

# Entwicklung einer Typologie von Modellierungsstrategien Lehramtsstudierender: Das Projekt TypMol

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  
Maximilian Felix Göhner

im Jahr 2023



*Für Karla.*



## **Selbstständigkeitserklärung**

Diese Arbeit entstand zwischen November 2017 und Juli 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. Die Arbeit wurde selbstständig verfasst und alle Hilfsmittel wurden entsprechend aufgeführt. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden. Die Arbeit wurde im Rahmen des DFG-geförderten Projekts TypMol (Projektnummer: 327507949; Projektleiter Prof. Dr. Moritz Krell) angefertigt.

1. Gutachter Prof. Dr. Dirk Krüger, Didaktik der Biologie, Freie Universität Berlin
2. Gutachter Prof. Dr. Moritz Krell, IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Christian-Albrechts-Universität zu Kiel

Disputation am 18.10.2023



## Danksagung

Ich danke...

...**Moritz Krell** für die großartige Vorarbeit und Leitung des Projekts „TypMoL“, das immer konstruktive, motivierende Feedback, welches jede Reviewschleife erträglich gemacht hat, seine unendliche Geduld und die entspannten Mittagspausen mit den allerbesten Nudeln und Zimtschnecken.

...**Dirk Krüger** für den tiefen und humorvollen Einblick in die naturwissenschaftsdidaktische Forschung, die verspielten und nachdenklich machenden Rätsel, unsere ausschweifenden, abstrakten Diskussionen, die vielen Anregungen in den Kolloquien und eine modellhaft schöne Zeit in der Arbeitsgruppe.

...**Jonna Kirchhof** und **Maike Barnebeck** für ihre unglaublich kreative und fleißige Zuarbeit als die besten studentischen Hilfskräfte, die man sich wünschen konnte.

...**Annette Upmeier zu Belzen** für den universitätsübergreifenden, perspektivenschaffenden Austausch und die Möglichkeit auch Studierende der HU zu beforschen.

...allen **teilnehmenden Studierenden**, ohne die diese Arbeit nie möglich gewesen wäre und die einiges an Fragebögen über sich ergehen lassen mussten.

...**Tom, Leroy, Nubia, Bianca, René, Kristin** und **Christine** für ihre offenen Ohren, offenen Augen, und den kollegialen Austausch und die gemeinsamen Spaziergänge durch Dahlem.

...**meiner Familie** für die liebevolle Unterstützung.

...der **Deutschen Forschungsgemeinschaft** für die Förderung des Projekts *TypMoL* (Projektnummer 327507949).

*“Essentially, all models are wrong, but some are useful.”*

*George Box (Box & Draper, 1987, S. 74)*

## Inhaltsverzeichnis

1.	Einleitung .....	16
2.	Theoretischer Rahmen.....	17
2.1	Modelle und das Modellieren in den Naturwissenschaften .....	18
2.1.1	Naturwissenschaftstheoretische Überlegungen zu Modellen .....	21
2.1.2	Naturwissenschaftstheoretische Überlegungen zum Modellieren .....	23
2.2	Modelle und das Modellieren im naturwissenschaftlichen Unterricht.....	26
2.2.1	Möglichkeiten für den Einsatz von Modellen und des Modellierens im Unterricht.....	26
2.2.2	Curriculare Vorgaben zu Modellen und dem Modellieren .....	28
2.2.3	Empirische Arbeiten zum Einsatz von Modellen und des Modellierens im Unterricht ..	30
2.3	Modellierkompetenz .....	34
2.3.1	Exkurs zum Kompetenzbegriff.....	34
2.3.2	Einordnung der Modellierkompetenz in die professionelle Kompetenz von Lehrkräften	36
2.3.3	Modellierkompetenz definieren.....	37
2.3.4	Dimensionalität und Extensionalität der Modellierkompetenz.....	38
2.4	Die Dimensionen der Modellierkompetenz in der empirischen Forschung und ihre Operationalisierung für die vorliegende Untersuchung.....	42
2.4.1	Modell(ier)verstehen .....	42
2.4.2	Modellierprozesse.....	46
2.4.3	Modellierprodukt .....	50
3.	Ziele der Untersuchung .....	52
4.	Methodischer Zugang.....	53
4.1	Qualitative Inhaltsanalyse.....	54
4.2	Publikation 1: Methodologischer Exkurs zu Gütekriterien und Maßnahmen der Qualitätssicherung bei qualitativ inhaltsanalytischen Verfahren .....	56
4.3	Typologisierung .....	59
4.4	Forschungskontext .....	61
4.5	Stichprobe .....	62
4.6	Datenaufnahme .....	65

4.6.1	Modellverstehen .....	65
4.6.2	Modellierprozesse.....	65
4.6.3	Modellierprodukte .....	68
4.7	Datenauswertung.....	69
4.7.1	Modellverstehen .....	69
4.7.2	Modellierprozesse.....	70
4.7.3	Modellierprodukte .....	72
4.7.4	Quantitative Auswertung der Zusammenhänge.....	72
5.	Ergebnisse und Diskussion der Einzelpublikationen.....	73
5.1	Zusammenarbeit mit den MitautorInnen.....	74
5.2	Publikation 2: Fallanalyse zweier Modellierprozesse.....	75
5.3	Publikation 3: Entwicklung einer Typologie von Modellierungsstrategien.....	79
5.4	Publikation 4: Identifikation und Beschreibung von Hindernissen beim Modellieren.....	84
5.5	Publikation 5: Untersuchung des Zusammenhangs zwischen Modellierprozessen und den Samplingkriterien .....	91
5.6	Publikation 6: Untersuchung des Zusammenhangs zwischen Modellverstehen, Modellierprozessen und der Qualität des Modellierprodukts.....	95
6.	Übergeordnete Diskussion.....	102
6.1	Methodologische Diskussion .....	102
6.1.1	Stichprobe und Forschungskontext.....	103
6.1.2	Das Blackbox-Setting .....	104
6.1.3	Modellieren durch Zeichnen und Lautes Denken.....	105
6.1.4	Datenauswertung .....	106
6.2	Inhaltliche Diskussion.....	108
6.2.1	Individuelle Modellierprozesse .....	108
6.2.2	Typologie von Modellierungsstrategien .....	108
6.2.3	Dimensionalität und Extensionalität von Modellierkompetenz.....	109
6.2.4	Modellieren in der schulischen Praxis.....	111
7.	Fazit.....	112

7.1	Implikationen für die Forschung.....	112
7.2	Implikationen für die Lehrkräftebildung.....	113
8.	Literaturverzeichnis.....	114
9.	Anhang .....	I
9.1	Zusammenfassung und Abstract .....	II
9.2	Liste der aus dieser Dissertation hervorgegangenen Vorveröffentlichungen .....	III
9.2.1	Publikation 1 .....	IV
9.2.2	Publikation 2.....	XXIII
9.2.3	Publikation 3.....	XL
9.2.4	Publikation 4.....	LX
9.2.5	Publikation 5.....	LXXXVI
9.2.6	Publikation 6.....	C
9.3	Material & Daten .....	CXXXIV
9.3.1	Codelines & Zeichnungen der Modellierprodukte (geordnet nach Typ).....	CXXXIV
9.3.2	Forschungseinwilligung.....	CLIII
9.3.3	Hinweise zur Durchführung der Studie .....	CLIV
9.3.4	Kodierleitfaden Modellverstehen .....	CLVI

## Abbildungsverzeichnis

Abbildung 1: Hierarchie naturwissenschaftlicher Praxis (verändert nach Pasley et al., 2016).....	19
Abbildung 2: Modell zweier Covid-19 Ausbreitungsszenarien (Philips, 2020).....	27
Abbildung 3: Kompetenz als Kontinuum (Blömeke, König et al., 2015, S. 7).....	35
Abbildung 4: Das COACTIV-Modell zur professionellen Kompetenz von Lehrkräften am Beispiel des Fachs Mathematik (Baumert & Kunter, 2013, S. 29).....	36
Abbildung 5: Dimensionalität der Modellierkompetenz nach a) Nicolaou und Constantinou (2014, S. 54), b) Chiu und Lin (2019, S. 3) und c) Nielsen und Nielsen (2021a, S. 6).....	39
Abbildung 6: Prozessschema naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren (Göhner & Krell, 2018, adaptiert nach Krell et al., 2016). In der Abbildung sind die 3 Phasen und 9 spezifischen Handlungselemente des Modellierens dargestellt. ....	49
Abbildung 7: Anzahl der beschriebenen Maßnahmen der Qualitätssicherung in den untersuchten Artikeln, sowie Anteil der beschriebenen Maßnahmen, die einem Gütekriterium zugeordnet werden....	58
Abbildung 8: Modell empirisch begründeter Typenbildung .....	60
Abbildung 9: Das Blackbox-Setting der vorliegenden Untersuchung. Die Komponenten sind wie folgt beschriftet: a) schriftliche Aufgabe, b) beispielhaft beschriftete Tafel, c) Videokamera (1 von 3), d) Blackbox, e) Eimer mit Wasser, f) verschieden große Bechergläser, g) Messkolben.....	66
Abbildung 10: Darstellung des inneren Mechanismus der Black. Wasser wird durch einen Trichter in ein erstes Gefäß geleitet, das dieses gleichmäßig auf zwei weitere Gefäße verteilt. Wird ein bestimmter Wasserstand in einem der beiden unteren Gefäße erreicht, entleert sich das jeweilige Gefäß vollständig durch den Siphon. Da die beiden Siphons in unterschiedlichen Höhen angebracht sind, unterscheidet sich der eine Entleerung auslösende Wasserstand bei beiden Gefäßen (links 350 ml, rechts 550 ml). ....	67
Abbildung 11: Codeline der Tätigkeitssequenz von Angelina (angepasst nach Göhner & Krell, 2018). Zeitlich länger andauernde Events sind als größere Punkte dargestellt. ....	76
Abbildung 12: Codeline der Tätigkeitssequenz von Raphael (angepasst nach Göhner & Krell, 2018). Zeitlich länger andauernde Events sind als größere Punkte dargestellt. ....	77
Abbildung 13: Zustandsübergangsdiagramm der kombinierten Modellierprozesse aller 36 ProbandInnen. Die 19 Tätigkeiten des Modellierens, sowie Start und Ende der Modellierprozesse, sind als Knoten dargestellt. Die Übergänge zwischen den Tätigkeiten des Modellierens sind als gerichtete Kanten dargestellt, die mit ihrer Auftretenshäufigkeit beschriftet sind. Die Kantenrichtung ist durch die Kantenkrümmung angegeben und verläuft immer im Uhrzeigersinn. ....	81

Abbildung 14: Zustandsübergangsdiagramme der Modellierprozesse von einzelnen Probandinnen, die den sechs Typen von Modellierungsstrategien zugeordnet wurden. Von links oben nach rechts unten: Carlo (Typ 1), Angelina (Typ 2a), Kara (Typ 2b), Jonathan (Typ 2c), Ben (Typ 3a), Raphael (Typ 3b). .....	82
Abbildung 15: Sterndiagramm der quantifizierten Variablen Komplexität und Homogenität der Modellierprozesse. ProbandInnen, die dem gleichen Typ zugeordnet wurden, sind in derselben Farbe hervorgehoben und mit dem Durchschnittswert (Centroid) ihres Typen verbunden. ....	83
Abbildung 16: Gestapelte Säulen-Diagramme der Hindernisse unterteilt nach ihrem Auftreten in den Phasen des Modellierungsprozesses (verändert nach Göhner & Krell, 2021). Das obere Diagramm zeigt den Anteil jeder Phase am Auftreten des jeweiligen Hindernisses, das untere Diagramm den zeitlichen Anteil jeder Phase am Auftreten des jeweiligen Hindernisses. Zur besseren Sichtbarkeit der Verhältnisse wurden relative, prozentuale Werte dargestellt. Die Säulen sind zusätzlich mit den absoluten Werten beschriftet, die entweder die Häufigkeit (oben) oder die mit dem Hindernis verbrachten Minuten (unten) angeben.....	88
Abbildung 17: Beispiele für Modellierprodukte. Misty's Modellierprodukt integriert Konzept 1. Floyd's Modellierprodukt integriert die Konzepte 1 und 2. Jenny's Modellierprodukt integriert die Konzepte 1 und 3. ....	98

