese Indikatoren bilden die didaktisch-methodischen Bezüge für die

Die SuS kennen nach der Stunde die Zellorganellen und ihre Funktion, indem sie einen Text über die Pflanzenzelle lesen und eine Tabelle vervollständigen, in der sie zu jedem Zellorganell die Funktion notieren.



Die SuS erklären den Zusammenhang zwischen Struktur und Funktion der wichtigsten Zellorganellen der Pflanzenzelle (Vakuole, Chloroplasten, Mitochondrien), indem sie jeweils Bau und Funktion dieser Organellen tabellarisch beschreiben und begründen, warum die jeweiligen Reaktionen nicht in einem gemeinsamen Kompartiment stattfinden können.

Abb. 3. Lernziel zum Kompetenzbereich Fachwissen „Die Schüler/innen können die Zusammenhänge von Struktur und Funktion von Zellkompartimenten erklären“ (SenBJF, 2015, 17). Links: Statt des schwer erfassbaren Verbs „kennen“ wäre es hilfreich, den Operator „erklären“ aus dem Standard aufzugreifen und anschließend eher das Lernziel als den Lernprozess zu beschreiben. Zudem werden zwei Schwerpunkte (Fachwissen: Zellorganellen; Kommunikation: Informationen aus Texten entnehmen, Tabelle entwickeln) gesetzt. Rechts: Der Operator „erklären“ aus dem Standard wird aufgegriffen und so konkretisiert, dass über die Verknüpfung von Struktur und Funktion sowie die Begründung für die Kompartimentierung nachweisbar wird, ob die Schüler/innen tatsächlich den Zusammenhang zwischen Struktur und Funktion der Zellorganellen erklären können oder nicht. In diesem Fall ist die Tabelle ein Hilfsmittel, um die Informationen strukturieren und darauf aufbauend den Zusammenhang zwischen Struktur und Funktion erklären zu können. Statt des Lernprozesses wird hier ein Lernziel präzise beschrieben.



konkrete Gestaltung der Lehr-Lern-Prozesse, die als Gesprächsanlass für die Reflexion der Stunde genutzt werden können und sollen. Oft wird jedoch der Stundenverlauf vom Einstieg bis zur Sicherung stichpunktartig dargestellt, was schon aus dem Stundenverlaufsplan hervorgeht (Abb. 4). Stattdessen sollen die Indikatoren die Erwartungen an das Lernergebnis so konkretisieren, dass sie eine gute Orientierung geben, wie das Arbeitsmaterial, die Arbeitsaufträge, das Unterrichtsgespräch usw. zu gestalten sind. Drei bis maximal fünf Indikatoren, insbesondere aus der Erarbeitungs- und Sicherungsphase, genügen. Es wäre hilfreich, sie so operationalisiert zu formulieren, dass man direkt beobachten kann, ob die Schüler/innen den angestrebten Kompetenzzuwachs erfahren haben. Zu diesem Zweck hilft die Nutzung von Verben, die in Operatorenlisten (z.B. KMK, 2004) definiert sind, die Handlungen der Schüler/innen präzise beschreiben und an denen Lehrkräfte diagnostizieren können, ob Schüler/innen den angestrebten Kompetenzzuwachs erfahren haben.

#### Problem gelöst?

- Stehen Ihre Indikatoren in eindeutigen Zusammenhang zum Lernziel?
- Sind Ihre Indikatoren präzise, operationalisiert und somit eindeutig beobachtbar?
- Haben Sie sich auf maximal fünf wichtige Indikatoren beschränkt?

Kasten 3. Leitfragen zu den Indikatoren als Nachweis des Kompetenzzuwachses (Problem 3)

### 3.4 Problem 4: Die Sachstrukturanalyse ist lediglich eine Sachanalyse.

Fachwissen ist ein wichtiger Teil der Kompetenz von Lehrkräften (BAUMERT & KUNTER, 2006). Allerdings wird die Darstellung des biologischen Fachinhaltes in vielen Unterrichtsentwürfen missverstanden: Oft wird das Unterrichtsthema entweder nahezu gar nicht (GASSMANN, 2013) oder in großer Ausführlichkeit nur fachlich beschrieben, ohne dass sein Potenzial für die Kompetenzentwicklung analysiert wird (Abb. 5). Eine solche reine Fokussierung auf den Fachinhalt wird als Sachanalyse bezeichnet. Eine *Sachstrukturanalyse* im Sinne einer fach-

lichen Klärung hingegen meint eine didaktische Perspektive, d.h. den Blick auf die Vermittlung von Inhalten und deren Zusammenhänge zu richten (HEIDENREICH & GROPENGIESSER, 2021). Wenn sich in der Sachstrukturanalyse zeigt, welche Teilaspekte des Unterrichtsgegenstandes relevant bzw. weniger relevant sind, welche als grundlegend oder nur als ergänzend zu betrachten sind, welche komplex und besonders schwer zu verstehen bzw. welche einfach und leicht zu vermitteln sind und welche aufeinander aufbauen, dann ergeben sich wichtige Hinweise für die Entwicklung sinnvoller methodischer Schritte und geeigneter Arbeitsmaterialien. Dabei geht es darum, potentielle Lernschwierigkeiten zu identifizieren, die dem Unterrichtsgegenstand auf Grund seiner Komplexität, seines Umfangs oder seines Abstraktionsgrades innewohnen. Es wäre beispielsweise sinnvoll, Alltagsvorstellungen der Schüler/innen (z.B. HAMMANN & ASSHOFF, 2014) miteinzubeziehen. Zusätzlich sollte in einer Sachstrukturanalyse auch die bewusste Entscheidung getroffen werden, welche Fachbegriffe relevant sind und auf welche im Sinne einer didaktischen Reduktion verzichtet werden kann (DORFNER, FÖRTSCH & NEUHAUS, 2020). Bei sehr klarer und einfacher Sachstruktur kann eine Analyse sehr kurz und knapp ausfallen, bei sehr komplexen Inhalten in der Sekundarstufe II ist gegebenenfalls eine sehr umfangreiche Sachstrukturanalyse unverzichtbar. Sie sollte in jedem Fall auf das Niveau der Lerngruppe abgestimmt und auf fachliche und fachdidaktische Literatur gestützt sein.

#### Problem gelöst?

- Haben Sie für die fachwissenschaftliche Darstellung des Themas Literatur zitiert?
- Haben Sie analysiert, welche unterrichtsrelevanten Ebenen, Perspektiven, Teilthemen ihr Thema bietet, und daraus im Sinne einer didaktischen Reduktion eine begründete Auswahl getroffen?
- Haben Sie typische Schülervorstellungen und Lernschwierigkeiten zu Ihrem Thema berücksichtigt?
- Haben Sie sich begründet auf wenige, notwendige Fachbegriffe beschränkt, mit denen sich die Lernenden auseinandersetzen müssen?

Kasten 4. Leitfragen zur Sachstrukturanalyse (Problem 4)

Die Schülerinnen und Schüler...

~~legen ein Puzzle zu Spender- und Empfängerblut.~~

~~ordnen den Bestandteilen des Modells die Begriffe „Antigen“ und „Antikörper“ zu.~~

~~verstehen, dass nicht jede Blutgruppe jedem Menschen gespendet werden kann.~~

~~beschreiben mithilfe des Blutgruppen-Modells, welche Antigene und Antikörper sich im Blut von Personen mit den Blutgruppen A, B, AB und 0 befinden.~~

~~erkennen, dass die Blutgruppe 0 Menschen mit allen Blutgruppen gespendet werden kann.~~

~~erklären durch Zusammenlegen der passenden Antigene und Antikörper des Modells, warum Personen mit Blutgruppe 0 als Universalspender bezeichnet werden.~~

Abb. 4. Indikatoren zu einer Stunde mit folgendem Lernziel: „Die Schüler/innen können mithilfe von Blutgruppen-Modellen das ABO-System erklären.“ Um diagnostizieren zu können, ob das angestrebte Lernziel erreicht wurde, ist es hilfreich auf eine reine Beschreibung der Schüleraktivität zu verzichten und unklare Verben zu vermeiden (durchgestrichen). Mithilfe von Operatoren kann präzise beschrieben werden, woran Kompetenzzuwachs erkannt werden soll. Ob Schüler/innen etwas erkennen oder verstehen, lässt sich nicht direkt diagnostizieren, sondern indirekt, wenn sie z.B. den Zusammenhang zwischen Antigenen und Antikörpern bei Blutgruppe 0 erklären können.

Der Mensch hat nur einen und nicht zwei Blutkreisläufe. Mit dem großen Abschnitt werden im Körperkreislauf Organe, Gewebe und Zellen durchblutet und mit Sauerstoff und anderen wichtigen Substanzen versorgt. Im zweiten Abschnitt, dem kleinen Lungenkreislauf, gelangt der Sauerstoff, der mit der Atemluft eingeatmet wird, ins Blut (SADAVA et al., 2019). *Viele Schüler/innen haben eine unklare Vorstellung vom Begriff „Kreislauf“ und können häufig nicht benennen, dass der Mensch einen Kreislauf besitzt (KATTMANN, 2015). Aus diesem Grund habe ich in der Unterrichtssequenz statt der Begriffe „Körperkreislauf“ und „Lungenkreislauf“ wie von KATTMANN (2015) vorgeschlagen die Begriffe „Körperschleife“ und „Lungenschleife“ eingeführt. Auf diese Weise wird deutlich, dass zwei Schleifen gemeinsam einen einzigen Blutkreislauf bilden.*

Abb. 5. Auszug aus der Sachstrukturanalyse zu einer Stunde mit folgendem Lernziel: „Die Schüler/innen erläutern den Aufbau der Arterien und Venen und beschreiben den Zusammenhang von Struktur und Funktion der Blutgefäße.“ Oft stützt es die fachliche Qualität des Entwurfs, wenn für fachwissenschaftliche Aspekte eine Quelle angegeben wird. Es ist entscheidend, den fachlichen Inhalt nicht in zu großer Breite darzustellen, sondern auch auf mögliche Probleme in der didaktischen Umsetzung einzugehen (kursiv). Ausgehend von der fachlichen Darstellung wird eine häufige Schülervorstellung aus der biologiedidaktischen Literatur dargestellt und daraus eine didaktische Maßnahme abgeleitet.

Die Lerngruppe besteht aus 27 Schüler/innen, davon 14 Mädchen und 13 Jungen. Die Lerngruppe ist heterogen und zum Großteil relativ interessiert an Biologie. Vier Schüler/innen lernen Deutsch als Zweitsprache, weshalb sie für die Erarbeitung ein Wörterbuch nutzen können. Die Sitzordnung ist so angepasst, dass die vier Schüler/innen Sitznachbarn haben, welche sie während des Unterrichts um Rat fragen können. Die Schüler/innen können bereits die Wirkung der Konzentration von Neurotransmittern auf die chemische Synapse beschreiben und erklären. Die meisten Schüler/innen können jedoch keine Fließdiagramme entwickeln, um Informationen aus einem Fachtext zu veranschaulichen. Insbesondere fällt es ihnen schwer, wichtige Aussagen aus einem Text herauszufiltern, in eigenen Worten kurz zu paraphrasieren und dann in eine adressatengerechte grafische Form zu überführen.

Abb. 6. Auszug aus einer Lerngruppenanalyse zu einer Stunde mit dem Lernziel: „Die Schüler/innen können mithilfe eines Fachtextes ein adressatengerechtes Fließdiagramm zur Wirkung von Kokain auf die chemische Synapse erstellen.“ Vorwiegend irrelevante Informationen (z.B. Geschlechterverhältnis) können vermieden werden, wenn sie keinen erkennbaren Einfluss auf die Gestaltung der Stunde haben (durchgestrichen). Zudem sollten neben inhaltsbezogenen Aspekten (Wissen über Synapse) auch prozessbezogene Aspekte des ausgewählten Lernziels (Kompetenz, ein Fließdiagramm zu erstellen), analysiert werden (kursiv).