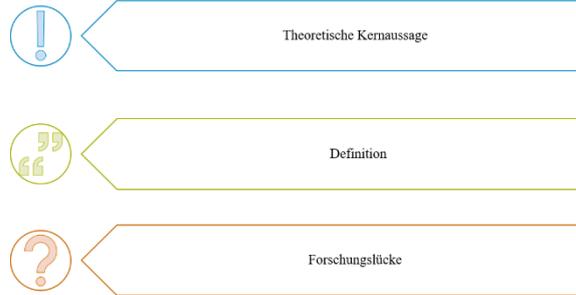
## Tabellenverzeichnis

Tabelle 1: Kompetenzmodell der Modellkompetenz (verändert nach Upmeyer zu Belzen & Krüger, 2010).....	44
Tabelle 2: Bezugspunkte, der in dieser Untersuchung durchgeführten, qualitativ inhaltsanalytischen Verfahren nach Stamann et al. (2016).....	54
Tabelle 3: Übersicht über die in den untersuchten Artikeln beschriebenen Gütekriterien und Maßnahmen der Qualitätssicherung. Das vollständige Kategoriensystem inklusive Kategorienbeschreibung und Kodierhinweisen ist im Anhang der Publikation (Göhner & Krell, 2020) beschrieben.....	57
Tabelle 4: Verteilung der ProbandInnen auf die Subgruppen des qualitativen Stichprobenplan.....	63
Tabelle 5: Übersicht über die teilnehmenden Lehramtsstudierenden, alphabetisch geordnet nach Pseudonym (verändert nach Göhner & Krell, 2021).....	64
Tabelle 6: Fragebogen zum Modellverstehen .....	65
Tabelle 7: Übungsaufgaben zum lauten Denken.....	68
Tabelle 8: Kategoriensystem mit 19 Tätigkeiten des Modellierens (Krell et al., 2019) sowie deren Verortung in Phasen des Modellierprozesses (E: Exploration; H: Herstellung; V: Vorhersage; vgl. Abbildung 1).....	70
Tabelle 9: Übersicht über die im Rahmen der vorliegenden Untersuchung entstandenen Publikationen .....	73
Tabelle 10: Angabe der Rolle der Mitwirkenden Autoren Göhner (G), Krell (K) und Bielik (B) an den präsentierten Publikationen adaptiert nach Brand et al. (2015). Die Kürzel „l“, „e“ und „s“ geben dabei den Grad der Verantwortlichkeit an (lead, equal, supporting). .....	74
Tabelle 11: Vergleich des Auftretens und des zeitlichen Anteils der in den Modellierungsprozessen von Angelina und Raphael beobachteten Tätigkeiten (Göhner & Krell, 2018) .....	77
Tabelle 12: Aussagen zum Modellverstehen von Angelina und Raphael .....	78
Tabelle 13: Typen von Modellierungsstrategien, deren Komplexität und Homogenität sowie die Anzahl von ProbandInnen deren Modellierprozesse dem jeweiligen Typ zugeordnet wurden (Göhner & Krell, 2021, 2022b).....	80
Tabelle 14: Kategoriensystem der identifizierten Hindernisse in den Modellierungsprozessen von Biologie-Lehramtsstudierenden (N = 36) an einer Blackbox, absteigend sortiert nach Anzahl der ProbandInnen mit den Hindernissen (Göhner & Krell, 2021).....	85

Tabelle 15: Kreuztabelle der Hinderniskategorien und Tätigkeiten aller 36 ProbandInnen, die zeigt, wie viele Kodierungen je Hinderniskategorie mit den identifizierten Tätigkeiten gemeinsam auftreten (Göhner & Krell, 2021).....	89
Tabelle 16: Deskriptive Statistiken zu den fünf Variablen (verändert nach Göhner & Krell, 2022a). Es sind jeweils die Durchschnittswerte und (in Klammern) die Standardabweichung angegeben. Die Variablen Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und allgemeine kognitive Fähigkeiten sind entsprechend der in der englisch-sprachigen Publikation genutzten Begriffe <i>scientific reasoning competencies</i> und <i>general cognitive abilities</i> respektive als SRC und GCA abgekürzt. ....	92
Tabelle 17: Pearson Korrelations-Koeffizienten der Zusammenhänge zwischen den relevanten Variablen (n. s. = $p > 0,05$ ).....	92
Tabelle 18: Übersicht aller ProbandInnen und derer quantitativen Werte hinsichtlich Modellverstehen, Modellierprozess und Modellierprodukt .....	96
Tabelle 19: Verteilung der Antwortniveaus hinsichtlich der fünf Aspekte des Modellverstehens aller 36 ProbandInnen (verändert nach Göhner et al., 2022).....	97
Tabelle 20: Übersicht über die Integration der drei Konzepte durch die ProbandInnen in den Modellierprodukten.....	99
Tabelle 21: Korrelationstabelle der Spearman'schen Rangkorrelation zwischen den Variablen dekontextualisiertes Modellverstehen, kontextualisiertes Modellverstehen, Komplexität der Modellierprozesse, Homogenität der Modellierprozesse und Anzahl der integrierten Konzepte in das Modellierprodukt.....	99
Tabelle 22: Vorschlag zur Zuordnung von Maßnahmen der Qualitätssicherung zu den klassischen Gütekriterien Objektivität, Reliabilität und Validität bei der Umsetzung qualitativ inhaltsanalytischer Verfahren.....	102

## Legende hervorgehobener Elemente

In der vorliegenden Untersuchung sind theoretische Kernaussagen, Definitionen und Forschungslücken wie rechts dargestellt hervorgehoben.



## Verzeichnis der theoretischen Kernaussagen

Kernaussage 1: Modelle sind zentral in den Naturwissenschaften.....	20
Kernaussage 2: Modelle sind vielfältig. ....	21
Kernaussage 3: Modelle sind in der Lehre unverzichtbar. ....	26
Kernaussage 4: Modelle haben viele Anwendungszwecke.....	27
Kernaussage 5: Curricula fordern epistemische Perspektiven auf Modelle und das Modellieren. ....	29
Kernaussage 6: Im Unterricht werden Modelle als Medien eingesetzt.....	32
Kernaussage 7: Lehrkräfte benötigen selbst Modellierkompetenz.....	33

## Verzeichnis der Definitionen

Definition 1: Modelle sind zweckgerichtet entwickelte epistemische Artefakte. ....	22
Definition 2: Modellieren ist ein zielgerichteter Prozess. ....	25
Definition 3: Kompetenzen nach Weinert (2001b). ....	34
Definition 4: Modellierkompetenz nach Krüger & Upmeyer zu Belzen (2021).....	37
Definition 5: Dimensionalität der Modellierkompetenz.....	41

## Verzeichnis der Forschungslücken

Forschungslücke 1: Prozessorientierte Studien zum Modellieren fehlen.....	48
Forschungslücke 2: Inwieweit sind die identifizierten Modellierungsstrategien übertragbar? .....	104
Forschungslücke 3: Welchen Einfluss haben Motivation und Volition auf Modellierkompetenz? ....	105
Forschungslücke 4: Wie kann Modellierverstehen erfasst werden? .....	106
Forschungslücke 5: Inwiefern sind Modellierprozesse situationsspezifisch? .....	111
Forschungslücke 6: Wie verändern sich Modellierkompetenz und schulische Praxis über die Zeit?.	111

## 1. Einleitung

*“We live in a society exquisitely dependent on science and technology, in which hardly anyone knows anything about science and technology. This is a clear prescription for disaster.”*

*Carl Sagan (1990, S. 264)*

Ein Desaster. Das ist die Erwartungshaltung, die der verstorbene Astronom, Physiker und Sachbuchautor Carl Edward Sagan in seinem Artikel „Why we need to understand science“ an eine von Wissenschaft und Technologie abhängige Gesellschaft formuliert, in der niemand etwas über Wissenschaft oder Technologie wisse. Während unsere Gesellschaft, die so wissenschafts- und technologieabhängig ist wie nie zuvor (Unesco, 2015), es zwar geschafft hat, einige der von Sagan genannten Desaster abzuwenden, bleiben andere Herausforderungen, beispielsweise der Klimawandel, weiterhin relevant. Um diesen Herausforderungen entgegenzutreten, rückt ein Verstehen von Wissenschaft und Technologie immer mehr in den Fokus aktueller Bildungsdebatten (Billingsley & Fraser, 2018; Driver et al., 1996; Gräber et al., 2002; Höttecke & Allchin, 2020; Millar, 1996). Unter dem Begriff der naturwissenschaftlichen Grundbildung (*scientific literacy*) wird dieses Verstehen definiert als “the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity” (National Academy of Sciences, 1996, S. 22). Außerdem umfasst sie die Fähigkeiten die Qualität naturwissenschaftliche Informationen evaluieren zu können, sowie die Fähigkeiten natürliche Phänomene beschreiben, erklären und vorhersagen zu können (National Academy of Sciences, 1996). Gerade die Covid-19-Pandemie zeigte als globales Phänomen, welche hohe Relevanz eine naturwissenschaftliche Grundbildung für unsere Gesellschaft haben kann und macht deutlich, wie wichtig guter naturwissenschaftlicher Unterricht ist: Unterricht, der – über das Erlernen naturwissenschaftlicher Inhalte hinaus - auch das Erlernen komplexer naturwissenschaftlicher Denkweisen, das Verstehen der Natur der Naturwissenschaften, die Umsetzung naturwissenschaftlicher Praxis und das Adressieren sozio-wissenschaftlicher Themen in den Blick nimmt (Hodson, 2014). Folglich erhalten die naturwissenschaftlichen Denk- (*scientific reasoning*) und Arbeitsweisen (*scientific inquiry*) weltweit in Curricula und Standarddokumenten immer höhere Relevanz (ACARA, 2015; BCMOE, 2019; KMK, 2005, 2020; NGSS Lead States, 2013b; VCAA, 2016). Empirische Studien zeigen allerdings, dass Lernziele, die über das Erlernen naturwissenschaftlicher Inhalte hinausgehen, bisher in der schulischen Realität kaum Anwendung finden (z. B. Weiss et al., 2003) und dass Lehrende in den Naturwissenschaften über ein geringes Verständnis naturwissenschaftlicher Denkweisen oder der Natur der Naturwissenschaften verfügen (Torrijos-Muelas et al., 2023). Folglich wird die Erfassung und Förderung der komplexen, naturwissenschaftlichen Denkprozesse von Osborne (2013) als *die* Herausforderung der naturwissenschaftsdidaktischen Forschung für das 21. Jahrhundert proklamiert. Die vorliegende Arbeit nimmt sich dieser Herausforderung an, indem sie eine zentrale Denk- und Arbeitsweise der Naturwissenschaften untersucht: Das Modellieren (Giere et al., 2006; Halloun, 2004).

## 2. Theoretischer Rahmen

Der theoretische Rahmen dieser Arbeit ist in vier Abschnitte gegliedert: In Abschnitt 2.1 wird zuerst die Relevanz von Modellen und des Modellierens in den Naturwissenschaften dargestellt. Dabei werden naturwissenschaftstheoretische Überlegungen für die Begriffe „Modell“ und „Modellieren“ beschrieben, und einzelne Definitionen und Konzeptualisierungen für die weitere Untersuchung ausgewählt.

In Abschnitt 2.2 werden Modelle und das Modellieren im naturwissenschaftlichen Unterricht thematisiert. Hierfür wird zuerst das Potenzial des Einsatzes von Modellen und des Modellierens im Unterricht beschrieben. Anhand ausgewählter Beispiele werden curriculare Vorgaben zum Umgang mit Modellen und dem Modellieren zusammengefasst. Anschließend wird anhand empirischer Studien zum Einsatz von Modellen und des Modellierens in der Schule gezeigt, dass das beschriebene Potenzial und die curricularen Vorgaben in der schulischen Realität nicht erreicht werden. Schließlich wird die Modellierkompetenz von Lehrkräften als ein wichtiger Ansatzpunkt identifiziert, um den Einsatz von Modellen und des Modellierens im naturwissenschaftlichen Unterricht zu fördern.

In Abschnitt 2.3 wird daher das theoretische Konstrukt Modellierkompetenz detailliert beschrieben. Nach einem kurzen Exkurs zum Kompetenzbegriff und der Zuordnung der Modellierkompetenz zur professionellen Kompetenz von Lehrkräften, werden hierfür die Begriffe Modellkompetenz und Modellierkompetenz abgegrenzt und verschiedene Vorschläge zur Struktur und Dimensionierung der Modellierkompetenz gesammelt. Für die vorliegende Untersuchung wird Modellierkompetenz dabei in drei Dimensionen unterteilt: Modellierverstehen, Modellierprozess und Modellierprodukt.

In Abschnitt 2.4 werden dann ausgewählte naturwissenschaftstheoretische Überlegungen und empirische Ergebnisse zu den drei Dimensionen Modellierverstehen, Modellierprozess und Modellierprodukt dargestellt, aus denen eine für die vorliegende Untersuchung relevante Definition und Operationalisierung der jeweiligen Dimension abgeleitet wird.

## 2.1 Modelle und das Modellieren in den Naturwissenschaften

*“In short, scientific methodology is primarily about making, testing and using conceptual models of patterns in physical realities.”*

*(Halloun, 2004, S. 29)*

Diese naturwissenschaftstheoretische Überlegung von Halloun (2004) beschreibt wissenschaftliche Methodik primär als das Entwickeln, Testen und Nutzen von konzeptuellen Modellen der physikalischen Realität. Giere (1999) geht noch einen Schritt weiter und schreibt „Scientific reasoning is to a large extent model-based reasoning.“ (Giere, 1999, S. 56); er formuliert damit, dass jegliches wissenschaftliches Denken als weitgehend modellbasiertes Denken angesehen werden kann. Beide Zitate zeigen, dass die Autoren Modellen und dem Modellieren in den Naturwissenschaften eine zentrale Rolle zusprechen, die disziplinunabhängig über alle Naturwissenschaften hinweg anerkannt wird (C. J. Suckling et al., 1978; Giere et al., 2006; Godfrey-Smith, 2006; Laubichler & Müller, 2007; Rowbottom, 2009).

### **Modelle und das Modellieren werden vielfältig eingesetzt**

Diese zentrale Rolle von Modellen und dem Modellieren ist dabei kaum verwunderlich: Interviewstudien mit NaturwissenschaftlerInnen zeigen, dass Modelle und das Modellieren in der naturwissenschaftlichen Praxis für vielfältige Zwecke eingesetzt werden. Die befragten NaturwissenschaftlerInnen nutzen Modelle für die Erklärung und Organisation von Observationen, Vorhersagen oder Tests; manche NaturwissenschaftlerInnen beschreiben, dass Modelle ihnen das Verstehen eines Systems auf einer weniger komplexen, abstrakteren oder visualisierten Ebene ermöglichen oder ein richtungsgebendes Rahmenwerk für die eigene Forschung darstellen können (Schwartz & Lederman, 2005). Auch Bailer-Jones (2002) fand in einer weiteren Interviewstudie mit neun NaturwissenschaftlerInnen aus Großbritannien ähnliche Überlegungen zu Modellen. Die teilnehmenden NaturwissenschaftlerInnen äußerten fünf konsensuelle Gedanken zu Modellen: (1) Modelle sind von zentraler Wichtigkeit für die Naturwissenschaften, (2) Definitionen des Modellbegriffs sind divers, (3) Modelle stellen durch das Weglassen und Vereinfachen von Eigenschaften eines zu modellierenden Sachverhaltes seine “Essenz” dar, (4) Modelle liefern Erkenntnisse über die Erklärung bestehender Daten hinaus und (5) Modelle werden empirisch getestet. In der naturwissenschaftsdidaktischen Forschung stehen außerdem drei Zwecke von Modellen im Vordergrund: Modelle können genutzt werden, um natürliche Phänomene retrodiktiv zu erklären (*abduktiver Schluss*; Krüger & Upmeier zu Belzen, 2021), prediktiv anhand des Modells Vorhersagen über das natürliche Phänomen abzuleiten (*deduktiver Schluss*; Krüger & Upmeier zu Belzen, 2021) oder naturwissenschaftliche Fachinhalte am Modell als Medium zu veranschaulichen (Krüger & Upmeier zu Belzen, 2021; Upmeier zu Belzen & Krüger, 2010).

**Modelle und das Modellieren als zentrale Denk- und Arbeitsweise der Naturwissenschaften**

In naturwissenschaftstheoretischen Überlegungen wird die genaue Verortung von Modellen und dem Modellieren im Verhältnis zum Begriff Theorie und anderen Denk- und Arbeitsweisen weiterhin diskutiert: Während einige AutorInnen den Umgang mit Modellen als nur eine von vielen Denk- und Arbeitsweisen beschreiben, beispielsweise als einen *style of scientific reasoning* (z. B. Kind & Osborne, 2017), argumentieren andere AutorInnen, dass Modelle und das Modellieren eine übergeordnete Rolle einnehmen (Bailer-Jones, 2002, 2003; Pasley et al., 2016). Gleichzeitig zeigt die oben bereits genannte Interviewstudie, dass es NaturwissenschaftlerInnen schwer fällt die Begriffe „Theorie“ und „Modell“ abzugrenzen. Während einige NaturwissenschaftlerInnen im Sprachgebrauch keine wirklichen Unterschiede zwischen dem Modell- und dem Theoriebegriff machen, gaben andere an, dass Modelle erst zu Theorien werden, wenn sie besser getestet wurden oder dass Theorien genauer oder allgemeiner seien als Modelle (Bailer-Jones, 2002). In der Hierarchie naturwissenschaftlicher Praxis (Abbildung 1) von Pasley et al. (2016), werden die Entwicklung von Erklärungen natürlicher Phänomene (synonym „Theorien“) und die Entwicklung und Anwendung von Modellen dagegen gleichgestellt.

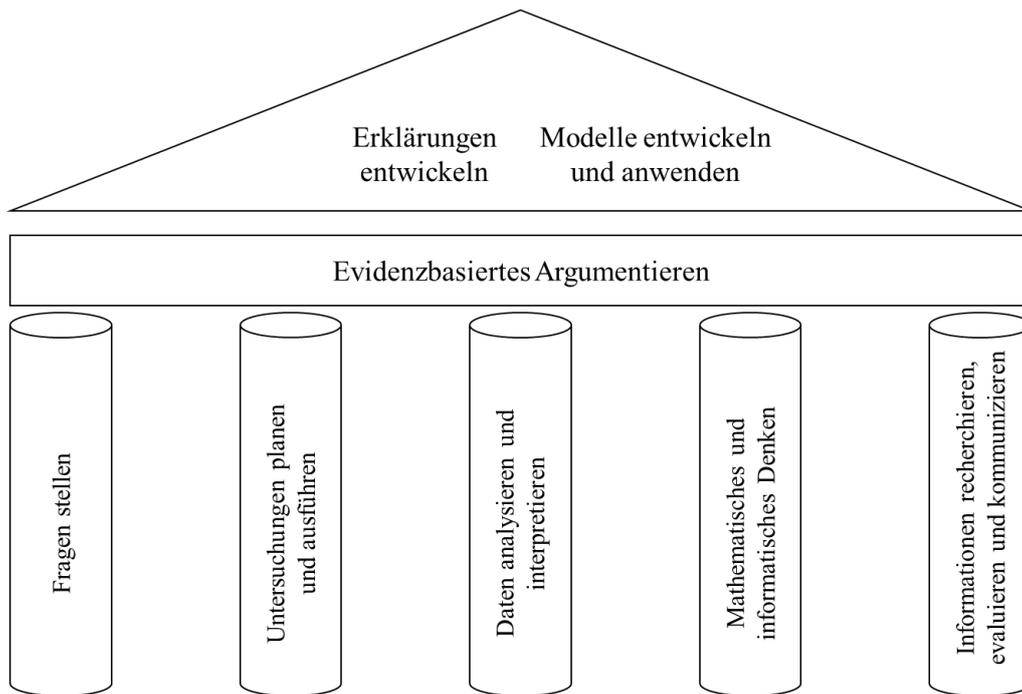


Abbildung 1: Hierarchie naturwissenschaftlicher Praxis (verändert nach Pasley et al., 2016)

Für Pasley et al. (2016) gelten die Entwicklung von Erklärungen und die Entwicklung und Anwendung von Modellen als die ultimativen Ziele der Naturwissenschaften, die über allen anderen naturwissenschaftlichen Denk- und Arbeitsweisen stehen, da sowohl die Entwicklung von naturwissenschaftlichen Erklärungen als auch die Entwicklung und Anwendung von Modellen auf validen und reliablen Evidenzen basieren, die aus den anderen naturwissenschaftlichen Denk- und Arbeitsweisen gewonnen werden (Pasley et al., 2016).

## Theoretischer Rahmen

Insgesamt wird deutlich, wie untrennbar naturwissenschaftliche Erklärungen und die Entwicklung und Anwendung von Modellen zusammenhängen (Gilbert & Justi, 2016; Lehrer & Schauble, 2006b; Passmore et al., 2014; Windschitl et al., 2008).



Modelle und das Modellieren sind für die Naturwissenschaften von zentraler, wenn nicht sogar übergeordneter Bedeutung und werden für vielfältige Zwecke eingesetzt.

Kernaussage 1: Modelle sind zentral in den Naturwissenschaften.

### 2.1.1 Naturwissenschaftstheoretische Überlegungen zu Modellen

Eine mathematische Funktion zur Vorhersage der Ausbreitung eines Virus (Philips, 2020), ein Flussdiagramm des ökologischen Wasserkreislaufs (Zangori et al., 2015), eine Computer-Simulation der Mondphasen (Kiroğlu et al., 2021), eine Fruchtfliege (Levy & Currie, 2015) oder eine Styroporkugel an einem Holzstab mit Knete (Maia & Justi, 2009). All diese völlig unterschiedlichen, teils fassbaren, teils unfassbaren Entitäten haben eines gemeinsam: Sie sind wissenschaftliche Modelle.

#### Die Formenvielfalt der Modelle

*„There is an additional trivial truth, which may strike some people as shocking:  
anything can be a model of anything else!“  
(Wartofsky, 1966, S. 4)*

So kommentierte der amerikanische Philosoph Wartofsky die große Vielfalt wissenschaftlicher Modelle in seinem Aufsatz „The model muddle“. Diese für Wartofsky triviale und doch gleichzeitig schockierende Aussage, alles könne ein Modell von etwas anderem sein, wird klar sobald man sich bewusst macht, dass ein Objekt zum Modell wird, sobald es von einem Subjekt dazu erklärt wird (Mahr, 2008, 2011). Die Formenvielfalt der Modelle ist somit nur durch die Fantasie der modellierenden Person beschränkt, die zwischen zwei Objekten eine Beziehung herstellen muss. Modelle können somit als haptische Objekte, mathematische Formeln, digitale Simulationen, lebende Organismen, Gedankenexperimente und vieles mehr vorliegen (Frigg & Hartmann, 2020; Knuuttila, 2011).



Modelle sind vielgestaltig und müssen immer im Kontext ihrer Entwicklung und Anwendung gesehen werden.

Kernaussage 2: Modelle sind vielfältig.

### Der Modellbegriff in dieser Arbeit

Die Abstraktheit des Modellbegriffs<sup>1</sup> und der Umstand, dass jedes Objekt potenziell ein Modell für etwas anderes darstellt, macht die Beantwortung der Frage „Ist Objekt x ein Modell?“ unmöglich. Damit ist auch eine ontologische Definition ausgeschlossen, die es erlauben würde anhand von Kriterien zu entscheiden, ob etwas ein Modell ist oder nicht (Mahr, 2008; Morrison & Morgan, 1999)<sup>2</sup>. Stattdessen liegen in der naturwissenschaftstheoretischen Forschung verschiedene Vorschläge zur Definition des Modellbegriffs vor, die Modelle eher als mediale oder eher als epistemische Artefakte charakterisieren: Medial-orientierte Definitionen (Bailer-Jones, 2003; Giere, 2004; Morrison & Morgan, 1999; van Fraassen, 2010) fokussieren sich eher auf den Aspekt der Repräsentation (Modelle *von* Etwas; Gouvea & Passmore, 2017; Upmeyer zu Belzen & Krüger, 2010); Beispielsweise schreibt Bailer-Jones (2003) „*a model is an interpretative description of a phenomenon (object or process) that facilitates perceptual as well as intellectual access to that phenomenon*“ (S. 61). Ein Modell ist für sie also eine interpretative Beschreibung eines Phänomens (Objekt oder Prozess), welches eine Wahrnehmung und intellektuelle Auseinandersetzung mit dem Phänomen ermöglicht. Epistemisch orientierte Definitionen hingegen (Frigg & Hartmann, 2020; Giere et al., 2006; Knuuttila, 2005, 2011; Nersessian, 1999) fokussieren sich auf den Aspekt der Forschung mit Modellen und betonen so die Rolle des Modells als Werkzeug der Erkenntnisgewinnung (Modelle *für* Etwas; Gouvea & Passmore, 2017; Upmeyer zu Belzen & Krüger, 2010). So beschreibt Knuuttila (2011) Modelle als „*concrete artefacts that are built by specific representational means and are constrained by their design in such a way that they facilitate the study of certain scientific questions, and learning from them by means of construction and manipulation*“ (S.1). In Anlehnung an Knuuttila (2011) folgt die vorliegende Untersuchung im Weiteren einer epistemisch orientierten Definition des Modellbegriffs:



Modelle sind zweckgerichtet entwickelte epistemische Artefakte, die spezifische Eigenschaften eines Phänomens abbilden oder vernachlässigen, um die Beantwortung einer spezifischen Frage zu ermöglichen, indem neue Erkenntnisse aus der Entwicklung und Manipulation des Modells abgeleitet werden.