### 3.5 Problem 5: Es werden viele irrelevante Auskünfte über die Lerngruppe mitgeteilt.

Oft werden Unterrichtsvoraussetzungen ausführlich beschrieben, auch wenn sie unbedeutend bleiben oder zu allgemein sind (Abb. 6). Demgegenüber fällt die Analyse der für das konkrete Unterrichtsvorhaben bedeutsamen Voraussetzungen (Kompetenzstand, kognitives Niveau, bekannte Fachbegriffe, Methodenwissen, sprachliche Probleme) dann oft lückenhaft und oberflächlich aus, was auch an einer fehlenden Standardexegese (Problem 2) liegen kann. Lehr-Lern-Prozesse sollten adaptiv an die jeweilige Lerngruppe und deren Bedürfnisse angepasst werden (KÖNIG, BUCHHOLTZ & DOHMEN, 2015). Zu diesem Zweck ist es nötig, die Lerngruppe hinsichtlich ihrer

inhalts- und ihrer prozessbezogenen Kompetenzentwicklung zu analysieren. Weitere Informationen zur Lerngruppe (Motivation, soziales Gefüge o.ä.) sind nur dann darzustellen, wenn sie einen möglichen Einfluss auf das Erreichen des angestrebten Kompetenzzuwachses haben. Dazu gehören u.a. auch methodische Aspekte, beispielsweise Vorerfahrungen im Umgang mit einer Methode.

### 3.6 Problem 6: Die Aufgaben und die Sicherungsphase passen nicht zum Lernziel.

Die Aufgabenstellungen bilden den Kern der Erarbeitungsphase und sind somit von zentraler Bedeutung für die Kompetenzförderung. Erstaunlicherweise werden zwar Medien, Methoden usw. oft in großer Ausführlichkeit beschrieben und zum Teil begründet, die Aufgabenstellungen in ihrem kognitiven Aktivierungspotential (z.B. ihrer inhaltlichen und sprachlichen Komplexität, Abstraktheit usw.; MAIER et al., 2014) jedoch sehr selten. Oft zeigt sich in Stunden zu den Kompetenzbereichen Erkenntnisgewinnung, Kommunikation und Bewertung,

#### Problem gelöst?

- Haben Sie den inhalts- und den prozessbezogenen Kompetenzstand der Lernenden analysiert?
- Beschränken Sie sich auf diejenigen Aspekte, die für die Förderung des angestrebten Kompetenzzuwachses relevant sind und somit konkrete Auswirkungen für den Unterrichtsverlauf haben (z.B. methodische Vorerfahrungen)?
- Wird deutlich, inwiefern das Lernziel formulierte Ziel eine Progression zum aktuellen Kompetenzstand darstellt?
- Sind Sie binnendifferenziert vorgegangen, indem Sie beispielsweise den Kompetenzstand der besonders leistungsstarken bzw. -schwachen Lernenden getrennt voneinander analysiert haben?
- Haben Sie sprachliche Probleme der Lernenden, die den vorgesehenen Lernprozess beeinträchtigen könnten, berücksichtigt?

Kasten 5. Leitfragen zur Lerngruppenanalyse (Problem 5)

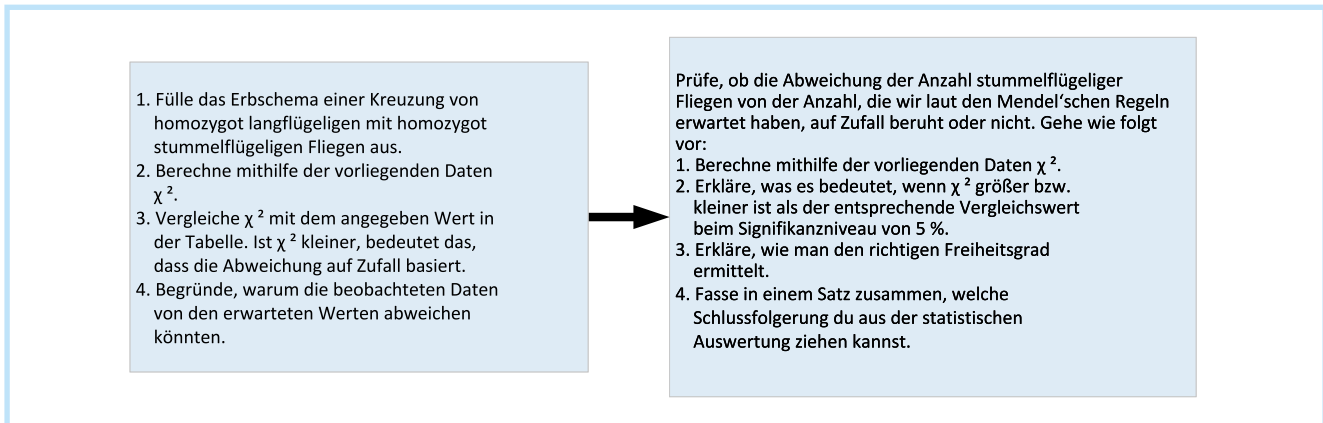


Abb. 7. Aufgabenstellungen in einer Stunde mit folgendem Lernziel: „Die Schüler/innen prüfen das Ergebnis einer Kreuzung von stummelflügeligen Drosophila-Mutanten mithilfe des  $\chi^2$ -Tests.“ Links: Die Aufgaben fokussieren nicht wie im Lernziel vorgesehen die statistische Auswertung (Kompetenzbereich Erkenntnisgewinnung), sondern die phänotypische Aufspaltung nach den Mendelschen Regeln (Kompetenzbereich Fachwissen). Die Besprechung von Aufgabe 4 wird laut Entwurf in der Sicherungsphase fokussiert. Rechts: Die Schüler/innen setzen sich in den Aufgaben ausschließlich mit Fragen der statistischen Auswertung auseinander, in der Sicherung werden alle vier Einzelschritte der Lernaufgabe besprochen und somit das Prinzip des  $\chi^2$ -Tests gesichert.

dass schon in den Aufgabenstellungen bzw. auf dem Arbeitsblatt und dann auch in der Sicherung eher das biologische Fachwissen im Vordergrund steht (Abb. 7, links). Die Aufgaben sollen den angestrebten Kompetenzzuwachs adressieren (Abb. 7, rechts) und dieser dementsprechend auch in der Sicherungsphase im Mittelpunkt stehen. Am Erwartungshorizont lässt sich gut erkennen, ob die angestrebten Lösungen tatsächlich dem Lernziel entsprechen.

#### Problem gelöst?

- Stehen die Aufgaben und der Erwartungshorizont in eindeutigem Zusammenhang zum Lernziel?
- Haben Sie die Gestaltung der Aufgaben (z.B. hinsichtlich ihrer Komplexität, Abstraktheit) begründet?
- Fordern die Aufgaben und das Arbeitsmaterial die Lernenden auf, über Aspekte des ausgewählten Kompetenzbereichs nachzudenken?
- Machen Sie den Lernenden Ihre Erwartungen an das Lernprodukt (Inhalt, Form, Umfang) transparent und nutzen Sie sie in der Sicherungsphase zur Diskussion?
- Erhalten die Lernenden in der Sicherungsphase eine Rückmeldung zu ihrer Kompetenzentwicklung?
- Fokussieren Sie in allen Phasen und Phasenübergängen eindeutig das Lernziel?

Kasten 6. Leitfragen zum Zusammenhang zwischen Phasenstruktur, Aufgaben und Lernziel (Problem 6)

## 4 Darstellungsweise im Unterrichtsentwurf (Probleme 2. Art)

Die letzten beiden Probleme betreffen die Art der Darstellung der Planungsüberlegungen im Unterrichtsentwurf. Dies betrifft Unklarheiten bei der Begründung von Planungsentscheidungen, zum einen durch fehlende oder wenig überzeugende fachdidak-

tische Argumente (Problem 7) und zum anderen durch wenig zielführende Alternativplanungen (Problem 8) als Aspekte legitimierenden Planungshandelns (Abb. 1).

### 4.1 Problem 7: Begründungen für wesentliche Planungsentscheidungen fehlen oder sind oberflächlich.

Unterrichtsentwürfe dienen nicht allein dazu, die geplante Stunde zu beschreiben, sondern die getroffenen didaktischen, methodischen und pädagogischen Planungsentscheidungen auch zu begründen. Oft erschöpft sich die Darstellung dieser Aspekte allerdings in einer mehr oder weniger umfangreichen Beschreibung der Unterrichtsschritte (Abb. 8, links). Nicht selten wird in Unterrichtsentwürfen eine Vielzahl von Unterrichtsbedingungen beschrieben, ohne deren tatsächliche Relevanz für den Unterricht in den Blick zu nehmen. Wichtige Entscheidungen werden entsprechend nicht oder nur oberflächlich begründet. Eine gute Planung würde jedoch erfordern, gute Argumente zu finden, um die geplante Lehr-Lern-Struktur zu legitimieren. Dazu können neben Verweisen auf vorhergehende Kapitel des Entwurfs auch fachdidaktische Quellen helfen (Abb. 8, rechts).

#### Problem gelöst?

- Fokussieren Sie in der Analyse Ihrer Lehr-Lern-Struktur (= methodisch-didaktische Begründung) die wichtigen Aspekte Ihrer geplanten Stunde?
- Nutzen Sie für die Begründung ausgewählter Planungsentscheidungen Aspekte, die Sie in anderen Kapiteln des Entwurfs analysiert haben (Kompetenzentwicklung, Sachstruktur des Themas, Lerngruppenvoraussetzungen)?
- Nutzen Sie allgemein- bzw. fachdidaktische Literatur, um wichtige Planungsentscheidungen zu begründen?

Kasten 7. Leitfragen zur Begründung von Planungsentscheidungen (Problem 7)

In der Erarbeitung sollen die SuS in Form eines Gruppenpuzzles die Schädel von fünf Hominiden (*A. africanus*, *A. boisei*, *H. erectus*, *H. pekinensis*, *H. neanderthalensis*) untersuchen. Zunächst erheben sie Daten zu vorgegebenen Kriterien (z.B. Umfang, Verhältnis Gesichtsschädel zu Stirnhöhe) und notieren die Daten in einer Tabelle. In einer zweiten Phase wechseln die SuS aus den Experten in die Stammgruppen und tauschen sich über ihre Ergebnisse aus. Aus dem Vergleich der Daten sollen die SuS eine begründete Vermutung über die Verwandtschaftsverhältnisse der fünf untersuchten Arten entwickeln.



In der Erarbeitung halte ich das Gruppenpuzzle aus zwei wichtigen Gründen für eine funktionale Lernform. Zum einen sind im Sinne eines schüleraktivierenden Unterrichts (Helmke, 2015) alle SuS involviert und verantwortlich für die Erarbeitung eines Lernproduktes. Zwar können sich in der ersten Phase einige SuS aus den fünf Fünfergruppen zurückhalten und die Beobachtung des jeweiligen Schädels den Gruppenmitgliedern überlassen, doch sind auch sie aufgefordert, die Ergebnisse zu dokumentieren (Gropengießer, 2013) und in der zweiten Phase den anderen SuS zu erklären. Es erfolgt also eine unmittelbare Beteiligung aller SuS. Zum anderen muss somit aus ressourcen- und zeitökonomischen Gründen jeder Schüler nur einen Schädel bearbeiten. Dies stellt einerseits eine Entlastung dar, andererseits wäre im Hinblick auf das Ziel der Stunde nichts gewonnen, wenn die SuS jeweils mehrere Schädel untersuchen würden, da die Kriterien des Vergleichs identisch blieben und somit keine vertiefte Auseinandersetzung mit dieser Arbeitsweise erfolgen würde (Wellnitz & Mayer, 2013). Das Gruppenpuzzle gewährleistet also in kurzer Zeit eine intensive Auseinandersetzung aller SuS mit einem Schädel und bietet aufgrund der Zwischensicherung allen SuS die Möglichkeit, aktiv in der zweiten Erarbeitungsphase mitzuwirken.

Abb. 8. Auszug aus der Analyse der Lehr-Lern-Struktur zu einer Stunde mit folgendem Lernziel: Die Schüler/innen vergleichen Hominidenschädel hinsichtlich ausgewählter Schädelmerkmale und ordnen die Schädel nach Verwandtschaftsnähe. Links: Der Unterrichtsverlauf wird chronologisch beschrieben; Rechts: Der Fokus liegt weniger auf der chronologischen Beschreibung, sondern auf der Begründung wichtiger methodischer (Gruppenpuzzle) sowie didaktischer Entscheidungen (Arbeitsweise Vergleichen) mithilfe allgemein- bzw. fachdidaktischer Literatur.