Definition 1: Modelle sind zweckgerichtet entwickelte epistemische Artefakte.

<sup>1</sup> Für eine etymologische Einordnung des Modellbegriffs siehe z. B. Mahr (2008).

<sup>2</sup> Allerdings liegen durchaus Arbeiten zu Merkmalen vor, die Objekten zugesprochen werden, die als Modell genutzt werden (z. B. Stachowiak, 1973; van der Valk et al., 2007).

### 2.1.2 Naturwissenschaftstheoretische Überlegungen zum Modellieren

Lange wurde die Praxis des naturwissenschaftlichen Modellierens kaum explizit beschrieben und in der naturwissenschaftstheoretischen Literatur fehlten Standards zum wissenschaftlichen Modellieren (Morrison & Morgan, 1999). In den letzten zwei Jahrzehnten hat sich in Forschung und Theoriebildung zum naturwissenschaftlichen Modellieren zwar einiges getan, trotzdem erscheint weitere Arbeit nötig, um ein ganzheitlicheres theoretisches Verständnis zum naturwissenschaftlichen Modellieren zu erreichen (Chiu & Lin, 2019). Ganz allgemein wird das Modellieren als zielorientierter, oft zyklischer Prozess beschrieben (Clement, 1989; Giere et al., 2006). Unter dem Begriff des Modellierens werden dabei verschiedene Vorgehensweisen zusammengefasst, wie beispielsweise die Entwicklung, Anwendung, Evaluation, Revision oder Auswahl von Modellen (z. B. Chiu & Lin, 2019). Hierbei sind analog zum Modellbegriff verschiedene Ziele denkbar. Zum einen kann das Modellieren darauf abzielen ein anschauliches, repräsentatives und inhaltlich korrektes Modell von einem Phänomen zu entwickeln. Zum anderen kann das Modellieren genutzt werden, um neue Erkenntnisse über ein Phänomen zu gewinnen, indem iterativ Modelle entwickelt, manipuliert und mit Beobachtungen des realen Phänomens abgeglichen werden (Schwarz et al., 2009). Mit dem Ziel der Erkenntnisgewinnung weist das Modellieren Ähnlichkeiten zum Problemlösen (Clement, 2008; vgl. Experimentieren als Problemlösen, Hammann et al., 2008) und naturwissenschaftlichen Denken auf (Clement, 1989).

#### **Modellieren als Problemlösen**

Der Begriff Problemlösen beschreibt kognitive Prozesse eines Individuums, auf die indirekt geschlossen werden kann, indem beispielsweise Verbalisierungen oder Zwischenprodukte des Problemlöseprozesses analysiert werden (Funke, 2011). Problemlösen wird hierbei definiert als „zielorientiertes Denken und Handeln in Situationen [...], für deren Bewältigung keine routinierten Vorgehensweisen verfügbar sind“ (Mayer, 2007, S. 178). Übertragen auf das Modellieren, wird erkennbar, dass auch hier mit der Entwicklung von Erklärungen oder zuverlässigen Vorhersagen in Bezug auf ein natürliches Phänomen ein Ziel vorliegt, aber keine feste Handlungsroutine, die zu diesem Ziel führt (Funke, 2011; Klahr, 2002; Mayer, 2007). Folglich spielt auch Kreativität eine große Rolle beim Modellieren (Mierdel & Bogner, 2019; Zimmerman, 2005). Modellieren ist dabei insbesondere vergleichbar mit dem sogenannten analogen Problemlösen (Funke, 2011), das sich aus dem Abrufen von relevanten Erfahrungen aus einem Quellbereich und der Übertragung dieser Erfahrung auf das Problem zusammensetzt. Hierbei zeichnet sich das Modellieren zusätzlich dadurch aus, dass die Qualität der Übertragung in Form eines Modells zyklisch getestet und optimiert werden kann (Clement, 2008).

### **Modellieren als naturwissenschaftliches Denken**

Im Kontext des naturwissenschaftlichen Denkens kann beim Modellieren angelehnt an Klahr und Dunbar (1988) zwischen Überlegungen in der Erfahrungswelt und der Modellwelt unterschieden werden. Modellierende beobachten ein natürliches Phänomen und untersuchen dieses gegebenenfalls in der Erfahrungswelt, indem sie das dem Phänomen zugrundeliegende natürliche System explorieren. Entsprechend des analogen Problemlösens (Funke, 2011) werden dann Vorwissen und Erfahrungen hinsichtlich des Phänomens aktiviert, aus denen Modellierende ein initiales Modell entwickeln, welches die als für das Phänomen relevant identifizierten Variablen des natürlichen Systems in der Modellwelt isoliert (Clement, 1989). Das entwickelte Modell kann hierbei als gegenständliches Produkt vorliegen oder aus einer mentalen Auffassung bestehen (Mahr, 2008; vgl. *mental model*; Clement, 2008). Dieses Modell kann dann auf interne Konsistenz oder eine angemessene Darstellung des natürlichen Phänomens geprüft werden (Clement, 1989; Mahr, 2008). Darüber hinaus kann eine gedankliche (*imagistic simulation*; Clement, 2009) oder materielle Manipulation des Modells genutzt werden, um retrodiktiv Erklärungen zum Verhalten des natürlichen Systems abzuleiten (abduktiver Schluss; Krüger & Upmeyer zu Belzen, 2021) oder um prediktiv Hypothesen darüber abzuleiten, wie sich das System unter bestimmten Bedingungen verhält (deduktiver Schluss; Krüger & Upmeyer zu Belzen, 2021). Reichen die abgeleiteten Erklärungen nicht aus, um das Verhalten des natürlichen Systems zu erklären oder wurden die abgeleiteten Hypothesen auf der Grundlage weiterer Beobachtungen des natürlichen Systems falsifiziert, kann auf eine mangelnde Passung zwischen Modell und System geschlossen werden, sodass das Modell weiterentwickelt oder verworfen werden muss. Hiermit findet beim Modellieren ein ständiger Vergleich zwischen Erfahrungswelt und Modellwelt statt (Leisner-Bodenthin, 2006). Dieser Wechsel zwischen Erfahrungswelt und Modellwelt wird auch in einigen Definitionen des Modellierens deutlich, so beschreibt Godfrey-Smith (2006, S. 726) das Vorgehen von Modellierenden wie folgt: “The modeler’s strategy is to gain understanding of a complex real-world system via an understanding of simpler, hypothetical system that resembles it in relevant respects.” Modellieren ist somit der Vorgang ein Verstehen über komplexe Systeme der realen Welt zu erlangen, indem man einfachere, hypothetische Systeme versteht (und konstruiert), die dem realen System in relevanten Aspekten gleichen.

### Der Modellierbegriff in dieser Arbeit

Zusammengefasst wird der Prozess des Modellierens idealtypisch als iterativer oder zyklischer, dynamischer und komplexer Prozess verstanden, der sich aus der Entwicklung; Evaluation, Revision oder Ersetzung von veränderbaren Modellen zusammensetzt (Cheng & Lin, 2015; Giere et al., 2006; Gilbert & Justi, 2016; Khan, 2011; Passmore et al., 2014). Für die vorliegende Untersuchung wurde die folgende Definition gewählt:



Modellieren ist ein zielgerichteter Prozess, der die Herstellung und Anwendung von Modellen umfasst, um Erklärungen oder Vorhersagen zu einem natürlichen Phänomen zu generieren.

Definition 2: Modellieren ist ein zielgerichteter Prozess.

Eine detaillierte Operationalisierung des Modellierens für die vorliegende Untersuchung erfolgt in Abschnitt 2.4 (S. 46).

## 2.2 Modelle und das Modellieren im naturwissenschaftlichen Unterricht

In diesem Abschnitt wird die Relevanz von Modellen für das Lehren und Lernen der Naturwissenschaften dargestellt und Möglichkeiten aufgezeigt, für welche Ziele Modelle und das Modellieren im naturwissenschaftlichen Unterricht eingesetzt werden können. Anschließend werden beispielhaft aktuelle curriculare Vorgaben aus Deutschland und Nordamerika betrachtet. Schließlich werden ausgewählte empirische Arbeiten vorgestellt, die untersuchen, inwieweit Modelle und das Modellieren im Unterricht eingesetzt werden, um bewerten zu können, ob die didaktischen Möglichkeiten genutzt und curriculare Vorgaben erfüllt werden.

### 2.2.1 Möglichkeiten für den Einsatz von Modellen und des Modellierens im Unterricht

*„But models are not ‘right answers’, rather, they are the methods and the products of science, and it is quite impossible to teach and learn science without using models.”*

*(Harrison & Treagust, 2000, S. 1013)*

In diesem Zitat machen Harrison und Treagust (2000) deutlich, dass sie die Naturwissenschaften ohne Modelle für weder lehr- noch lernbar halten. Modelle und das Modellieren nehmen somit auch in der Vermittlung naturwissenschaftlicher Inhalte und Kompetenzen eine zentrale Rolle ein (Harré, 1970; Harrison, 2001; Harrison & Treagust, 2000), sodass der Umgang mit Modellen und dem Modellieren auch für Nicht-WissenschaftlerInnen als höchst relevanter Bestandteil der naturwissenschaftlichen Grundbildung angesehen werden (Acher et al., 2007; Gilbert et al., 1998; Gilbert & Osborne, 1980).



Die Naturwissenschaften sind ohne Modelle weder lehr- noch lernbar.

Kernaussage 3: Modelle sind in der Lehre unverzichtbar.

Entsprechend ihrer zentralen Rolle und Vielfalt tragen Modelle und das Modellieren im naturwissenschaftlichen Unterricht potenziell zur Förderung vielfältiger kognitiver, meta-kognitiver, sozialer und epistemischer Fähigkeiten bei und erhöhen so insgesamt den Lernerfolg (Clement, 2000; Louca & Zacharia, 2012). Hierzu gehören auch solche Fähigkeiten, die als die vier zentralen Ziele der naturwissenschaftlichen Grundbildung aufgefasst werden (Hodson, 2014; Justi & Gilbert, 2002): Wenn Modelle als Produkte der Wissenschaft im Sinne eines Mediums verstanden werden, können SchülerInnen sie im Unterricht zur Erschließung naturwissenschaftlicher Inhalte nutzen (*model-based learning*; z. B. Forbes et al., 2017; Vo et al., 2015). Wenn SchülerInnen natürliche Phänomene selbst modellieren, können sie das Modellieren als komplexe naturwissenschaftliche Denkweise erlernen (*modeling-based inquiry/reasoning*; z. B. Campbell & Neilson, 2012; Neilson et al., 2010). Wenn SchülerInnen über

## Theoretischer Rahmen

Modelle und Modellieren reflektieren, z. B. hinsichtlich der hypothetischen Annahmen die einem Modell zugrunde liegen oder den Limitationen eines Modells, kann das Verstehen der Natur der Naturwissenschaften gefördert werden (*nature of science*; Gobert & Pallant, 2004; McComas, 2002). Schließlich können SchülerInnen auch mit Modellen über naturwissenschaftliche, gesellschaftsrelevante Fragen, wie den Klimawandel, diskutieren (*model-based argumentation*; z. B. Böttcher & Meister, 2011; Zangori et al., 2017). Ein aktuelles Beispiel eines Modells und dessen vielfältiger Anwendung zeigt Abbildung 2.

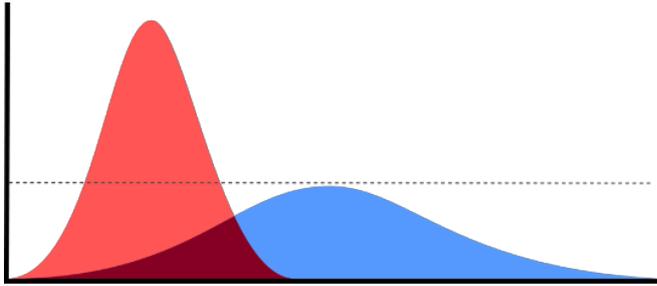


Abbildung 2: Modell zweier Covid-19 Ausbreitungsszenarien (Philips, 2020)

Das Modell, bestehend aus zwei Wachstumskurven und der Aufforderung „Flatten the curve!“, erfüllt potenziell alle vier Ziele naturwissenschaftlicher Grundbildung (Hodson, 2014; Justi & Gilbert, 2002): Es kann genutzt werden, um einen Appell an die Einhaltung der Hygienevorschriften zu *kommunizieren*, um die exponentiellen Ansteckungsmechanismen einer Pandemie zu *erklären* oder um den Effekt der Hygienevorschriften in Form einer niedrigeren Kurve *vorherzusagen*. Zudem können Handlungsempfehlungen oder politische Entscheidungen anhand des Modells *diskutiert* werden.



Modelle und das Modellieren können im naturwissenschaftlichen Unterricht zur Erreichung vielfältiger Ziele beitragen.

Kernaussage 4: Modelle haben viele Anwendungszwecke.

### 2.2.2 Curriculare Vorgaben zu Modellen und dem Modellieren

Der zentralen Bedeutung von Modellen und dem Modellieren in den Naturwissenschaften und den vielfältigen Möglichkeiten Modelle und das Modellieren im naturwissenschaftlichen Unterricht zu nutzen folgend, wird die Arbeit mit Modellen weltweit in Steuerungsdokumenten für den naturwissenschaftlichen Unterricht explizit gefordert (ACARA, 2015; BCMOE, 2019; KMK, 2005, 2020; NGSS Lead States, 2013b; VCAA, 2016). Dabei wird in den vielen<sup>3</sup> schulischen Steuerungsdokumenten ein klarer Bezug zu Modellen im Sinne epistemischer Werkzeuge und dem Modellieren im Sinne einer naturwissenschaftlichen Denkweise hergestellt. So wird das Modellieren in den deutschen Bildungsstandards für das Fach Biologie dem Kompetenzbereich „Erkenntnisgewinnung“ zugeordnet. In den Bildungsstandards für den Mittleren Schulabschluss wird hervorgehoben, dass „gerade das Modellieren bzw. kritische Reflektieren des Modells bedeutsamer Teil der naturwissenschaftlichen Erkenntnisgewinnung“ ist (KMK, 2005, S. 11). SchülerInnen der 10. Klasse sollen beispielsweise in der Lage sein, „Modelle praktisch [zu] erstellen“, „Sachverhalte mit Modellen [zu] erklären“ und „Hypothesen [zu] erstellen mit einem Modell“ (KMK, 2005, S. 17). In den Bildungsstandards für die allgemeine Hochschulreife wird von SchülerInnen weitergehend gefordert „Phänomene der Natur, der Technik und des Alltags aus naturwissenschaftlicher Perspektive zu beobachten, mithilfe zunehmend abstrakter und komplexer Modelle zu beschreiben und naturwissenschaftliche Fragestellungen aus diesen abzuleiten“, „Hypothesen zu bilden [und] diese zum Beispiel durch [...] Modelle [...] zu prüfen“ und „die Methoden der Erkenntnisgewinnung, wie zum Beispiel [...] Modelle in den Naturwissenschaften zu reflektieren und die Vor- und Nachteile, sowie die Grenzen dieser Methoden zu bewerten“ (KMK, 2020, S. 9). Dabei sollen fachspezifische Modelle charakterisiert, ausgewählt und hypothesengeleitet zur „Untersuchung von Sachverhalten“ genutzt werden (KMK, 2020, S. 15).

Auch in den wohl weltweit einflussreichsten Standards für naturwissenschaftliche Bildung, den US-amerikanischen *Next Generation Science Standards* (NGSS Lead States, 2013a, 2013b, 2017) wird die Entwicklung und Anwendung von Modellen (*developing and using models*; NGSS Lead States, 2013b) als essentielle naturwissenschaftliche Praxis beschrieben. Es werden sehr explizit epistemische Ideen zu Modellen und zum Modellieren beschrieben, die Modelle als epistemische Werkzeuge und das Modellieren als Prozess der naturwissenschaftlichen Erkenntnisgewinnung positionieren:

---

<sup>3</sup> Ein Gegenbeispiel für ein Curriculum, in welchem Modelle lediglich als Medien aufgefasst werden, liefert z. B. das Department of Basic Education Republic of South Africa (2011)

## Theoretischer Rahmen

*„Models include diagrams, physical replicas, mathematical representations, analogies, and computer simulations. Although models do not correspond exactly to the real world, they bring certain features into focus while obscuring others. All models contain approximations and assumptions that limit the range of validity and predictive power, so it is important for students to recognize their limitations. In science, models are used to represent a system (or parts of a system) under study, to aid in the development of questions and explanations, to generate data that can be used to make predictions, and to communicate ideas to others. Students can be expected to evaluate and refine models through an iterative cycle of comparing their predictions with the real world and then adjusting them to gain insights into the phenomenon being modeled. As such, models are based upon evidence. When new evidence is uncovered that the models can't explain, models are modified.“*

*(NGSS Lead States, 2013b, S. 6)*

SchülerInnen sollen beispielsweise Modelle von Originalen unterscheiden können (Jahrgangsstufen 1-2), Modelle entwickeln oder anwenden, um Phänomene vorherzusagen (ab den Jahrgangsstufen 3-5), Modelle entwickeln oder anwenden, um Daten zu generieren und damit Annahmen zu einem Phänomen zu prüfen (Jahrgangsstufen 6-8) oder mehrere verschiedene Modelle entwickeln und reflektieren und basierend auf den Vorhersagen und Limitationen der Modelle flexibel zwischen diesen wechseln (Jahrgangsstufen 9-12; NGSS Lead States, 2013b).



Curriculare Vorgaben betonen epistemische Perspektiven auf den Umgang mit Modellen und dem Modellieren.

Kernaussage 5: Curricula fordern epistemische Perspektiven auf Modelle und das Modellieren.

Insgesamt zeigen die beiden Curricula, stellvertretend für viele Curricula weltweit, somit sehr komplexe, epistemische Anforderungen an SchülerInnen im Hinblick auf den Umgang mit Modellen und dem Modellieren. Entsprechend sind auch die Lehrkräfte gefordert, Modelle und das Modellieren in diesem Sinne in ihren Unterricht zu integrieren und die curricular formulierten Ziele zu erreichen. Im folgenden Abschnitt wird daher geprüft, ob diese Vorgaben erfüllt werden, indem ausgewählte empirische Arbeiten dargestellt und zusammengefasst werden, die den tatsächlichen Einsatz von Modellen und dem Modellieren im Unterricht untersuchen.

### 2.2.3 Empirische Arbeiten zum Einsatz von Modellen und des Modellierens im Unterricht

Zum Einsatz von Modellen und des Modellierens in der Schule liegen diverse Studien vor. Während einige Studien thematisieren wie sich modell-basiertes Lernen auf den Lernerfolg in Bezug auf bestimmte Inhalte verhält (*model-based learning*; z. B. Broman et al., 2018; Maia & Justi, 2009) und damit vornehmlich Fachwissen adressieren, werden im Folgenden primär Studien dargestellt, die den curricular geforderten Umgang mit Modellen und dem Modellieren im Sinne der epistemischen Perspektive auf Modelle und das Modellieren untersuchen.

#### **Wie werden Modelle im Unterricht eingesetzt?**

Grundsätzlich weisen viele Studien zum Einsatz von Modellen im Unterricht darauf hin, dass die Integration von Modellen in den Unterricht für Lehrkräfte herausfordernd ist und sich oft auf eine mediale Nutzung von Modellen beschränkt (Krell & Krüger, 2013; T. D. Lee & Jones, 2018; Nelson & Davis, 2012; Nielsen & Nielsen, 2021a, 2021b).

In den Arbeiten von Nielsen und Nielsen (2021a, 2021b) zum praktischen Einsatz von Modellen durch dänische Lehrkräfte der Sekundarstufe, geben die Lehrkräfte in Gruppendiskussionen oder Fragebögen an, den Einsatz von Modellen vornehmlich inhaltlich zu begründen (d. h. an Fachwissen auszurichten). Modelle werden als Medien für die Kommunikation und Erklärung von Sachverhalten genutzt, die bei SchülerInnen das Verstehen von fachlichen Inhalten fördern sollen (Nielsen & Nielsen, 2021a). In der Interviewstudie geben Lehrkräfte zum Einsatz von Modellen im Unterricht zudem an, dass sie ihren SchülerInnen vor allem Instruktionen zur deskriptiven Beschreibung von Modellen geben. Gleichzeitig werden Instruktionen zum Ableiten von Vorhersagen aus dem Modell selten verwendet (Nielsen & Nielsen, 2021b). Auch Krell und Krüger (2013) untersuchten in einer Fragebogenstudie mit 146 Berliner Lehrkräften, wie Lehrkräfte selbst angeben, Modelle im Biologieunterricht einzusetzen. Ihre Ergebnisse deuten darauf hin, dass Modelle insgesamt im Biologieunterricht der befragten Biologielehrkräfte stark vertreten sind. Es werden häufig der Zweck von Modellen und weniger häufig Gründe für alternative Modelle thematisiert. Zusätzlich konnte beobachtet werden, dass Biologielehrkräfte mit zwei naturwissenschaftlichen Fächern angeben, stärker Modelle einzusetzen als solche Biologielehrkräfte mit anderen Fächerkombinationen. Zudem wirkt sich, im Gegensatz zum Schultyp, auch die Erfahrung der Biologielehrkräfte positiv auf die Stärke des Modelleinsatzes aus (Krell & Krüger, 2013). Zu ähnlichen Ergebnissen kommen auch T. D. Lee und Jones (2018). Sie untersuchten, wie und welche visuellen Modelle zum ökologischen Wasserkreislauf angehende und unterrichtende Grundschullehrkräfte für ihren Unterricht auswählen. Ihre Ergebnisse zeigen keinen Unterschied in Auswahl und Vorgehensweise der beiden Gruppen. Angehende und praktizierende Grundschullehrkräfte tendierten dazu, visuelle Modelle für den Unterricht auszuwählen, die ästhetisch ansprechend, sowie möglichst einfach gehalten sind und spezifische Elemente des Wasserkreislaufs explizit beschriftet darstellen. Visuelle Modelle, die

komplex und dynamisch sind, wurden seltener ausgewählt. Die AutorInnen kommen zu dem Schluss, dass Grundschullehrkräfte den Zweck der Visualisierung bestimmter Komponenten des Wasserkreislaufs verfolgen, statt Modelle als Werkzeuge der Erkenntnisgewinnung für die dargestellten komplexen Systeme zu nutzen (T. D. Lee & Jones, 2018). Miller und Kastens (2018) konnten in einer Interventionsstudie allerdings auch zeigen, dass der Umgang von Lehrkräften mit Modellen im Unterricht erfolgreich gefördert werden kann. Vor der Intervention nutzten die beiden in der Studie betrachteten Physiklehrkräfte Modelle in ihrem Unterricht primär als Medien zur Förderung von Fachwissen, indem SchülerInnen beispielsweise am Modell vorhandenes Wissen replizieren (vgl. Passmore et al., 2009). Nach einer Intervention, die vor allem auf eine Reflektion der eigenen Instruktionen in Bezug auf Modelle abzielte, gingen die Physiklehrkräfte vermehrt dazu über, praktisches Modellieren in ihren Unterricht zu integrieren und Modelle als Werkzeuge der Erkenntnisgewinnung zu nutzen (Miller & Kastens, 2018). Darüber hinaus zeigen die Ergebnisse einer Fallstudie von Nelson und Davis (2012), dass Grundschullehrkräfte erst mit entsprechender Förderung in der Lage sind von SchülerInnen entwickelte Modelle kriteriengeleitet und standardisiert zu bewerten. Hierbei nutzen sowohl die untersuchten Lehrkräfte in der Studie von Nelson und Davis (2012) als auch die untersuchten Lehrkräfte in einer weiteren Studie von Harlow et al. (2013) primär inhaltlich-orientierte Kriterien, z. B. die adäquate Beschriftung von Modellelementen oder die korrekte und häufige Verwendung von Fachbegriffen, zur Bewertung von Modellen, da die Lehrkräfte ihre Lehraufgabe vor allem im Sinne der Präsentation einer richtigen Lösung wahrnehmen (Harlow et al., 2013).

### **Wie wird das Modellieren im Unterricht umgesetzt?**

Hinsichtlich des Modellierens kommen Baumfalk et al. (2018) zu dem Schluss, dass bisher wenige Arbeiten unternommen wurden, um die valide und reliable Erfassung der Unterrichtspraxis zum naturwissenschaftlichen Modellieren zu ermöglichen. Die existierenden Studien deuten allerdings auch darauf hin, dass praktisches Modellieren in der schulischen Realität kaum und wenn dann nur sehr eingeschränkt umgesetzt wird (Campbell et al., 2015; Khan, 2011; Krell & Krüger, 2013; Mendonça & Justi, 2011, 2013, 2014; Nielsen & Nielsen, 2021b; Werner et al., 2019).