Alternativ hätte die Erarbeitung der Merkmale des Schweinebandwurms und seines Lebenszyklus statt in Form des Lerntempoduetts auch als Gruppenpuzzle stattfinden können, wobei die als thematische Vertiefung angedachte Erarbeitung der verschiedenen Formen des Parasitismus als weitere Gruppe aufgeführt worden wäre. *Ich habe mich jedoch aus zwei Gründen dagegen entschieden: Zum einen kennt die Lerngruppe Gruppenpuzzle noch nicht, so dass die Einführung der Methode Zeit kosten würde. Außerdem wäre dieser dritte Aspekt deutlich abstrakter als die anderen beiden und würde vom Schweinebandwurm wegführen, so dass es ein Ungleichgewicht zwischen den drei Teilthemen gäbe.*

Abb. 9. Auszug aus der Analyse der Lehr-Lern-Struktur einer Stunde mit folgendem Lernziel: „Die Schüler/innen erarbeiten sich arbeitsteilig den Schweinebandwurm als Parasiten und dessen Lebenszyklus und recherchieren dafür Sachinformationen aus Fachtexten.“ Es wird eine Alternative für eine methodische Entscheidung genannt, die aber erst dadurch die getroffene Planungsentscheidung stärkt, dass sie mit Verweis auf die Vorkenntnisse der Lerngruppe und die Unterrichtsstruktur analysiert und somit begründet verworfen wird (kursiv).

#### 4.2 Problem 8: Es werden keine oder irrelevante Alternativen beschrieben.

Die eigene Argumentation kann auch dadurch gestärkt werden, dass Alternativen in Erwägung gezogen werden und vor dem Hintergrund der Lernvoraussetzungen und der Unterrichtsziele begründet dargestellt wird, warum diese verworfen wurden (Abb. 9). Auch in diesem Zusammenhang ist es ratsam, sich auf allgemein- bzw. fachdidaktische Literatur zu beziehen.

##### Problem gelöst?

- Beschreiben Sie für wichtige didaktische und methodische Planungsentscheidungen mögliche Alternativen, wenn dies zur Begründung Ihrer Entscheidung erforderlich ist?
- Begründen Sie jeweils, warum die angegebene Alternative zugunsten der getroffenen Planungsentscheidung verworfen wurde?

Kasten 8. Leitfragen zu möglichen Alternativen (Problem 8)

#### 5 Abschließende Hinweise

Das Verfassen eines Unterrichtsentwurfs ist eine herausfordernde Aufgabe. Da Unterrichtsentwürfe naturgemäß in Vorbereitung eines Unterrichtsbesuchs entwickelt werden, rücken sie im Rahmen des daran anschließenden Reflexionsgesprächs zumeist zugunsten der Unterrichtsdurchführung in den Hintergrund. Das ist verständlich, bestärkt aber auch den irrigen Eindruck, dass ausführliche Unterrichtsentwürfe eine eigentlich unnötige, zeitaufwändige Arbeit darstellen. Mit den oben dargestellten Beispielen wird versucht, diesem Eindruck entgegenzuwirken.

Unabhängig davon, ob man einen ausführlichen Unterrichtsentwurf schreiben muss oder im Alltag Unterricht plant: Die dargestellten Beispiele können eine Hilfe dabei sein, Unterricht vertieft fachdidaktisch begründet zu planen.

##### Förderhinweis

Das Projekt K2Teach (*Know how to teach*) wird im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und

Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1802 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

## Literatur

Die Quellenangaben finden sich in der Online-Ergänzung.



*LEROY GROSSMANN*, [leroy.grossmann@fu-berlin.de](mailto:leroy.grossmann@fu-berlin.de), hat den Vorbereitungsdienst für das Lehramt an Integrierten Sekundarschulen und

*Gymnasien in den Fächern Biologie und Deutsch in Berlin absolviert und ist wissenschaftlicher Mitarbeiter in der Didaktik der Biologie an der Freien Universität Berlin.*

*ALEXANDER LOTZ ist Lehrer an der Georg-Christoph-Lichtenberg-Schule in Leipzig-Grünau für Biologie und Chemie und war zuvor sechs Jahre lang Fachseminarleiter für Biologie am 4. Schulpraktischen Seminar in Treptow-Köpenick, Berlin.*

*STEFAN MULKE war Lehrer für Biologie und Sport am Rheingau-Gymnasium in Berlin und war 21 Jahre lang Fachseminarleiter für Biologie am 1. Schulpraktischen Seminar in Steglitz-Zehlendorf, Berlin.*

*Prof. Dr. DIRK KRÜGER war Gymnasiallehrer für Biologie und Mathematik und ist Professor für Didaktik der Biologie an der Freien Universität Berlin.* ■□



# Probleme beim Planen?

Acht Vorschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie

//////  
LEROY GROßMANN – ALEXANDER LOTZ – STEFAN MULKE – DIRK KRÜGER  
//////

## Online-Ergänzung

## Literatur

- BAUMERT, J. & KUNTER, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- BLÖMEKE, S., GUSTAFSSON, J.-E. & SHAVELSON, R.J. (2015). Beyond Dichotomies: Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223(1), 3–13.
- DÖBRICH, P. & ABS, H.J. (2008). *Evaluation der zweiten Phase der Lehrerbildung. Pädagogische Entwicklungsbilanzen mit Studienseminaren in Hessen*. Schulverwaltung. Hessen, Rheinland-Pfalz, 13(3), 70–73.
- DORFNER, T., FÖRTSCH, C. & NEUHAUS, B. J. (2020). Use of technical terms in German biology lessons and its effects on students' conceptual learning, *Research in Science & Technological Education*, 38(2), 227–251.
- FUNKE, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- GASSMANN, C. (2013). *Erlebte Aufgabenschwierigkeit bei der Unterrichtsplanung: Eine qualitativ-inhaltsanalytische Studie zu den Praktikumsphasen der universitären Lehrerbildung*. Wiesbaden: Springer.
- GROPENGIESSER, H. & KATTMANN, U. (2013). Didaktische Rekonstruktion. In: H. GROPENGIESSER, U. HARMS & U. KATTMANN (Hg.), *Fachdidaktik Biologie* (S. 18–23). Hallbergmoos: Aulis.
- GROßMANN, L. & KRÜGER, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen, *Erkenntnisweg Biologiedidaktik*, 19, 21–39.
- GROßMANN, L. & KRÜGER, D. (2021). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1). <https://doi.org/10.1007/s40573-022-00141-w>
- HAMMANN, M. & ASSHOFF, R. (2014). *Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten*. Seelze: Klett Kallmeyer.
- HEIDENREICH, T. & GROPENGIESSER, H. (2021). Vorstellungen zum inhaltlichen Teil der Unterrichtsplanung – die Fachliche Klärung als Herausforderung für Lehramtsstudierende. In: S. KAPELARI, A. MÖLLER & P. SCHMIEMANN (Hg.), *Lehr- und Lernforschung in der Biologiedidaktik Bd.9, Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen* (S. 59–76), Innsbruck/Wien: Studienverlag.
- HELMKE, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett Kallmeyer
- KATTMANN, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht*. Hallbergmoos: Aulis.
- KMK (2004). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie* (Beschluss der Kultusministerkonferenz vom 01.12.1989 i. d. F. vom 05.02.2004).
- KMK (2019). *Standards für die Lehrerbildung: Bildungswissenschaften* (Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019).
- KÖNIG, J., BREMERICH-VOS, A., BUCHHOLTZ, C., LAMMERDING, S., STRAUSS, S., FLADUNG, I., et al. (2017). Die Bedeutung des Professionswissens von Referendarinnen und Referendaren mit Fach Deutsch für ihre Planungskompetenz (PlanvoLL-D). In S. WERNKE & K. ZIERER (Hg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?!: Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 121–133). Bad Heilbrunn: Klinkhardt.
- KÖNIG, J., BUCHHOLTZ, C. & DOHMEN, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18(2), 375–404.
- LERSCH, R. (2010). Didaktik und Praxis kompetenzfördernden Unterrichts. In: K. FAULSTICH-CHRIST, R. LERSCH & K. MOEGLING (Hg.), *Kompetenzorientierung in Theorie, Forschung und Praxis* (S. 31–60). Immenhausen: Prolog.
- MAIER, U., BOHL, T., DRÜKE-NOE, C., HOPPE, H., KLEINKNECHT, M. & METZ, K. (2014). Das kognitive Anforderungsniveau von Aufgaben analysieren und modifizieren können: Eine wichtige Fähigkeit von Lehrkräften bei der Planung eines kompetenzorientierten Unterrichts, *Beiträge zur Lehrerinnen- und Lehrerbildung* 32(3), 340–358.
- MEISERT, A. (2016). Eine Langzeitplanung anlegen. In: C. SCHAAPER (Hg.), *Werkzeugkoffer Pädagogisches Handeln. Ein Handbuch für den Start in den Lehreralltag* (S. 85–103). Göttingen: Vandenhoeck & Ruprecht.
- SADAVA, D., HILLIS, D.M., HELLER, H.C. & HACKER, S.D. (2019). *Purves Biologie*. Heidelberg: Springer.
- SenBJF (Senatsverwaltung für Bildung, Jugend und Familie Berlin) (Hg.) (2015). *Rahmenlehrplan Biologie, Teil C, Jahrgangsstufen 7–10*.

- STRIETHOLT, R. & TERHART, E. (2009). Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung. *Zeitschrift für Pädagogik*, 55(4) 622–645.
- VOGELSANG, C. & RIESE, J. (2017). Wann ist eine Unterrichtsplanung ‚gut‘? - Planungsperformanz in Praxisratgebern zur Unterrichtsplanung. In S. WERNKE & K. ZIERER (Hg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?!: Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 47–61). Bad Heilbrunn: Klinkhardt.
- WEINGARTEN, J. (2019). *Wie planen angehende Lehrkräfte ihren Unterricht? Empirische Analysen zur kompetenzorientierten Gestaltung von Lernangeboten*. Münster: Waxmann.
- WELLNITZ, N. & MAYER, J. (2013). Erkenntnismethoden in der Biologie - Entwicklung und Evaluation eines Kompetenzmodells. *ZfdN*, 19, 315–345.
- WERNKE, S. & ZIERER, K. (2017). Die Unterrichtsplanung – Ein in Vergessenheit geratener Kompetenzbereich?! In S. WERNKE & K. ZIERER (Hg.), *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?!: Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 7–16). Bad Heilbrunn: Klinkhardt.
- WIGGINS, G.P. & MCTIGHE, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- WITTWER, J., KRATSCHMAYR, L. & VOSS, T. (2020). Wie gut erkennen Lehrkräfte typische Fehler in der Formulierung von Lernzielen? *Unterrichtswissenschaft*, 48, 113–128.



## Beitrag 10

**Großmann, L.** (2022). Coole Kriminelle? Von ungestraften Regelbrüchen in der Natur. In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 102-108). Berlin: Cornelsen Verlag.

Dieser Beitrag ist Teil eines Sammelbandes, der käuflich erworben werden kann: ISBN 978-3-06-420376-1

## Beitrag 11

**Großmann, L.** (2022). Charles Darwin und Alfred Russel Wallace. In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 116-122). Berlin: Cornelsen Verlag.

Dieser Beitrag ist Teil eines Sammelbandes, der käuflich erworben werden kann: ISBN 978-3-06-420376-1

## Beitrag 12

**Großmann, L.** & Krüger, D. (2022). Guckst du noch oder beobachtest du schon? In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 39-45). Berlin: Cornelsen Verlag.

Dieser Beitrag ist Teil eines Sammelbandes, der käuflich erworben werden kann: ISBN 978-3-06-420376-1

## Beitrag 13

**Großmann, L.,** Vogt, N. & Tietjen, B. (2023). Bietet Schwarmverhalten einen Überlebensvorteil für Fische? Testen von Hypothesen mithilfe eines Modells. *Unterricht Biologie*, Nr. 482, 47(1), S. 5-9.

Dieser Beitrag ist Teil Hefts, das käuflich erworben werden kann:

<https://www.friedrich-verlag.de/shop/modellieren-und-forschen-53482?wkz=20V021>

## Beitrag 14

Vogt, N., **Großmann, L.** & Krüger, D. (angenommen). Logisches Schließen im Biologieunterricht – Wenn Aufgaben evolvieren. *MNU-Journal*.

# Logisches Schließen im Biologieunterricht – Wenn Aufgaben evolvieren

NUBIA VOGT – LEROY GROBMANN – DIRK KRÜGER

*Die Evolutionsbiologie verfügt im Vergleich zu anderen Teildisziplinen der Biologie über eine einzigartige Schlussweise für die Gewinnung neuer Erkenntnisse, die im Unterricht für die Sekundarstufe II jedoch kaum Beachtung findet. Im Beitrag wird basierend auf Ergebnissen einer empirischen Untersuchung die Entwicklung einer Leistungsaufgabe zum historisch-evolutionären Schlussfolgern hin zu einer Übungsaufgabe beschrieben und ein Unterrichtsvorschlag zum Einsatz der Übungsaufgabe vorgestellt.*

## 1 Einleitung

Was unterscheidet einen Evolutionsbiologen wie CHARLES DARWIN von einem Genetiker wie THOMAS HUNT MORGAN oder einem Neuro- und Verhaltensbiologen wie ERIC KANDEL? Auf den ersten Blick sicher ihre Fachgebiete, ihre Untersuchungsobjekte und ihre methodischen Zugänge. Viel wichtiger jedoch: Während MORGAN seine Taufliegen (*Drosophila melanogaster*) und KANDEL seine Meeresschnecken (*Aplysia californica*) heranzüchtete und mit Hilfe von Beobachtungen und Experimenten direkt untersuchen konnten, sind die Untersuchungsobjekte von Evolutionsbiolog/inn/en wie DARWIN meist bereits ausgestorben und viele naturwissenschaftliche Methoden daher nicht anwendbar. Evolutionsbiolog/inn/en bleibt nur eine Möglichkeit, Fragestellungen zu untersuchen: das *historisch-evolutionäre Schlussfolgern*.

Dabei handelt es sich um eine einzigartige Schlussweise in der Biologie, da die Vergangenheit mit Experimenten nicht direkt erfasst werden kann. Es muss also ein rückwärts gerichteter Erklärmodus, die *Abduktion*, genutzt werden, der die Biologie als historische Naturwissenschaft auszeichnet (MAYR, 2004). Das Thema Evolution ist in den Bildungsstandards implementiert und soll über verschiedene Kontexte in den Unterricht eingebracht werden (KMK, 2005, 2020). Bisher liegt der Fokus im Unterricht stark auf der Vermittlung des grundlegenden Fachwissens (OSBORNE, COLLINS, RATCLIFFE, MILLAR, & DUSCHL, 2003), obwohl zu einem wissenschaftspropädeutischen Biologieunterricht auch das abstraktere Nachdenken über historische Erklärungen gehören sollte (Historisch-evolutionäres Schlussfolgern; KIND & OSBORNE, 2017; KROHS, 2004; MAYR, 1988). TROMMLER UND HAMMANN (2020) sprechen hier auch vom Typus ätiologischer Erklärungen, also Erklärungen des Ursprungs oder der Entstehung.

In diesem Beitrag werden die Ergebnisse einer empirischen Untersuchung vorgestellt, in der mithilfe einer Leistungsaufgabe die Vorstellungen erfasst wurden, die Studierende über die Schlussweise von Evolutionsbiolog/inn/en haben. Diese Leistungsaufgabe wurde anschließend in eine Übungsaufgabe überführt und wird in einem Unterrichtsvorschlag genutzt, um mit Schüler/inne/n über das historisch-evolutionäre Schlussfolgern zu reflektieren.

## 2 Historisch-evolutionäres Schlussfolgern

Das *historisch-evolutionäre Schlussfolgern* wird angewandt, um vergangene Ereignisse und deren Entstehung zu erklären und die Gewissheit dieser Erklärung einzuschätzen. Das Besondere an der Evolutionsforschung ist ihr rückwärts gerichteter Erklärmodus (LÜBECK, 2020; SÜBMUTH, 2007). Mit ihm ist die Suche nach einer in der Vergangenheit liegenden

Ursache für ein beobachtetes Phänomen gemeint (JOHNSON & KREMS, 2001; KAMPOURAKIS & NIEBERT, 2018; ROCKSÉN, 2016, TROMMLER & HAMMANN, 2020). Diese Art des Schließens wird *Abduktion* genannt und ist abzugrenzen von den logischen Schlussweisen *Deduktion* und *Induktion*. Ebenso ist das abduktive Erklären abzugrenzen vom Erklären im medialen Sinne, bei dem ein Sachverhalt in didaktischer Vermittlungsabsicht verdeutlicht wird (KRÜGER & UPMEIER ZU BELZEN, 2021). Die Vermittlungsabsicht ist Teil des fachdidaktischen Wissens von Lehrkräften (EHRAS & DITTMER, 2019), die vorwiegend das mediale Erklären im Biologieunterricht umsetzen. Eine Untersuchung der Biologie-Lehrpläne aller deutschen Bundesländer konnte zeigen, dass die Auseinandersetzung mit der evolutionsspezifischen Schlussweise im Unterricht dagegen nicht angestrebt wird (VOGT & KRÜGER, 2022). Durch die explizite Beschäftigung auch mit logischen Schlussweisen könnte jedoch ein verbessertes Verständnis der Lernenden für den historischen Prozess, mit dem in der Biologie Erkenntnisse gewonnen werden, erreicht werden (FURTA, SEIDEL, IVERSON, & BRIGGS, 2012).

Die drei logischen Schlussweisen Induktion, Deduktion und Abduktion lassen sich wie folgt unterscheiden: Bei der Induktion wird aus zahlreichen Einzelbeobachtungen eine verallgemeinernde Regel (Abb. 1, (c)) entwickelt, die sich auf vergangene, aktuelle und zukünftige Beobachtungen anwenden lässt. Induktives Schlussfolgern ist damit erkenntniserweiternd, aber nur in dem Sinne, dass eine bereits gemachte Beobachtung auf eine größere Menge an potentiellen Beobachtungen erweitert wird. Die Induktion ist logisch nicht einwandfrei, denn diese Erweiterung kann fehlerhaft sein, wenn ein unerwarteter Fund die induktiv entwickelte Regel widerlegt (LANGLET, 2016). Bei der Deduktion wird aus einer allgemeinen Regel und einem beobachteten Ereignis ein Ergebnis abgeleitet (Abb. 1, (b)). Ist die Regel wahr, muss auch das erwartete Ergebnis eintreten. Deduktives Schlussfolgern ist damit logisch einwandfrei (LANGLET, 2016), ist jedoch auch nicht erkenntniserweiternd, solange die bestehende Regel bestätigt wird. Abduktion wird bei der rückwärts gerichteten Suche nach Ursachen im Evolutionskontext genutzt. Hier wird ein Phänomen, zum Beispiel ein unerwarteter Knochenfund, entweder mit Hilfe von bekanntem Regelwissen oder kreativ durch Schaffung neuer Regeln durch ein vorhergehendes Ereignis erklärt (Abb. 1, (a); KRÜGER & UPMEIER ZU BELZEN, 2021). In dem Moment, wo eine Ursache kreativ gefunden wird, wird neues Regelwissen aufgestellt. Damit ist Abduktion eine Schlussweise, die neues Wissen in die Naturwissenschaft bringen kann.

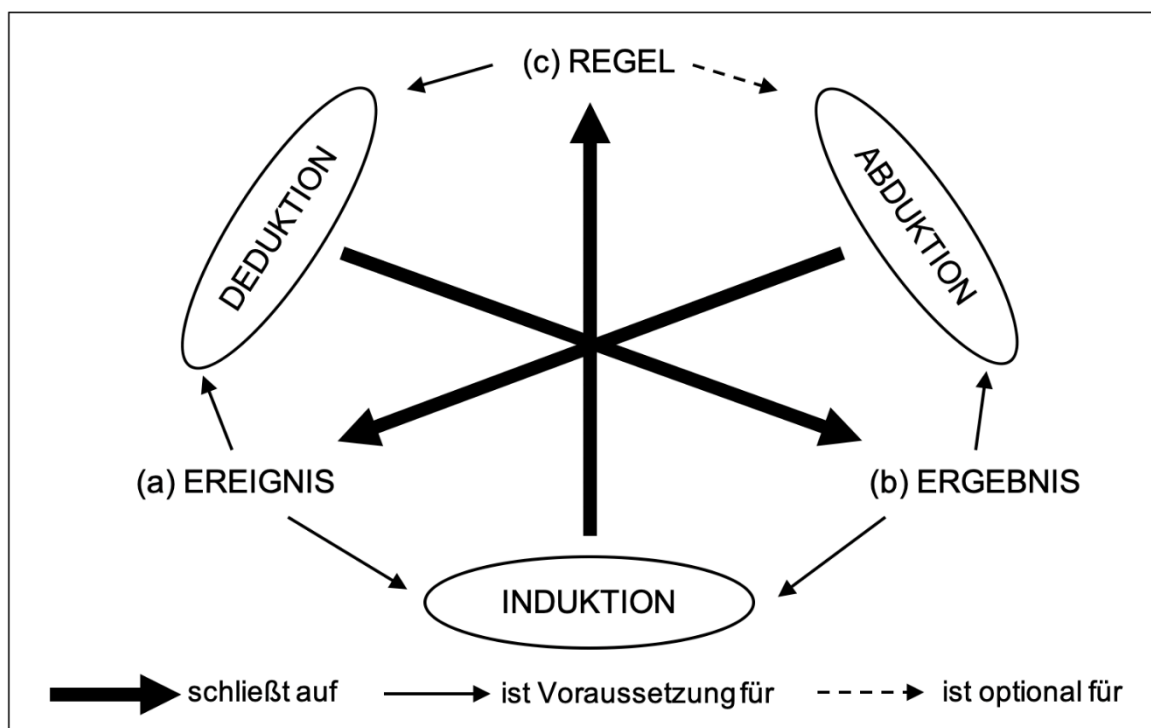


Abb. 1. Zusammenhang der drei logischen Schlussweisen (vgl. KRÜGER UND UPMEIER ZU BELZEN, 2021)

In der Evolutionsforschung können keine Vorhersagen über Ereignisse in der Zukunft getroffen werden (Prädikt), sondern nur Aussagen über historische Ereignisse (Retrodikt; KATTMANN, 2009; OH, 2019). Im Evolutionskontext kommt dem Begriff ‚erklären‘ dann eine erkenntniserweiternde Funktion zu, wenn beim abduktiven Schließen die bestmögliche Erklärung für ein Phänomen nicht nur durch Rückgriff auf bestehende theoretische Erwägungen, sondern durch kreative Erweiterungen gefunden wird (JOHNSON & KREMS, 2001).

### 3 Empirische Untersuchung

Im Vorfeld der hier vorgestellten Untersuchung wurde auf Basis des theoretischen Vorschlags zur Strukturierung von *scientific reasoning* von KIND UND OSBORNE (2017) ein Multiple-Choice (MC)-Testinstrument entwickelt, das die Kompetenzen in verschiedenen Arbeits- und Denkweisen (*styles of scientific reasoning*) des Kompetenzbereichs Erkenntnisgewinnung in einzelnen Aufgaben abbilden und so zugänglich machen sollte. Mithilfe der Aufgaben sollte untersucht werden, inwiefern Kompetenzen bezüglich der *styles* voneinander abzugrenzen sind (VOGT & KRÜGER, 2020). Die Aufgaben wurden systematisch konstruiert (TERZER, HARTIG, & UPMEIER ZU BELZEN, 2013) und der Prozess wurde mit Untersuchungen zur Sicherung der Validität der Testwertinterpretationen begleitet (AERA, APA, & NCME, 2014; VOGT & KRÜGER, 2020; Abb. 2).