In der bereits beschriebenen Interviewstudie von Nielsen und Nielsen (2021b) geben Lehrkräfte an, ihre SchülerInnen selten beim Modellieren anzuleiten. Khan (2011) konnte in einer Fallstudie mit vier Lehrkräften der Naturwissenschaften detailliert zeigen, dass bestimmte Aspekte des Modellierens kaum oder gar nicht im Unterricht umgesetzt werden. So werden Modelle selten eingesetzt, um Datensätze zu vergleichen, Vorhersagen abzuleiten, Zusammenhänge zwischen Variablen des Modells herzustellen oder einzelne Variablen zu isolieren. Eine Weiterentwicklung oder Modifikation „fertiger“ Modelle, systematische und zyklische Wechsel zwischen Modellevaluation und Modellentwicklung und das Nutzen von Erklärungen zur Verbesserung der von SchülerInnen produzierten Modelle wurden gar nicht umgesetzt (Khan, 2011). Auch in der Studie von Krell und Krüger (2013) geben Lehrkräfte an, dass

insbesondere das Testen von Modellen durch die Ableitung und Überprüfung von Hypothesen im Unterricht eher selten umgesetzt wird (Krell & Krüger 2013), obwohl gerade ein Verständnis dieses Testens von Modellen als ein Schlüssel zur Förderung von Modellverstehen betrachtet wird (Krell & Krüger 2016; S. Lee & Kim 2014). Campbell et al. (2015) identifizieren fünf spezifische Anwendungsmöglichkeiten von Modellen und dem Modellieren im Unterricht (*modeling pedagogies*; Campbell et al., 2013). Sie unterscheiden zwischen explorativem Modellieren (SchülerInnen untersuchen ein existierendes Modell), expressivem Modellieren (SchülerInnen beschreiben ein Phänomen, indem sie ein Modell entwickeln oder ein existierendes Modell nutzen), experimentellem Modellieren (SchülerInnen stellen Hypothesen am Modell auf und testen diese am dazugehörigen Phänomen), evaluierendes Modellieren (SchülerInnen vergleichen mehrere Modelle eines Phänomens und wählen ein für den jeweiligen Zweck angemessenes Problem aus) und zyklischem Modellieren (SchülerInnen entwickeln, evaluieren und verbessern wiederholt ein Modell, z. B. im Rahmen eines Projekts; Campbell et al., 2015; Campbell et al., 2013). In einem Review von 81 Interventionsstudien, in denen modell-basiertes Lernen umgesetzt wurden, konnten Campbell et al. (2015) dann zeigen, dass die *modeling pedagogies* experimentelles, evaluierendes und zyklisches Modellieren, welche Modellierprozesse beschreiben, in denen Modelle kritisch reflektiert oder als Werkzeuge zur Erkenntnisgewinnung eingesetzt werden, in den untersuchten Studien am seltensten umgesetzt wurden. Exploratives und expressives Modellieren trat dagegen häufiger auf (Campbell et al., 2015). Ähnlich zu diesem Einsatz in Forschungskontexten, konnte auch in der Videostudie von Werner et al. (2019) beobachtet werden, dass Aufgabenstellungen, die Modelle als Werkzeuge zur Erkenntnisgewinnung nutzen oder eine kritische Reflexion von Modellen fordern nur selten vorkommen. Eine Vergleichsstudie zur Modellierpraxis und den dazugehörigen Subpraxen Modellkonstruktion, Modellnutzung, Modellevaluation und Modellrevision von GrundschülerInnen zwischen U.S.-amerikanischen und deutschen Grundschulklassen zeigt außerdem kleine nationale Unterschiede auf: So wurden in den observierten deutschen Unterrichtseinheiten die Modelle häufiger mit Lernendenvorstellungen über das Phänomen abgeglichen, während in den U.S.-amerikanischen Unterrichtseinheiten Modelle etwas häufiger evaluiert und überarbeitet wurden (Böschl et al., 2023). Allerdings weisen die Autoren primär darauf hin, dass insgesamt wenige Unterrichtssituationen genutzt wurden, um die verschiedenen Subpraxen des praktischen Modellierens anzuwenden und zu üben, da in den betrachteten Unterrichtseinheiten überwiegend fertige Repräsentationen zu lernender Phänomene als Modell im Unterricht eingesetzt wurden, mit dem Zweck inhaltliches Wissen zu vermitteln.



Modelle werden im Unterricht primär medial eingesetzt. Praktisches Modellieren spielt nur eine geringfügige Rolle.

Kernaussage 6: Im Unterricht werden Modelle als Medien eingesetzt.

### Was bestimmt den Einsatz von Modellen und dem Modellieren im Unterricht?

Die beschriebenen Studien zum Einsatz von Modellen und zum praktischen Modellieren im Unterricht zeigen, dass das Potenzial von Modellen und dem Modellieren für die Unterrichtsgestaltung zur Förderung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung nicht ausgenutzt wird. Fachdidaktische Zielsetzungen (Hodson, 2014; Justi & Gilbert, 2002) und curriculare Vorgaben (z. B. KMK, 2005; NGSS Lead States, 2013b) werden damit nicht erreicht. In naturwissenschaftstheoretischen und empirischen Arbeiten werden folglich verschiedene Gründe für die Nicht-Erreichung curricularer Ziele und die weiterhin primär inhalts-orientierte Unterrichtsrealität in Bezug auf Modelle und das Modellieren diskutiert: Manche AutorInnen kritisieren die weiterhin inhaltschweren Curricula und nehmen an, dass solche Inhaltsverpflichtungen und bestimmte Formen der Leistungsmessung, wie der Einsatz von Multiple-Choice Instrumenten, Lehrkräfte darin behindern, Modelle und das Modellieren angemessen im Unterricht umzusetzen (Nielsen & Nielsen, 2021b). Viele AutorInnen gehen jedoch davon aus, dass die Perspektiven der Lehrkräfte auf Modelle und ihre eigenen Fähigkeiten des Modellierens die Unterrichtsplanung und -umsetzung zentral beeinflussen (Harlow et al., 2013; Vo et al., 2015, 2019). So zeigen Harlow et al. (2013), dass nur wenige Lehrkräfte, obwohl sie ihre SchülerInnen als kreative Denker sehen, Kreativität auch als wichtige Voraussetzung für das Modellieren beschreiben; einige Lehrkräfte sehen kreative Einfälle beim Modellieren sogar als Hindernis, welches die SchülerInnen von der korrekten Lösung abhält. Vo et al. (2015) untersuchten Lehrkräfte auf ihre epistemischen Ideen zu Modellen und dem Modellieren hin. Sie konnten beobachten, dass Lehrkräfte, die bestimmte epistemische Ideen betonen, im Hinblick auf Modelle und das Modellieren besseren Unterricht planen und umsetzen. Es liegt nahe, dass Lehrkräfte selbst über die entsprechenden Fähigkeiten im Umgang mit Modellen und dem Modellieren verfügen müssen (Günther et al., 2017, 2019). Insbesondere sollten Lehrkräfte in der Lage sein, die Vorstellungen und Fähigkeiten von SchülerInnen zu Modellen und dem Modellieren im Unterricht zu diagnostizieren, sowie entsprechende Lerngelegenheiten planen und umsetzen zu können (Elby & Hammer, 2010; Sung & Oh, 2018). Kurz, Lehrkräfte sollten im Umgang mit Modellen und dem Modellieren *kompetent* sein!



Lehrkräfte müssen selbst über ausreichende Fähigkeiten, d.h. Modellierkompetenz, verfügen, um Modelle und das Modellieren, entsprechend der curricularen Forderungen, im Unterricht einsetzen zu können.

Kernaussage 7: Lehrkräfte benötigen selbst Modellierkompetenz.

## 2.3 Modellierkompetenz

Um sich den Vorstellungen und Fähigkeiten zum Umgang mit Modellen und dem Modellieren von Lehrkräften und SchülerInnen theoriebasiert nähern zu können, wird in diesem Abschnitt zuerst der allgemeine Kompetenzbegriff definiert. Anschließend wird eine erste Definition der Modellierkompetenz vorgenommen und in die professionelle Kompetenz von Lehrkräften eingeordnet, bevor im nächsten Abschnitt detailliert auf naturwissenschaftstheoretische Überlegungen und empirische Untersuchungen zur Struktur, Dimensionalität und Extensionalität der Modellierkompetenz eingegangen wird.

### 2.3.1 Exkurs zum Kompetenzbegriff

Der Kompetenzbegriff gilt in Deutschland als primärer Zielbegriff der öffentlichen Schul- und Lehrkräftebildung (BMBF, 2007). Die Kompetenzorientierung stellt daher seit dem Jahrtausendwechsel das zentrale Paradigma für das Lernen in Schul- und Arbeitskontexten dar (Dubs, 2004; Max, 1999). Als besonders etablierte Definition des Kompetenzbegriffs gilt in der naturwissenschaftsdidaktischen Forschung der Vorschlag von Weinert (2001b, S. 27–28):



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.

Definition 3: Kompetenzen nach Weinert (2001b).

#### **Kompetenzen sind situationsabhängig**

Kompetenzen sind allerdings nicht direkt beobachtbar, sondern als latente Konstrukte zu verstehen, auf die man durch die Beobachtung des Verhaltens einer Person bei der Bearbeitung einer Aufgabe (Performanz) schließen kann. Dabei gibt die jeweilige Aufgabe, Weinert beschreibt sie als „variable Situation“, den Kontext vor, an welchem die Performanz sichtbar wird. Das Kompetenzmodell von Blömeke, König et al. (2015) macht diesen Umstand besonders deutlich (Abbildung 3).

## Theoretischer Rahmen

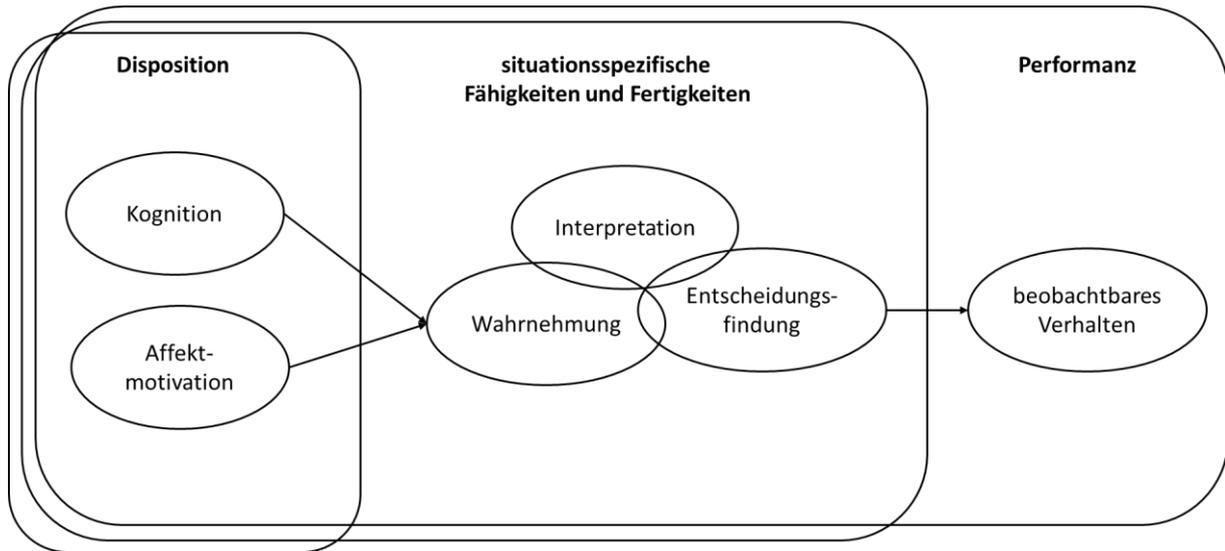


Abbildung 3: Kompetenz als Kontinuum (Blömeke, König et al., 2015, S. 7)

In diesem Kompetenzmodell wird Kompetenz als Kontinuum beschrieben. Eine Situation wird von einer Person wahrgenommen und interpretiert, wobei kognitive und affekt-motivationale Dispositionen aktiviert werden, die dann zu einer Entscheidungsfindung führen und in der Umsetzung eines beobachtbaren Verhaltens münden. Kompetenz stellt dann in soweit ein Kontinuum dar, als dass sich Performanzen auf der Basis von Merkmalen der Person (Disposition) und der Situation (Kontext<sup>4</sup>) unterscheiden (Blömeke, König et al., 2015). Dies impliziert allerdings auch, dass multiple Performanzen beobachtet werden müssen, die unterschiedliche situationsspezifische Merkmale abdecken, um valide auf die tatsächliche Kompetenz einer Person schließen zu können, selbst wenn die personenspezifischen Merkmale als stabil angenommen werden. Die theoretische Annahme, Kompetenzen seien abhängig von motivationalen, volitionalen und sozialen Faktoren und der jeweiligen variablen Situation findet sich dabei sowohl in Weinerts Kompetenzdefinition (Weinert, 2001b) als auch im Kompetenzmodell von Blömeke wieder; in der empirischen Forschung werden Kompetenzen aus pragmatischen Erwägungen allerdings oft als rein kognitive Konstrukte aufgefasst (Klieme et al., 2007). Entsprechend liegen kaum Studien zu motivationalen, volitionalen oder sozialen Faktoren vor und auch Studien zu situationsspezifischen Merkmalen sind vergleichsweise rar.