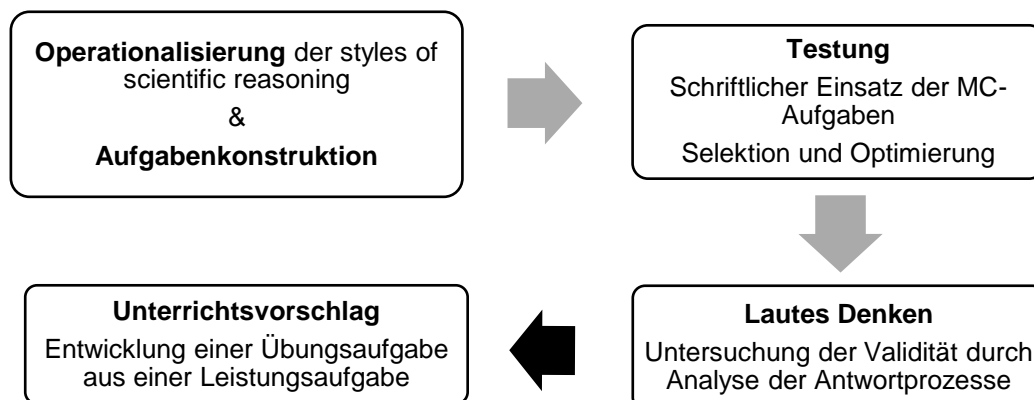


Abb. 2. Prozess der Aufgabenentwicklung und -nutzung (verändert nach VOGT & KRÜGER, 2020)

Eine der Aufgaben zum historisch-evolutionären Schlussfolgern bezieht sich auf die Abduktion. Der Kontext der Aufgabe thematisiert das Aussterben der Panzerfische am Ende des Devons (Box 1). In den drei Antwortmöglichkeiten werden die logischen Schlussweisen Deduktion (A), Induktion (B) und Abduktion (C) angeboten. Um zu vermeiden, dass Antwortmöglichkeit C über sprachliche Attribute als richtige Antwort erkannt wird, wurde hier mit einem breit gefassten Hypothesenbegriff gearbeitet.



Im Zeitalter des Devons entwickelten sich in gewaltigen Korallenriffen wärmeabhängige Panzerfische. Den Forschenden der marinen Ökologie ist bekannt, dass die Panzerfische am Ende des Devons ausgestorben sind. Eine Regel besagt, dass ein Temperaturabfall zum Sterben wärmeabhängiger Arten führt. Es besteht die Hypothese, dass ein Temperaturabfall durch die globale Abkühlung ursächlich das Aussterben der Panzerfische bewirkt hat.

Welchen logischen Schluss führten die Forschenden beim Aufstellen der Hypothese durch?

A) Sie schließen auf Basis der Regel und aus der Ursache eines Temperaturabfalls, dass dies das Aussterben der wärmeabhängigen Panzerfische bewirkt hat.

B) Sie schließen aus der Ursache eines Temperaturabfalls und der Wirkung des Aussterbens der wärmeabhängigen Panzerfische auf die Regel.

C) Sie schließen aus dem Ergebnis des Aussterbens der wärmeabhängigen Panzerfische und auf Basis der Regel, dass als Ursache dafür ein Temperaturabfall stattgefunden hat.

#### Box 1. Aufgabe zum historisch-evolutionären Schlussfolgern

Das Ergebnis der schriftlichen Testung mit Lehramtsstudierenden der Biologie ( $N = 57$ ) zeigte, dass die Lernenden überwiegend die Deduktion (Antwortmöglichkeit A; 51 %) als die von den Forschenden angewandte logische Schlussweise wählten. Die Abduktion (C) sahen 39 % als die richtige Schlussweise an. Nur 10 % entschieden sich für die Induktion (B). Aufgrund der vergleichsweise unwahrscheinlichen Antwortmöglichkeit B kann es möglich sein, dass die restliche Verteilung der Antworten durch eine hohe Ratefrequenz zustande kommt. Daher wurden Validierungsuntersuchungen ergriffen, um das Ergebnis der schriftlichen Testung besser interpretieren zu können.

Eine Validierungsmöglichkeit ist die Untersuchung der Antwortprozesse bei der Lösung der Aufgabe, die Lehramtsstudierende während des lauten Denkens verbalisieren (SANDMANN, 2014). Dabei wurden in diesem Fall die Vorstellungen zum logischen Schließen in der Evolutionsbiologie bei Lehramtsstudierenden ( $N = 6$ ) erfasst. Die qualitativen Daten aus dem lauten Denken sollten sowohl Hinweise für die Verständlichkeit der Aufgabe liefern, als auch Vorstellungen der Studierenden zur Abduktion. Das Ergebnis des lauten Denkens bestätigt die Tendenz, die Induktion (B) als genutzte Schlussweise auszuschließen. Die anderen Antworten verteilen sich auf Antwortmöglichkeiten A ( $n = 2$ ) und C ( $n = 4$ ), wobei zwei Personen, die sich für die Abduktion entschieden, äußerten, dass sie geraten haben. Es wird deutlich, dass die Aufgabe für viele Studierende unverständlich erscheint: „Ich bin mir gerade unsicher zwischen a) und c). Ja, ich hab hier keine Ahnung, ich würde jetzt erstmal raten“ (Person 2). Nur eine Person konnte ihre Entscheidung für die Abduktion als Schlussweise angemessen artikulieren: „Da [als] Ergebnis des Ganzen ja die Panzerfische ausgestorben sind, [...] und auf Basis der Regel, dass der Temperaturabfall zum Sterben wärmeabhängiger Arten führt, haben sie dann als Ursache dafür [angenommen], dass ein Temperaturabfall stattgefunden haben muss“ (Person 4). Die anderen Studierenden konnten keine angemessene Begründung für ihre Entscheidung liefern.

## 4 Diskussion und Implikationen für die Praxis

Die Ergebnisse aus dem lauten Denken zeigen Schwierigkeiten auf, die Studierende mit logischen Schlussweisen haben. Das Verständnis sowie der anschließende Ausschluss der Induktion als Schlussweise in der Evolutionsbiologie scheinen möglich, die Unterscheidung zwischen Deduktion und Abduktion hingegen schwieriger. Die Studierenden entschieden sich in der schriftlichen Testung häufiger für die Deduktion, was in der verstärkten Nennung und Übung des hypothetisch-deduktiven Erkenntniswegs im Unterricht begründet sein könnte (GROPENIEBER, HARMS, & KATTMANN, 2016). Materialien für den Evolutionsunterricht beziehen sich zudem häufig auf Fachinhalte wie die verschiedenen Evolutionstheorien oder Verwandtschaftsanalysen (z.B. KLEMMSTEIN, 2022; SCHAAL & SCHIMPF, 2020), und fördern

kein Nachdenken über die Schlussweise in der Evolutionsbiologie, obwohl seit Jahrzehnten die Forderung nach einer historischen Ausrichtung des Biologieunterrichts besteht (KATTMANN, 1995), die es explizit zu machen gilt (KMK, 2005, 2020). Die hier vorgestellte MC-Aufgabe wurde als Grundlage für die Entwicklung einer Übungsaufgabe für den Evolutionsunterricht genutzt, die die Ausbildung eines Verständnisses der Abduktion fördern soll. Durch ihre eher offene Strukturierung kann sie zum kreativen Nachdenken und Diskutieren anregen und zum langfristigen Wissenserwerb beitragen (lernförderliche Funktion; ABRAHAM & MÜLLER, 2009).

Wir empfehlen, die Schüler/innen in Einzelarbeit zunächst den Informationstext „Bäume im Herbst“ (Box 2) lesen zu lassen und ihnen Zeit zu geben, die Hilfekarten mit den Erläuterungen zur Induktion, Deduktion und Abduktion (Box 3 - 5) zu verstehen. Anschließend können die Aufgaben, auch in Partnerarbeit, bearbeitet werden. Die Aufgaben können als Kopiervorlage genutzt werden, wenn sie in schwarz-weiß kopiert werden, sodass die farbliche Kennzeichnung (Erwartungshorizont für Aufgabe 1) entfällt.

### Bäume im Herbst

Eine Person beobachtet über mehrere Jahre hinweg im Park vor ihrem Haus, dass im Herbst (vorhergehendes Ereignis) die dort stehenden Bäume ihre Blätter verlieren (nachfolgendes Ergebnis). Sie schlussfolgert daraus, dass zur Herbstzeit alle Bäume ihre Blätter verlieren (Regel). Die Person hat also die wiederkehrende Beobachtung des Ereignisses (Herbst) in Kombination mit dem Ergebnis (Bäume verlieren ihre Blätter) zu einer Regel verallgemeinert. Diese Art des Schlussfolgerns nennt man *Induktion*.

Eine andere Person nutzt die Regel und möchte diese überprüfen. Sie leitet aus der Regel diese Hypothese ab: Wenn es Herbst wird (vorhergehendes Ereignis), dann verlieren alle Bäume im Mischwald vor meinem Haus ihre Blätter (nachfolgendes Ergebnis). Im Herbst stellt die Person fest, dass die Nadelbäume weiterhin grün bleiben. Die Person hat ihre Hypothese widerlegt, also führt deduktives Schließen in diesem Fall dazu, dass die Regel geändert werden muss: Wenn es Herbst wird (Ereignis), dann verlieren Laubbäume ihre Blätter (Ergebnis). Die Art des Schlussfolgerns nennt man *Deduktion*.

Eine weitere Person gräbt einen versteinerten Baum aus, an dem keine Blätter erkennbar sind (Ergebnis). Sie zieht zur Erklärung des Fundes die umformulierte Regel heran, dass im Herbst Laubbäume ihre Blätter verlieren. Die Person schlussfolgert, dass der versteinerte Baum ein Laubbaum sein könnte (Ereignis), der im Herbst seine Blätter verloren hatte. Die Person hat also eine bestehende Regel auf ein beobachtetes Ergebnis (Baum ohne Blätter) angewendet und damit eine Erklärung gefunden (Ereignis), die die fehlenden Blätter des versteinerten Baumes (Ergebnis) erklärt. Diese Erklärung ist nur eine mögliche Erklärung, da das Ereignis in der Vergangenheit liegt und sich daher nicht abschließend geprüft werden kann. Es wäre ebenso denkbar, dass der versteinerte Baum aufgrund eines Schädlingsbefalls oder einer vorhergehenden Dürre keine Blätter trägt. Diese Art des Schlussfolgerns nennt man *Abduktion*.

Wenn Sie mehr über die drei logischen Schlussweisen erfahren möchten, schauen Sie sich die Hilfekarten an!

Box 2. Bäume im Herbst

### Induktion: Das Prinzip, Regeln zu entdecken

Die Induktion ist eine Methode des logischen Schließens, bei der von einem speziellen Fall auf eine allgemeine **Regel** geschlossen wird. Ein Fall besteht aus der Beobachtung, dass ein **Ereignis** zu einem bestimmten **Ergebnis** führt. Beobachtet man entsprechende Fälle öfter, leitet man eine allgemeine **Regel** daraus ab. Bei der Induktion geht es darum, eine allgemeingültige **Regel** zu entdecken. Die entdeckte Regel ist wahrscheinlich, aber nicht sicher gültig. Selbst eine hohe Anzahl beobachteter gleicher Fälle bietet keine Sicherheit, dass alle folgenden Fälle der entdeckten Regel entsprechen.

#### Zeitachse

**1. Wenn Ereignis → 2. Dann Ergebnis → 3. allgemeine Regel**

### Box 3. Hilfekarte Induktion

### Deduktion: Das Prinzip, Regeln zu untersuchen

Die Deduktion ist eine Methode des logischen Schließens, bei der mit Hilfe einer allgemeinen **Regel** auf einen speziellen Fall geschlossen wird. Dazu formuliert man auf Basis der **Regel** Hypothesen. Diese Hypothesen beschreiben einen speziellen Fall: Ein beobachtetes **Ereignis** führt laut **Regel** zu einem konkreten **Ergebnis**. Der Fall muss anschließend überprüft werden. Bei der Deduktion geht es darum, die **Regel** zu bestätigen oder zu widerlegen. Ist die Regel wahr, führt das Ereignis zum vorhergesagten Ergebnis.

#### Zeitachse

**1. Es gibt Regel → 2. Wenn Ereignis → 3. vorausgesagtes Ergebnis**

### Box 4. Hilfekarte Deduktion

### Abduktion: Das Prinzip, Erklärungen für Phänomene zu finden

Die Abduktion ist eine Methode des logischen Schließens, bei der für ein **Ergebnis**, man spricht dann auch von einem **Phänomen**, eine Erklärung gesucht wird. Die Erklärung liefert dann das **Ereignis**, dass zum **Phänomen** führt. Dafür sucht man nach **Regeln**, die das beobachtete **Phänomen** erklären helfen. Findet man keine **Regel** zur Erklärung des **Phänomens**, muss durch kreative Ideen eine neue **Regel** für eine Erklärung gefunden werden. Bei der Abduktion geht es darum, Erklärungen für Phänomene zu finden, was zu einer neuen **Regel** führen kann. Das Besondere an der Abduktion ist, dass beim Erklären Aussagen über Ereignisse getroffen werden, die in der Vergangenheit liegen. Die gefundene Erklärung für das Phänomen ist möglich, aber nicht sicher. Alternative Erklärungen können ebenso zutreffend sein.

#### Zeitachse

**3. Erkläre mit Ereignis → 1. Es taucht Phänomen auf → 2. Regel**

### Box 5. Hilfekarte Abduktion

Die aus der ursprünglichen Leistungsaufgabe entwickelte, hier vorgestellte Übungsaufgabe kann gezielt das Verständnis über einen epistemischen Aspekt des historisch-evolutionären Schlussfolgerns - die Abduktion - fördern, indem sie weniger auf die korrekte Lösung der eingebetteten MC-Aufgabe abzielt als eher auf das Verständnis und die reflektierte Betrachtung der Schlussweise von Evolutionsbiolog/inn/en.

### Aufgaben:

1. Entscheiden Sie jeweils für die folgenden Beispiele, ob es sich um eine Induktion, Deduktion oder Abduktion handelt. Nutzen Sie dafür die Informationen auf den Hilfefkarten, indem Sie in jedem Beispiel die drei Bestandteile (Regel, Ereignis, Ergebnis) des logischen Schlusses farblich markieren.

#### Beispiel 1

Eine Forscherin vergleicht den **Körperbau von 1000 Fliegen** und stellt fest, dass **150 Fliegen Ähnlichkeiten** aufweisen. Sie vermutet, dass diese 150 Fliegen derselben Art zuzuordnen sind. Zur Bestätigung sammelt sie weitere Fliegen, von denen sie vermutet, dass sie der Beschreibung der Fliegenart entsprechen. Die zusätzlichen Fliegen haben den erwarteten Körperbau. Sie formuliert daraufhin eine **Regel zur Beschreibung des Körperbaus** bei dieser Fliegenart.

#### Beispiel 2

Forschende haben in den vergangenen Jahrzehnten Kriterien aufgestellt, mit Hilfe derer man **Zellen unter dem Mikroskop als Nervenzellen identifizieren** kann. Sehen sie im Präparat nun eine **Zelle, die diese vorher definierten Merkmale aufweist**, können sie daraus schließen, **dass es sich um eine Nervenzelle handelt**.

#### Beispiel 3

Forschende haben einen Schädel eines frühen Huftieres mit intaktem Gebiss gefunden. **Die Zähne sind flach, tief im Kiefer verankert und zeigen deutliche Kratzspuren**. Die Forschenden wissen, **dass Pflanzenfresser flache Mahlzähne haben und dass die Kieselsäure, die in Gräsern enthalten ist, den Zahnschmelz derartig abschleifen kann**. Sie vermuten daher, dass sich das **Huftier von pflanzlicher Kost mit Grasanteilen ernährt** hat.

#### Beispiel 4

Forschende sind sich **einig, dass das Säugen des Nachwuchses mit Milch und ein Fell Merkmale für Säugetiere sind**. Betrachtet man ein **Tier aus der Klasse der Säugetiere**, muss es demnach seinen **Nachwuchs säugen und ein Fell aus Haaren haben**.

### Box 6. Beispiele für logisches Schließen

2. a) Entscheiden Sie für das folgende Beispiel, welchen logischen Schluss die Forschenden beim Aufstellen der Vermutung durchführten (A, B oder C).  
b) Ordnen Sie den drei Antwortmöglichkeiten die logischen Schlussweisen Induktion, Deduktion und Abduktion zu.

Im Zeitalter des Devons entwickelten sich in gewaltigen Korallenriffen wärmeabhängige Panzerfische, die am Ende des Devons jedoch ausgestorben sind.

Eine Regel besagt, dass ein Temperaturabfall zum Sterben wärmeabhängiger Arten führt. Die Forschenden schließen daher, dass ein Temperaturabfall durch die globale Abkühlung das Aussterben der Panzerfische bewirkt hat.

- 1) Sie schließen auf Basis der Regel und aus der Ursache eines Temperaturabfalls, dass dies das Aussterben der wärmeabhängigen Panzerfische bewirkt hat.
- 2) Sie schließen aus dem Ergebnis des Aussterbens der wärmeabhängigen Panzerfische und auf Basis der Regel, dass als Ursache dafür ein Temperaturabfall stattgefunden hat.
- 3) Sie schließen aus der Ursache eines Temperaturabfalls und dem Ergebnis des Aussterbens der wärmeabhängigen Panzerfische auf die Regel, dass ein Temperaturabfall zum Sterben wärmeabhängiger Arten führt.

### Box 7. Aufgaben

3. Formulieren Sie das folgende Szenario (Box 8) für die Abduktion um, vergleichbar mit dem Informationstext „Bäume im Herbst“ (Box 2).

### Schneehasen im Winter

Induktion:

Eine Person, die in den Alpen lebt, beobachtet über mehrere Jahre hinweg, dass sich bei den dort lebenden Hasen das Fell von braun zu weiß färbt, wenn der Winter beginnt. Sie schlussfolgert daraus, dass zum Beginn des Winters alle Hasen ein weißes Fell bekommen.

Deduktion:

Eine andere Person lebt an einem Wald und möchte die Regel testen. Sie erwartet, dass zu Beginn des Winters Hasen im Wald ihre Fellfarbe wechseln. Im Winter beobachtet die Person jedoch nur Feldhasen mit durchgehend braunem Fell. Sie formuliert die Regel um, dass nur Schneehasen einen Fellwechsel vollziehen.

Abduktion:

Eine weitere Person beobachtet im Januar bei einer Alpenwanderung einen weißen Hasen (...).

Box 8. Aufgabe Schneehase

*Lösungen:*

1. 1) Induktion, 2) Deduktion, 3) Abduktion, 4) Deduktion
2. a) Abduktion (B)  
b) 1) Deduktion, 2) Abduktion, 3) Induktion
3. Eine weitere Person beobachtet im Januar bei einer Alpenwanderung einen weißen Hasen. Sie erklärt sich mithilfe der neuen Regel, dass der beobachtete Hase ein Schneehase sein könnte, der zu Beginn des Winters seine Fellfarbe geändert hat. Diese Erklärung ist jedoch nur eine mögliche Erklärung, denkbar wäre zum Beispiel auch ein Albino-Hase mit Gendefekt, der keine Farbpigmente herstellen kann, oder eine weitere Hasenart, die das ganze Jahr hindurch ein weißes Fell hat.

## Literatur

ABRAHAM, U., & MÜLLER, A. (2009). Aus Leistungsaufgaben lernen. *Praxis Deutsch: Zeitschrift für den Deutschunterricht*, 36(214), 4-12.

AERA, APA, & NCME (Hg.). (2014). *Standards for educational and psychological testing. National Council on Measurement in Education*. Washington, DC: American Educational Research Association.

EHRAS, C., & DITTMER, A. (2019). Kennzeichen guten Erklärens im Biologieunterricht - Wie Schüler/innen die Erklärung komplexer biologischer Phänomene wahrnehmen. *Erkenntnisweg Biologiedidaktik* 17, 79-94.

FURTAK, E. M., SEIDEL, T., IVERSON, H., & BRIGGS, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of educational research*, 82(3), 300-329.

- GROPENGBIEBER, H., HARMS, U., & KATTMANN, U. (2016). *Fachdidaktik Biologie*. Halbergmoos: Aulis.
- JOHNSON, T., & KREMS, J. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive Science*, 25, 903-939.
- KAMPOURAKIS, K., & NIEBERT, K. (2018). Explanation in Biology Education. In K. KAMPOURAKIS & M. J. REISS (Hg.), *Teaching Biology in Schools: Global Research, Issues, and Trends* (S. 237–248). New York, USA: CRC Press.
- KATTMANN, U. (1995). Konzeption eines naturgeschichtlichen Biologieunterrichts: Wie Evolution Sinn macht. *Zeitschrift für Didaktik der Naturwissenschaften*, 1(1), 29–42.
- KATTMANN, U. (2009). *Evolution und Schöpfung – Kreationismus als Herausforderung für den Biologieunterricht*. Paper presented at the Bundeskongress der MNU, Regensburg.
- KIND, P., & OSBORNE, J. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, 101(1), 8-31.
- KLEMMSTEIN, W. (2022). Die Evolution des Menschen entdecken. *Unterricht Biologie Kompakt*(476), 3-46.
- KMK. (2005). *Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München, Neuwied: Luchterhand.
- KMK. (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife. Beschluss der Kultusministerkonferenz vom 18.06.2020*. München, Neuwied: Luchterhand.
- KROHS, U. (2004). Die Struktur biologischer Theorien. In U. KROHS (Hg.), *Eine Theorie biologischer Theorien: Status und Gehalt von Funktionsaussagen und informationstheoretischen Modellen* (163-198). Berlin: Springer.
- KRÜGER, D., & UPMEIER ZU BELZEN, A. (2021). Kompetenzmodell der Modellierkompetenz – Die Rolle abduktiven Schließens beim Modellieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 27(1), 127-137.
- LANGLET, J. (2016). Kultur der Naturwissenschaften. In H. GROPENGBIEBER, U. HARMS, & U. KATTMANN (Hg.), *Fachdidaktik Biologie* (S. 80-97). Halbergmoos: Aulis.
- LÜBECK, M. (2020). *'Basiskonzepte' der Erkenntnisgewinnung im Biologieunterricht: Ein Nachschlagewerk mit Aufgabenbeispielen*. Münster: Waxmann Verlag.
- MAYR, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*. Cambridge: Harvard University Press.
- MAYR, E. (2004). *What makes biology unique?: Considerations on the autonomy of a scientific discipline*. Cambridge: Cambridge University Press.
- OH, P. (2019). Features of Modeling-Based Abductive Reasoning as a Disciplinary Practice of Inquiry in Earth Science. Cases of Novice Students Solving a Geological Problem. *Science & Education*, 28, 731-757.
- OSBORNE, J., COLLINS, S., RATCLIFFE, M., MILLAR, R., & DUSCHL, R. (2003). What “ideas-about-science” should be taught in school science? A delphi study of the expert community. *Journal of Research in Science Teaching*, 40(7), 692-720.
- ROCKSÉN, M. (2016). The many roles of “explanation” in science education: a case study. *Cultural Studies of Science Education*, 11.

SANDMANN, A. (2014). Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. KRÜGER, I. PARCHMANN, & H. SCHECKER (Hg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179-188). Berlin, Heidelberg: Springer Spektrum.

SCHAAL, S., & SCHIMPF, F. (2020). Evolution der Honigbiene. Honigbieren und Wildbienen auf der Spur: *Unterricht Biologie kompakt*, 454, 12-17.

SÜBMUTH, R. (2007). Die Evolutionstheorie, ihre Bedeutung und ihre Grenzen. *Imago Hominis*, 14(1), 13 – 45.

TERZER, E., HARTIG, J., & UPMEIER ZU BELZEN, A. (2013). Systematische Konstruktion eines Tests zu Modellkompetenz im Biologieunterricht unter Berücksichtigung von Gütekriterien. *Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)*, 19, 51-76.

TROMMLER, F. & HAMMANN, M. (2020). The relationship between biological function and teleology: Implications for biology education. *Evolution: Education & Outreach*, 13:11.

VOGT, N., & KRÜGER, D. (2020). Styles of Scientific Reasoning - Entwicklung eines Testinstruments zur Erfassung prozeduralen und epistemischen Wissens. *Erkenntnisweg Biologiedidaktik*, 19, 41-56.

VOGT, N., & KRÜGER, D. (2022). Vergleichende Analyse von scientific reasoning-Aspekten in Biologie-Lehrplänen für die Sekundarstufe II. *Zeitschrift für Didaktik der Biologie (ZDB) - Biologie Lehren und Lernen*, 26, 36-53.