---

<sup>4</sup> In der Literatur wird der Begriff „Kontext“ nicht einheitlich verwendet; eine Einordnung des Begriffs für die vorliegende Arbeit erfolgt in Abschnitt 2.3.4.

2.3.2 Einordnung der Modellierkompetenz in die professionelle Kompetenz von Lehrkräften

Wie in Abschnitt 2.3.1 beschrieben ist die Entwicklung der zur Planung, Umsetzung und Reflexion von Unterricht benötigten professionellen Kompetenzen bei angehenden und praktizierenden Lehrkräften zentrales Ziel der Lehrkräftebildung. Diese professionellen Kompetenzen umfassen angelehnt an Shulman (1986) neben allgemein-pädagogischem Professionswissen (*pedagogical knowledge*), motivationalen Orientierungen und persönlichen Überzeugungen auch Fachwissen (*content knowledge*) und fachdidaktisches Wissen (*pedagogical content knowledge*; Abbildung 4; Baumert & Kunter, 2013).

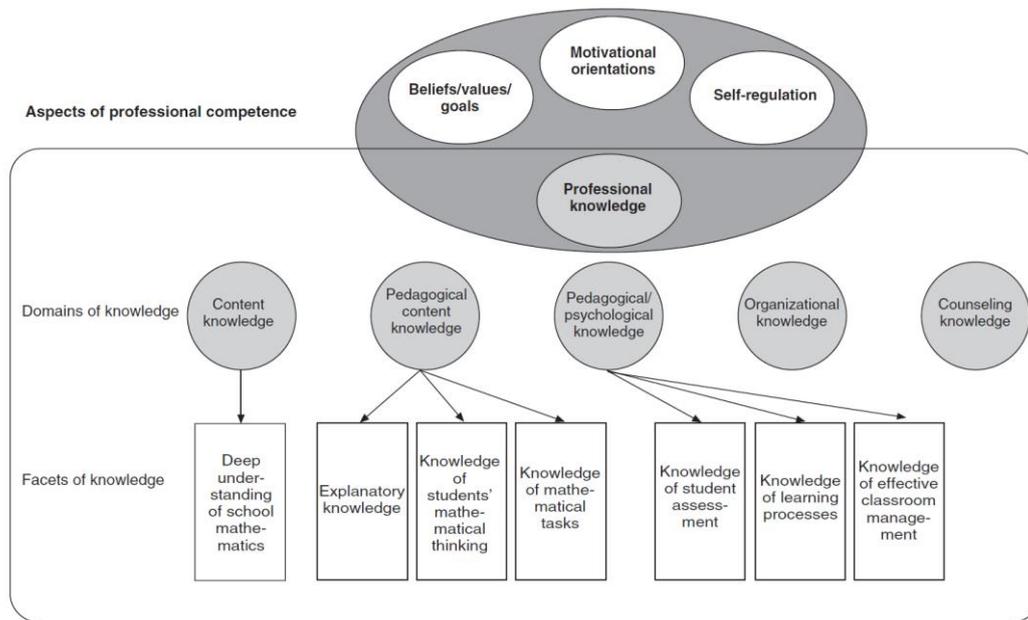


Abbildung 4: Das COACTIV-Modell zur professionellen Kompetenz von Lehrkräften am Beispiel des Fachs Mathematik (Baumert & Kunter, 2013, S. 29)

Auch Kompetenzen im Umgang mit Modellen und dem Modellieren gelten als Bestandteil der professionellen Kompetenz von Lehrkräften naturwissenschaftlicher Fächer (Günther et al., 2017, 2019; Henze et al., 2007, 2008; Justi & van Driel, 2005; Krell & Krüger, 2016; Oh & Oh, 2011). So wird Wissen zum Inhalt naturwissenschaftlicher Modelle, Wissen zum adäquaten Einsatz von Modellen im Unterricht und Wissen zu typischen SchülerInnenvorstellungen in Bezug auf Modelle, sowie die Fähigkeit im Unterricht praktisches Modellieren umzusetzen, dem fachdidaktischen Wissen von Lehrkräften naturwissenschaftlicher Fächer (*pedagogical content knowledge*) zugeordnet (Henze et al., 2007, 2008; Justi & van Driel, 2005). Als Bestandteil des Fachwissens (*content knowledge*) benötigen Lehrkräfte zudem ein generelles Verständnis über Modelle und das Modellieren in den Naturwissenschaften (Justi & van Driel, 2005; Oh & Oh, 2011). Entsprechend wird von angehenden Biologie-Lehrkräften im fachspezifischen Kompetenzprofil Biologie der ländergemeinsamen inhaltlichen Anforderungen in der Lehrkräftebildung gefordert, dass sie „mit basalen Arbeits- und Erkenntnismethoden der Biologie“

vertraut sind und „über Kenntnisse und Fertigkeiten sowohl im hypothesengeleiteten Experimentieren und Modellieren, im kriteriengeleiteten Beobachten als auch im hypothesengeleiteten Vergleichen“ verfügen (KMK, 2019, S. 22). Hierzu gehört insbesondere auch ein epistemischer Umgang mit Modellen und dem Modellieren (Upmeier zu Belzen & Krüger, 2010).

### 2.3.3 Modellierkompetenz definieren

In der deutschsprachigen Forschung hat sich für Kompetenzen im Umgang mit Modellen und dem Modellieren die Definition von Upmeier zu Belzen und Krüger (2010) etabliert, die an Weinert (2001b) angelehnt ist. Unter dem Begriff Modellkompetenz werden die Fähigkeiten, mit Modellen zweckbezogen Erkenntnisse gewinnen zu können und über Modelle mit Bezug auf ihren Zweck urteilen zu können, die Fähigkeiten, über den Prozess der Erkenntnisgewinnung durch Modelle und Modellieren in der Biologie zu reflektieren sowie die Bereitschaft, diese Fähigkeiten in problemhaltigen Situationen anzuwenden“ verstanden (S. 49.). Insbesondere die Fähigkeiten, Modelle zur Vorhersage von Daten einzusetzen und sie als hypothetische Entitäten zu verstehen, werden als bedeutsame Elemente ausgeprägter Modellkompetenz betrachtet (Gouvea & Passmore, 2017; Upmeier zu Belzen & Krüger, 2010). Um auch das abduktiv-schließende Erklären miteinzubeziehen und den Prozess des Modellierens stärker in den Fokus zu rücken, wurde dieses Kompetenzmodell der Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) von den AutorInnen erweitert und in das Kompetenzmodell der Modellierkompetenz umbenannt (Krüger & Upmeier zu Belzen, 2021). Die Erweiterung umfasst nun auch das abduktiv-schließende Erklären, bei dem kreativ modellierend eine Erklärung für ein bereits beobachtetes natürliches Phänomen abgeleitet wird<sup>5</sup>. Zudem wurde die Definition der Modellierkompetenz präzisiert (Krüger & Upmeier zu Belzen, 2021, S. 8). In der vorliegenden Untersuchung wird sich daher an der aktuellen Definition orientiert:



Modellierkompetenz umfasst die Fähigkeiten, beim Herstellen von Modellen einen theoriegeleiteten oder kreativen Erkenntnisprozess zu initiieren, bei der Anwendung von Modellen zweckbezogen Erkenntnisse zu gewinnen, über Modelle mit Bezug auf ihren Zweck zu urteilen und über den Prozess der Erkenntnisgewinnung durch Modelle und das Modellieren zu reflektieren.

Definition 4: Modellierkompetenz nach Krüger & Upmeier zu Belzen (2021)

---

<sup>5</sup> Obwohl das abduktive Schließen in der vorliegenden Untersuchung nur teilweise berücksichtigt wird, orientiert sich die vorliegende, Untersuchung aufgrund ihrer Prozessorientierung im Weiteren an dem von Krüger und Upmeier zu Belzen (2021) vorgeschlagenen Begriff der Modellierkompetenz

### 2.3.4 Dimensionalität und Extensionalität der Modellierkompetenz

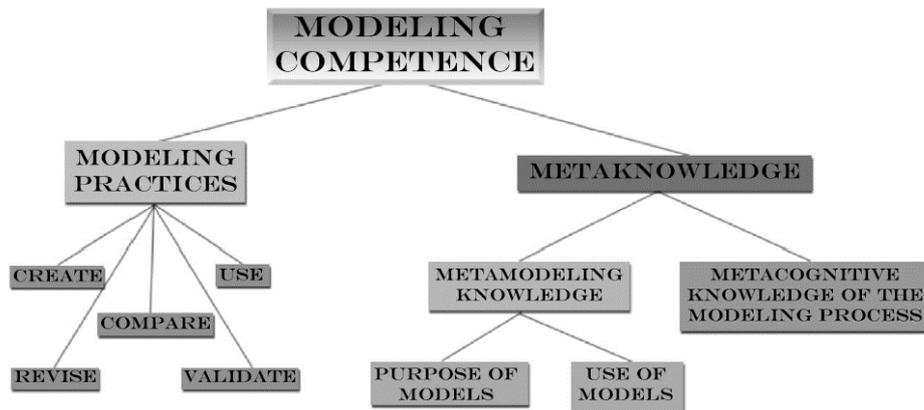
Auch Modellierkompetenz gilt, in Anlehnung an Blömeke, König et al. (2015), als latentes theoretisches Konstrukt, auf das durch eine spezifische Performanz einer Person nur indirekt geschlossen werden kann. Hinsichtlich der genauen Struktur der Modellierkompetenz liegen daher in der aktuellen naturwissenschaftsdidaktischen Forschungslage verschiedene Vorschläge vor, die sich hinsichtlich ihrer Dimensionalität und Extensionalität unterscheiden.

Im Rahmen ihrer Dimensionalität werden verschiedene Teilkompetenzen oder Dimensionen einer Kompetenz unterschieden, die zum Beispiel eher ontologische, epistemische oder prozedurale Facetten betreffen (Kind & Osborne, 2017). Durch die Untersuchung einer Kompetenz mit verschiedenen Instrumenten können einzelne Dimensionen oder Teilkompetenzen voneinander abgegrenzt oder Zusammenhänge zwischen auf den ersten Blick theoretisch unzusammenhängenden Teilkompetenzen entdeckt werden. Um eine Kompetenz, hier die Modellierkompetenz, darüber hinaus gegenüber anderen Konstrukten (bspw. anderen *styles of scientific reasoning*; Kind & Osborne, 2017) abgrenzen zu können, stellt sich auch die Frage nach der Extensionalität der Modellierkompetenz. Hierbei werden weitere, potenziell relevante Einflussfaktoren auf eine Kompetenz, bspw. andere Persönlichkeitsmerkmale (inklusive anderen Kompetenzen) oder Situationsmerkmale untersucht, die zwar nicht direkt der jeweiligen Kompetenz zugeordnet werden, aber im Zusammenhang mit der zu erfassenden Kompetenz stehen könnten. Eine solche Betrachtung der Dimensionalität und Extensionalität der Modellierkompetenz kann insbesondere dazu beitragen, das theoretische Verständnis über den Kompetenzbegriff zu erweitern (Schwichow & Nehring, 2018) und so weitere Forschungsprozesse anzuleiten (Nehring & Schwichow, 2020; Shavelson, 2013) oder passende Interventionen zu entwickeln (Schwichow & Nehring, 2018).