*NUBIA VOGT ist Biologin (M. Sc.) und war bis Oktober 2022 wissenschaftliche Mitarbeiterin in der Didaktik der Biologie an der Freien Universität Berlin (n.vogt@fu-berlin.de; Schwendenerstraße 1, 14195 Berlin).*

*LEROY GROßMANN hat den Vorbereitungsdienst für das Lehramt an Integrierten Sekundarschulen und Gymnasien in den Fächern Biologie und Deutsch in Berlin absolviert und ist wissenschaftlicher Mitarbeiter in der Didaktik der Biologie an der Freien Universität Berlin.*

*Prof. DR. DIRK KRÜGER war Gymnasiallehrer für Biologie und Mathematik und ist Professor für Didaktik der Biologie an der Freien Universität Berlin.*

□

## Lebenslauf

### Persönliche Daten

---

Name: Leroy Großmann  
Nationalität: Deutsch

### Ausbildung

---

- seit 06/2019 **Wissenschaftlicher Mitarbeiter (Praedoc) an der Freien Universität Berlin**  
Fachbereich Biologie, Chemie, Pharmazie – Institut für Biologie – Didaktik der Biologie  
(AG Prof. Dr. Dirk Krüger)  
voraussichtlicher Abschluss: Dr. rer. nat.
- 02/2018 – 06/2019 **Vorbereitungsdienst für das Lehramt an Gymnasien**  
in den Fächern Deutsch und Biologie am Gottfried-Keller-Gymnasium (Berlin)  
Abschluss: Zweites Staatsexamen
- 10/2015 – 09/2017 **Master-Studium der Deutschen Philologie und Biologie (Lehramt)**  
an der Freien Universität Berlin  
Abschluss: Master of Education (M. Ed.)
- 04/2015 – 09/2017 **Studentische Hilfskraft an der Freien Universität Berlin**  
Fachbereich Biologie, Chemie, Pharmazie – Institut für Biologie – Didaktik der Biologie  
(AG Prof. Dr. Dirk Krüger)
- 10/2012 – 09/2015 **Bachelor-Studium der Deutschen Philologie und Biologie (Lehramt)**  
an der Freien Universität Berlin  
Abschluss: Bachelor of Arts (B. A.)
- 10/2011 – 09/2012 **Bachelor-Studium der Deutschen Philologie und Chemie (Lehramt)**  
an der Humboldt-Universität zu Berlin
- 09/2002 – 06/2011 **Abitur am Gymnasium Wellingdorf (Kiel)**  
Gesamtnote: Sehr gut (1,5)

### Praktische Erfahrungen

---

- 10/2017 – 12/2017 **Praktikum am Institut für Schulqualität der Länder Berlin und Brandenburg e.V. (ISQ)**  
Entwicklung und Durchführung einer Hotliner-Schulung zum Selbstevaluationsportal
- 10/2016 – 03/2017 **Teilnahme am Projektseminar „Let’s Play Schule“ (Kreidestaub e.V.)**  
mit einwöchiger Durchführung einer Projektwoche an der Neuen Mittelschule Langenzersdorf bei Wien (Österreich) [ausgezeichnet mit dem Hans-Sauer-Preis 2017]
- 04/2016 – 09/2016 **Leitung des Projektseminars „Prinzip Lernreise“ (Kreidestaub e.V.)**
- 10/2015 – 03/2016 **Teilnahme am Projektseminar „Prinzip Lernreise“ (Kreidestaub e.V.)**



### Lehrveranstaltungen

---

WiSe 2023/24	Schulpraktikum (Master) [LV-Nr. 23601e] Begleit-/Nachbereitungsseminar (Master) [LV-Nr. 23602b] Begleit-/Nachbereitungsseminar (Master) [LV-Nr. 23602c]
SoSe 2023	Vorbereitungsseminar Praxissemester (Master) [LV-Nr.23600c] Ausgewählte Themen des Biologieunterrichts: Lernaufgaben (Master; gemeinsam mit Dirk Krüger) [LV-Nr.23610c]
WiSe 2022/23	Schulpraktikum (Master) [LV-Nr. 23601e]
SoSe 2022	Vorbereitungsseminar Praxissemester (Master) [LV-Nr. 23600d]
WiSe 2021/2022	Schulpraktikum (Master) [LV-Nr. 23601e]
SoSe 2021	Ausgewählte Themen des Biologieunterrichts: Mysterys (Master; gemeinsam mit Dirk Krüger) [LV-Nr.23610a]
WiSe 2020/2021	Schulpraktikum (Master) [LV-Nr.23611e]
SoSe 2020	Vorbereitungsseminar Praxissemester (Master) [LV-Nr.23600c] Ausgewählte Themen des Biologieunterrichts: Mysterys (Master; gemeinsam mit Dirk Krüger) [LV-Nr.23610a]
WiSe 2019/2020	Schulpraktikum (Master) [LV-Nr.23611g]

### Monographie

---

**Großmann, L.** (2019). *Modellverstehen und Modellieren an einer Blackbox. Eine Videoanalyse aus der biologiedidaktischen Forschung*, BestMasters, Springer Spektrum, Wiesbaden.

### Veröffentlichungen in Zeitschriften und Sammelbänden (mit Review-Verfahren)

---

- [11] **Großmann, L.**, Koberstein-Schwarz, M., Scholl, D., Meisert, A. & Krüger, D. (eingereicht). Forschungsfelder kartieren, Forschungslücken entdecken: Bibliometrische Analysen am Beispiel der naturwissenschaftsdidaktischen Unterrichtsplanungsforschung.
- [10] Lieberei, T., **Großmann, L.**, Dawborn-Gundlach, M., Krüger, D. & Krell, M. (eingereicht). How Considering Multiple Sources of Validity Evidence Can Help to Address Challenges in the Development of Pedagogical Content Knowledge (PCK) Multiple-Choice Items.
- [9] Welter, V., Dawborn-Gundlach, M., **Großmann, L.** & Krell, M. (eingereicht). Adapting a Self-Efficacy Scale to the Task of Teaching Scientific Reasoning: Collecting Evidence for its Psychometric Quality Using Rasch Measurement.
- [8] **Großmann, L.\***, Koberstein-Schwarz, M., Scholl, D., Krüger, D. & Meisert, A. (eingereicht). Establishing common ground in empirical research on science teachers' lesson planning – A scoping review. [\* geteilte Erstautorenschaft]

- [7] **Großmann, L.** & Krüger, D. (in Druck). Erkenntnisgewinnung (v)erklärt. Wie plant man hypothesengeleiteten Biologieunterricht? In: B. Reinisch, D. Krüger & D. Mahler (Hrsg.), *Biologiedidaktische Nature-of-Science-Forschung: Zukunftsweisende Praxis*. Berlin: Springer.
- [6] Lieberei, T., Welter, V.D., **Großmann, L.** & Krell, M. (2023). Findings from the Expert-Novice Paradigm on Differential Response Behavior among Multiple-Choice Items of a PCK Test -Implications for Test Development. *Frontiers in Psychology* (Section Educational Psychology). <https://doi.org/10.3389/fpsyg.2023.1240120>
- [5] **Großmann, L.** & Krüger, D. (2023). Assessing the Quality of Science Teachers' Lesson Plans: Evaluation and Application of a Novel Instrument. *Science Education*, 1-37. <https://doi.org/10.1002/sc.21832>
- [4] **Großmann, L.** & Krüger, D. (2023). Identifying Performance Levels of Enacted Pedagogical Content Knowledge in Trainee Biology Teachers' Lesson Plans. In: G. S. Carvalho, A. S. Afonso & Z. Anastácio (Hrsg.), *Fostering Scientific Citizenship in an Uncertain World. Selected Papers from the ESERA 2021 Conference* (S. 95-116). Cham: Springer. [https://doi.org/10.1007/978-3-031-32225-9\\_7](https://doi.org/10.1007/978-3-031-32225-9_7)
- [3] **Großmann, L.** & Krüger, D. (2022). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar\*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, 28, 4. <https://doi.org/10.1007/s40573-022-00141-w>
- [2] **Großmann, L.** & Krüger, D. (2022). Students' Conceptions as a Neglected Perspective in Trainee Teachers' Biology Lesson Plans. In: K. Korfiatis & M. Grace (Hrsg.), *Current Research in Biology Education. Contributions from Biology Education Research* (S. 181-193). ERIDOB. Cham: Springer. [https://doi.org/10.1007/978-3-030-89480-1\\_14](https://doi.org/10.1007/978-3-030-89480-1_14)
- [1] **Großmann, L.** & Krüger, D. (2020). Entwicklung und Anwendung eines Kategoriensystems zur Analyse des fachdidaktischen Wissens angehender Biologie-Lehrkräfte in schriftlichen Unterrichtsplanungen, *Erkenntnisweg Biologiedidaktik*, 19, 21-39.

## Veröffentlichungen mit Vorschlägen für die Unterrichts- und Seminarpraxis

---

### Herausgeberschaft:

- [2] **Großmann, L.**, Reinisch, B. & Krüger, D. (Hrsg.) (in Vorbereitung). *Biosphäre Lernaufgaben*. Berlin: Cornelsen Verlag.
- [1] **Großmann, L.**, Nessler, S.H. & Krüger, D. (Hrsg.) (2022). *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung*. Berlin: Cornelsen Verlag.

### Beiträge für die Unterrichtspraxis:

- [7] Vogt, N., **Großmann, L.** & Krüger, D. (im Druck). Logisches Schließen im Biologieunterricht – Wenn Aufgaben evolvieren. *MNU-Journal*.
- [6] **Großmann, L.**, Vogt, N. & Tietjen, B. (2023). Bietet Schwarmverhalten einen Überlebensvorteil für Fische? Testen von Hypothesen mithilfe eines Modells. *Unterricht Biologie*, Nr. 482, 47(1), S. 5-9.

- [5] **Großmann, L.** (2022). Charles Darwin und Alfred Russel Wallace. In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 116-122). Berlin: Cornelsen Verlag.
- [4] **Großmann, L.** & Krüger, D. (2022). Guckst du noch oder beobachtest du schon? In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 39-45). Berlin: Cornelsen Verlag.
- [3] **Großmann, L.** (2022). Coole Kriminelle? Von ungestraften Regelbrüchen in der Natur. In: L. Großmann, S.H. Nessler, & D. Krüger (Hrsg.), *Biosphäre Mysterys SI/SII. Rätselhafte Wege der Erkenntnisgewinnung* (S. 102-108). Berlin: Cornelsen Verlag.
- [2] Bösche, R. & **Großmann, L.** (2020). *COVID-19-Pandemie - Erarbeitung im Mystery. RAAbits Biologie Mittlere Schulformen*, EL 40, 1-17, online unter: <https://www.raabits.de/unterrichtsmaterial/biologie/humanbiologie/13459/covid-19-pandemie-erarbeitung-im-mystery>
- [1] **Großmann, L.**, Mathesius, S. & Dannemann, S. (2018). Kastanienblätter unter der Lupe. Der Roskastanienminiermotte auf der Spur. In: *Biologie 5-10* (24), Friedrich Verlag, Seelze, S. 26-29.

### Beiträge für die Seminarpraxis:

- [6] Böhnke, A., Dietz, D., **Großmann, L.**, Wienmeister, A., & Zoppke, T (2023). Neue videobasierte Lehr-Lern-Gelegenheiten für die erste und zweite Phase der Lehrkräftebildung im FOCUS Videoportal. Fachdidaktische und pädagogisch-psychologische Perspektiven zur Reflexion von Unterrichtsvideos. In: L. Mientus, C. Klempin & A. Nowak (Hrsg.), *Reflexion in der Lehrkräftebildung Empirisch – Phasenübergreifend – Interdisziplinär* (S. 309–316). Universitätsverlag Potsdam: Potsdam. <https://doi.org/10.25932/publishup-59171>
- [5] **Großmann, L.**, Lotz, A., Mulke, S. & Krüger, D. (2022). Probleme beim Planen?! Acht Ratschläge zum Gelingen eines Unterrichtsentwurfs im Fach Biologie, *MNU-Journal*, 75 (05), 368-376.
- [4] Böhnke, A., Jordan, A., **Großmann, L.**, Haase, S., Helbig, K., Müller, J., Achour, S., Krüger, D. & Thiel, F. (2022). Das FOCUS Videoportal der Freien Universität Berlin – Videobasierte Lerngelegenheiten für die erste und zweite Phase der Lehrkräftebildung. In R. Junker, V. Zucker, M. Oellers, T. Rauterberg, S. Konjer, N. Meschede & M. Holodynski (Hrsg.), *Lehren und Forschen mit Videos in der Lehrkräftebildung* (S. 37-55), Münster: Waxmann. <https://doi.org/10.31244/9783830995111>
- [3] **Großmann, L.** & Krüger, D. (2022). Biologieunterricht erfolgreich planen – ein Kriterienraster zum Schreiben von Unterrichtsentwürfen, *SEMINAR – Lehrerbildung und Schule*, 2022 (1), 91-110.
- [2] Böhnke, A., Becker, J., **Großmann, L.**, Haase, S., Jordan, A., Müller, J., Achour, S., Krüger, D., Ramseger, J. & Thiel, F. (2021). Das FOCUS Videoportal der Freien Universität Berlin. In Bundesministerium für Bildung und Forschung (Hrsg.), *Lehren und Forschen mit Videos in der Lehrkräftebildung* (S. 13-14). BMBF.
- [1] **Großmann, L.** & Nessler, S.H. (2021). Biologieunterricht in Blogs kompetenzorientiert reflektieren – Verfassen von Mikroartikeln im Praxissemester. In: M. Kubsch, S. Sorge, J. Arnold und N. Graulich (Hrsg.). *Lehrkräftebildung neu gedacht – Ein Praxisbandbuch für die Lehre in den Naturwissenschaften und deren Didaktiken* (S. 166-175). Münster: Waxmann.

### Posterpräsentationen (\* zeigt an, wer das Poster präsentiert hat)

---

- [5] Christmann, C.\*, Binder, K., Blauza, S., Büchter, T., Eichler, A., **Großmann, L.**, Kremer, K., Krüger, D., Peters, A., Romeike, R. & Heuckmann, B. (2023). *Mit Simulationen Kompetenzen für das 21. Jahrhundert fördern: MINT-übergreifendes Verständnis von Modellieren und Modellierkompetenz, Risiko und Risikokompetenz sowie Unsicherheit und den kompetenten Umgang mit Unsicherheit*. Posterpräsentation auf der 24. Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, 18.-21.09.2023, Ludwigsburg.
- [4] **Großmann, L.\*** & Krüger, D. (2022). *Erkenntnisgewinnung oder Gewinnung von Erkenntnissen?* Posterpräsentation auf der Tagung „Biologiedidaktische Nature of Science-Forschung: Zukunftsweisende Praxis“, 08.09.-09.09.2022, Berlin.
- [3] Lieberei, T.\*, **Großmann, L.**, Krüger, D. & Krell, M. (2022). *Fachdidaktisches Wissen von angehenden Biologielehrkräften im Kompetenzbereich Erkenntnisgewinnung*. Posterpräsentation auf der 23. Internationalen Frühjahrsschule der Fachsektion Didaktik der Biologie (FDdB) im VBIO, 21.-23.03.2022, Bielefeld.
- [2] **Großmann, L.\*** & Krüger, D. (2020). *Analyse von Staatsexamentwürfen im Fach Biologie*. Posterpräsentation auf der 22. Internationalen Frühjahrsschule der Fachsektion Didaktik der Biologie (FDdB) im VBIO, 24.-27.02.2020, Halle.
- [1] **Großmann, L.\*** & Krüger, D. (2019). *Förderung von biologiespezifischen Handlungskompetenzen bei Referendarinnen durch Videozirkel*. Posterpräsentation auf der 22. Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, 09.-12.09.2019, Wien.

### Vorträge (\* zeigt an, wer den Vortrag gehalten hat)

---

- [11] **Großmann, L.**, & Krüger, D. (angenommen). *Supporting Pre-Service Science Teachers in Lesson Planning: A Scoring Rubric*. Vortrag auf der Tagung der American Educational Research Association (AERA), 11.-14.04.2024, Philadelphia.
- [10] Koberstein-Schwarz, M., **Großmann, L.**, Scholl, D., Krüger, D. & Meisert, A., (angenommen). *Establishing common ground in empirical research on science teachers' lesson planning: A scoping Review*. Vortrag auf der 97. Tagung der National Association for Research in Science Teaching (NARST), 17.-20.03.2024, Denver. [geteilte Erstautorenschaft]
- [9] Koberstein-Schwarz, M.\*, **Großmann, L.\***, Scholl, D., Krüger, D. & Meisert, A., (2023). *Unterrichtsplanungskompetenz in den Naturwissenschaftsdidaktiken – ein Scoping Review*. Vortrag auf der 24. Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, 18.-21.09.2023, Ludwigsburg.
- [8] Lieberei, T.\*, **Großmann, L.**, Welter, V. D. E., Krüger, D. & Krell, M. (2023). *Entwicklung und Evaluation eines Instruments zur Erhebung fachdidaktischen Wissens angehender Biologielehrkräfte im Kompetenzbereich Erkenntnisgewinnung: PCK\_SR-bio*. Vortrag auf der 24. Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, 18.-21.09.2023, Ludwigsburg.
- [7] **Großmann, L.\*** & Krüger, D. (2022). *Unterrichtsplanung phasenübergreifend kriterienorientiert reflektieren*. Vortrag auf der Tagung „Reflexion in der Lehrkräftebildung. Empirisch – Phasenübergreifend – Interdisziplinär“ an der Freien Universität Berlin und der Universität Potsdam, 05.-07.10.2022 (online).

## Lebenslauf

---

- [6] **Großmann, L.\*** & Krüger, D. (2022). *Development and validation of a scoring rubric to measure pre-service biology teachers' lesson plan quality*. Vortrag auf der 13. Tagung der Conference of European Researchers in Didactics of Biology (ERIDOB), 29.08.-02.09.2022, Nikosia.
- [5] **Großmann, L.\***, Lieberei, T., Dawborn-Gundlach, M., van Driel, J., Krüger, D. & Krell, M. (2022). *Measuring Pre-service Biology Teachers' PCK about Scientific Reasoning – A case of validity*. Vortrag auf der 95. Tagung der National Association for Research in Science Teaching (NARST), 27.-30.03.2022 (online).
- [4] **Großmann, L.\*** & Krüger, D. (2021). *Biologieunterricht erfolgreich planen – Ein Kriterienraster zum Schreiben von Unterrichtsentwürfen*. Vortrag im Kooperationsrat zur Koordinierung der Lehrkräftebildung des Landes Berlin, 12.11.2021 (online).
- [3] **Großmann, L.\*** & Krüger, D. (2021). *Wie planen Biologie-Referendar\*innen ihren Unterricht? Eine Untersuchung des Zusammenhangs zwischen Planungsqualität und fachdidaktischem Wissen*. Vortrag auf der 23. Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBiO, 13.-16.09.2021 (online).
- [2] **Großmann, L.\*** & Krüger, D. (2021). *What determines sophisticated lesson planning in biology? Investigating trainee teachers' enacted pedagogical content knowledge*. Vortrag auf der 14. Tagung der European Science Education Research Association (ESERA), 30.08.-03.09.2021 (online).
- [1] **Großmann, L.\*** & Krüger, D. (2019). *K2Teach – Grundlegende Handlungskompetenzen für eine adaptive Unterrichtspraxis im Studium erwerben – Einsatz von Videos in der Lehrkräftebildung*. Vortrag im Kooperationsrat zur Koordinierung der Lehrkräftebildung des Landes Berlin, 25.11.2019, Berlin.

## Workshops

---

- [5] Fricke, K., **Großmann, L.** & Reinisch, B. (2023). *Rätselhafte Wege der Erkenntnisgewinnung: Mit Mysterys über die Natur der Naturwissenschaften reflektieren*. Workshop im Rahmen der Landestagung des Verbandes zur Förderung des MINT-Unterrichts (MNU), Berlin/Brandenburg, 05.09.2023, Freie Universität Berlin.
- [4] **Großmann, L.**, Peters, A., Romeike, R. & Krüger, D., (2023). *Modellieren im Biologieunterricht*, Workshop auf der 1. Stakeholderkonferenz im Rahmen des Projekts siMINT, 17.-18.03.2023, Freie Universität Berlin.
- [3] Böhnke, A., Bormuth, M., Dietz, D. **Großmann, L.**, Haase, S., Jordan, A., Wienmeister, A. & Zoppke, T. (2022). *Videobasierte Lehr- Lerngelegenheiten für die erste und zweite Phase der Lehrkräftebildung im FOCUS Videoportal. Fachdidaktische und pädagogisch-psychologische Perspektiven zur Reflexion von Unterrichtsvideos*. Workshop auf der Tagung „Reflexion in der Lehrkräftebildung. Empirisch – Phasenübergreifend – Interdisziplinär“ an der Freien Universität Berlin und der Universität Potsdam, 05.-07.10.2022 (online).
- [2] Huch, S., Reinisch, B., **Großmann, L.**, Mecklenburg, L., Müller, S. & Mahler, D. (2022). *Ein Blick in das Universitätsleben – Abstecher in die Biologiedidaktik*. Online-Workshop für den Girls' Day, 28.04.2022, Freie Universität Berlin.
- [1] Huch, S., Reinisch, B., Göhner, M., **Großmann, L.**, Mückai, R. & Nessler, S. (2021). *Ein Blick in das Universitätsleben – Abstecher in die Biologiedidaktik*. Online-Workshop für den Girls' Day, 22.04.2021, Freie Universität Berlin.

### Abschlussarbeiten (\* mit Erstbetreuung)

---

- [16] Abduktives und deduktives Schließen an einer Blackbox ohne Regelwissen (Masterarbeit, 2023)
- [15] Erfassung und Förderung der Modellierkompetenz von Biologiestudierenden anhand einer biologischen Problemstellung (Masterarbeit, 2023)
- [14\*] Biologie-Lehramtsstudierende im Videozirkel: Unterrichtssituationen auswählen, fachdidaktisch beurteilen und Handlungsalternativen entwickeln (Masterarbeit, 2023)
- [13] Einfluss des abduktiven Schließens auf den deduktiven Einsatz von Modellen (Masterarbeit, 2023)
- [12\*] Qualitative Überprüfung von Stundenplanungen angehender Biologie-Lehrpersonen in den Kompetenzbereichen Erkenntnisgewinnung und Fachwissen mithilfe eines Kriterienrasters (Masterarbeit, 2022)
- [11\*] Wie planen angehende Biologie-Lehrkräfte Unterricht zu den Kompetenzbereichen Fachwissen und Erkenntnisgewinnung? Evaluation eines Kriterienrasters für Unterrichtsentwürfe (Masterarbeit, 2022)
- [10\*] Qualitative Überprüfung einer digitalen Selbstlernumgebung zu den Arbeitsweisen Beobachten und Experimentieren im Biologieunterricht (Masterarbeit, 2021)
- [9\*] Evaluation eines Kriterienrasters zur Beurteilung der Qualität von Biologie-Stundenentwürfen in den Kompetenzbereichen Kommunikation und Bewertung (Masterarbeit, 2021)
- [8\*] Wie planen angehende Biologie-Lehrkräfte Unterricht zu den Kompetenzbereichen Fachwissen und Erkenntnisgewinnung? Qualitative Überprüfung eines Kriterienrasters für Unterrichtsentwürfe (Masterarbeit, 2021)
- [7\*] Untersuchung von Unterrichtsentwürfen von Biologie-Referendar\*innen zum Kompetenzbereich "Erkenntnisgewinnung" (Masterarbeit, 2021)
- [6] Theoretisch fundierte Qualitätsbeurteilung von Mysterys für den Biologieunterricht (Masterarbeit, 2020)
- [5\*] Erkenntnisgewinnung versus Fachwissen – Untersuchung von Unterrichtsentwürfen im Fach Biologie mit Blooms Lernzieltaxonomie (Masterarbeit, 2020)
- [4\*] Fachdidaktisches Wissen angehender Biologielehrkräfte im Themenfeld „Neurobiologie“ und Kompetenzbereich „Kommunikation“ – Qualitative Inhaltsanalyse von Staatsexamensentwürfen im Fach Biologie (Masterarbeit, 2020)
- [3\*] Fachdidaktisches Wissen angehender Biologielehrkräfte im Kompetenzbereich „Bewerten“ – Qualitative Inhaltsanalyse von Staatsexamensentwürfen im Fach Biologie (Masterarbeit, 2020)
- [2\*] Fachdidaktisches Wissen angehender Biologielehrkräfte im Themenfeld „Ökologie“ – Qualitative Inhaltsanalyse von Staatsexamensentwürfen im Fach Biologie (Masterarbeit, 2020)
- [1\*] Fachdidaktisches Wissen angehender Biologielehrkräfte im Themenfeld „Neurobiologie“ – Qualitative Inhaltsanalyse von Staatsexamensentwürfen im Fach Biologie (Bachelorarbeit, 2020)