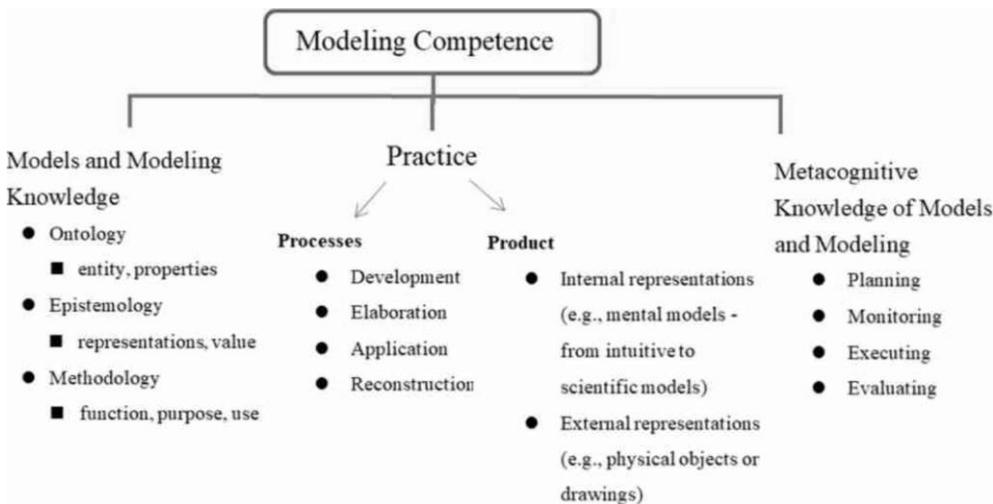
In Bezug auf ihre Dimensionalität unterscheiden gängige Kompetenzmodelle zwei oder drei Dimensionen der Modellierkompetenz (Abbildung 5; Chiu & Lin, 2019; Nicolaou & Constantinou, 2014; Nielsen & Nielsen, 2021a).

# Theoretischer Rahmen

a)



b)



c)

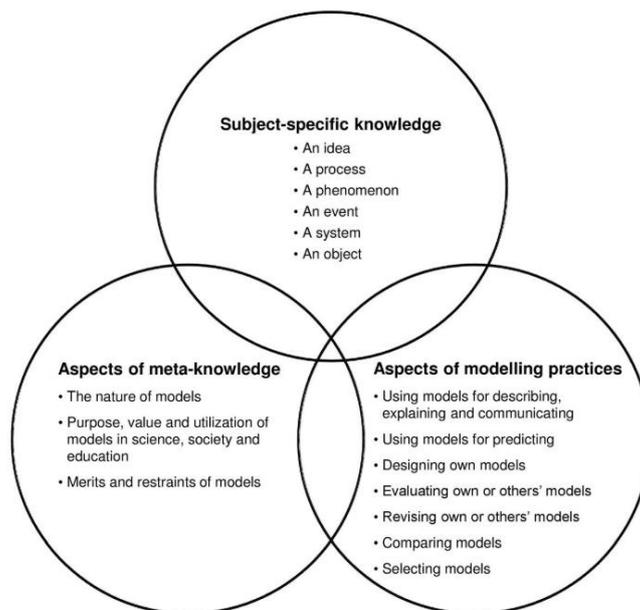


Abbildung 5: Dimensionalität der Modellierkompetenz nach a) Nicolaou und Constantinou (2014, S. 54), b) Chiu und Lin (2019, S. 3) und c) Nielsen und Nielsen (2021a, S. 6)

Wie auf der vorherigen Seite dargestellt unterteilen Nicolaou und Constantinou (2014) die Modellierkompetenz in die zwei Dimensionen *modeling practices*, und *metaknowledge* (Abbildung 5a). Nach ihrer Konzeptualisierung beschreibt die Dimension *modeling practices* die Fähigkeiten des praktischen Modellierens, welche die Entwicklung, Anwendung, Evaluation, Revision und den Vergleich von Modellen umfassen. Die Dimension *metaknowledge* beschreibt das Wissen über Modelle und das Modellieren. Dieses Wissen über Modelle und das Modellieren kann weiter aufgegliedert werden in Wissen über den Zweck und die Nutzung von Modellen (*metamodeling knowledge*) und Wissen über den Prozess des Modellierens (*metacognitive knowledge of the modeling process*; Nicolaou & Constantinou, 2014). Chiu und Lin (2019) schlagen dagegen drei Dimensionen der Modellierkompetenz vor (Abbildung 5b). Hierbei differenzieren Chiu und Lin (2019) die von Nicolaou und Constantinou (2014) vorgeschlagenen Dimensionen weiter aus. Insbesondere unterteilen sie das Wissen über Modelle und das Modellieren (*models and modeling knowledge*, statt *metamodeling knowledge*) in ontologische, epistemische und methodologische (*procedural*; vgl. Kind & Osborne, 2017) Wissensbestände und grenzen dieses eher deklarative Wissen vom metakognitiven Wissen über die Planung, Überwachung, Umsetzung und Evaluation von Modellierprozessen ab. Zusätzlich unterteilen sie auch die Praxis in einen Prozess- und einen Produktaspekt. Der Prozessaspekt umfasst die Entwicklung, Elaboration, Anwendung und Rekonstruktion von Modellen. Der Produktaspekt unterscheidet zwischen internen Repräsentationen, zu denen bspw. gedankliche Modelle und Theorien gehören, und externen Repräsentationen wie physikalischen Objekten und Zeichnungen. Auch Nielsen und Nielsen (2021a) definieren drei Dimensionen der Modellierkompetenz (Abbildung 5c), legen dabei allerdings zusätzlichen Fokus auf *subject-specific knowledge*, also Wissen über den zu modellierenden Fachinhalt.

Der Vergleich der drei dargestellten Vorschläge zur Struktur und Dimensionalität lässt viele Gemeinsamkeiten, aber auch einige Unterschiede erkennen. Die Unterschiede lassen vor allem eine trennscharfe Abgrenzung der Modellierkompetenz zu anderen Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung oder einzelnen Personen- und Situationsmerkmalen schwierig erscheinen („unscharfe Extensionalität“; Franke, 2005). Dies zeigt sich hier beispielsweise in der Einordnung von Fachwissen, welches entweder als Dimension der Modellierkompetenz (z. B. Nielsen & Nielsen, 2021a) oder als externer Einflussfaktor auf Modellierkompetenz (z. B. Chiu & Lin, 2019) verstanden werden kann. Die Rolle von Fachwissen wird dabei sowohl im Zusammenhang mit Modellen und dem Modellieren als auch im Zusammenhang mit naturwissenschaftlichem Denken allgemein diskutiert (vgl. Domänenspezifizität; Chang, 2018; Glaser et al., 1992; Hetmanek et al., 2018; Ruppert et al., 2017). Auch der in Abschnitt 2.3.1 bereits erwähnte Kontextbegriff fördert die Unschärfe der Extensionalität, da Kontextgebundenheit zwar als zentrale Eigenschaft für den Kompetenzbegriff angesehen wird (Klieme et al., 2007; Max, 1999; Rychen & Salganik, 2003), der Kontextbegriff allerdings in der naturwissenschaftsdidaktischen Forschung keineswegs eine einheitliche Bedeutung innehat (Gilbert, 2006; Weinert, 2001a). So umfasst der Kontextbegriff zum einen Beschreibungen der Situationsmerkmale, beispielsweise in Form einer

## Theoretischer Rahmen

Charakterisierung einer Aufgabenstellung oder detaillierten Deklaration einer Testsituation. Zum anderen kann er auch für die Beschreibung von Personenmerkmalen eingesetzt werden, wie beispielsweise das Fachwissen, den besuchten Studiengang oder auch komplexe Lebenssituationen (Finkelstein, 2005; Nachtigall et al., 2008). Für ein einheitliches Begriffsverständnis im Rahmen der vorliegenden Untersuchung wird Kontext hier als eine spezifische Aufgabenstellung verstanden (van Vorst et al., 2015). Der Kontext ist somit “the particular form a task takes” (Finkelstein, 2005, S. 1192). In der empirischen Forschung zur Modellierkompetenz gibt es Hinweise auf den Einfluss verschiedener Situations- und Personenmerkmale: Zu den Situationsmerkmalen gehören dabei zum Beispiel in einer Aufgabe präsentierte Modelle (Al-Balushi, 2011; Krell et al., 2012, 2014a), Vorwissen über andere Modelle (Treagust et al., 2002) oder erkenntnistheoretische Ziele in der Aufgabenstellung oder in Lehrendenimpulsen (Ke & Schwarz, 2020; Krell, 2019). Verschiedene Personenmerkmale werden ebenfalls im Zusammenhang mit Modellierkompetenz untersucht; so wird erwartet, dass allgemeine kognitive Fähigkeiten (Nehring et al., 2015), fachliches bzw. domänenspezifisches Vorwissen (Ruppert et al., 2017, 2019), Kreativität (Mierdel & Bogner, 2019) sowie Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung (Krell et al., 2014b) bei der Umsetzung von Modellierprozessen unterstützen.

Insgesamt lässt der Vergleich der drei dargestellten Vorschläge zur Struktur und Dimensionalität der Modellierkompetenz aber auch drei konsensuelle Dimensionen erkennen, die für die vorliegende Untersuchung maßgeblich sein werden:



Modellierkompetenz umfasst drei Dimensionen: Das Verständnis von Modellen und des Prozesses des Modellierens (Modellierverstehen bzw. *metamodeling knowledge*), Fähigkeiten des praktischen Modellierens (Modellierprozesse bzw. *modeling practices*), sowie die Qualität des entwickelten Modells (Modellobjekt, vgl. Mahr, 2008; bzw. *modeling product*).

Definition 5: Dimensionalität der Modellierkompetenz.

Es wird bezüglich der Dimensionen der Modellierkompetenz zumeist angenommen, dass alle drei positiv zusammenhängen; so wird beispielsweise erwartet, dass ein ausgeprägtes Modellverstehen die Umsetzung von Modellierprozessen unterstützt (Cheng & Lin, 2015; Louca & Zacharia, 2012; Schwarz et al., 2009). Im Folgenden werden die drei beschriebenen Dimensionen der Modellierkompetenz im Detail dargestellt.

## 2.4 Die Dimensionen der Modellierkompetenz in der empirischen Forschung und ihre Operationalisierung für die vorliegende Untersuchung

Im diesem Abschnitt werden, ausgehend von den drei Dimensionen der Modellierkompetenz, Modellierverstehen, Modellierprozess und Modellierprodukt, für die vorliegende Untersuchung relevante theoretische und empirische Arbeiten dargestellt und zusammengefasst. Darauf aufbauend wird für jede der drei Dimensionen eine Operationalisierung für die vorliegende Untersuchung ausgewählt.

### 2.4.1 Modell(ier)verstehen

Im Hinblick auf eine Operationalisierung des Modellierverstehens existieren verschiedene Vorschläge in der naturwissenschaftsdidaktischen Literatur. Diese definieren unterschiedliche Aspekte (Schwarz et al., 2009), Dimensionen (Crawford & Cullin, 2005) oder Themengebiete (Treagust et al., 2002) innerhalb des Modellierverstehens. Allerdings nehmen die existierenden Überlegungen zur Operationalisierung von Modellierverstehen vor allem das Wissen über Modelle (Modellverstehen oder *meta-modeling knowledge*; Nicolaou & Constantinou, 2014) in den Fokus, während das Wissen über den Prozess des Modellierens (Modellierverstehen oder *metacognitive knowledge of the modeling process*; Nicolaou & Constantinou, 2014) kaum oder nur indirekt berücksichtigt wird. Aktuell wird in der naturwissenschaftsdidaktischen Forschung daher vorgeschlagen, das Wissen über das Modellieren gesondert zu betrachten (Lazenby et al., 2020), bisher liegen jedoch keine erprobten Instrumente vor, die dies ermöglichen. In der vorliegenden Untersuchung wird entsprechend der naturwissenschaftsdidaktischen Forschungslage und in Anbetracht des Mangels an erprobten Instrumenten zur Erfassung des Modellierverstehens primär das Wissen über Modelle thematisiert. Zur deutlichen Abgrenzung werden im Folgenden die Begriffe „Modellverstehen“ und „Modellierverstehen“ genutzt.

### **Empirische Forschung zum Modellverstehen**

Zum Modellverstehen als kognitive Komponente der Modellierkompetenz liegen vergleichsweise viele Studien vor (z. B. Cheng & Lin, 2015; Crawford & Cullin, 2004; Gobert & Pallant, 2004; Grosslight et al., 1991; Grünkorn, 2014; Grünkorn & Krüger, 2012; Henze et al., 2007; Justi & van Driel, 2005; Krell & Krüger, 2013, 2016; Treagust et al., 2002; Treagust et al., 2004; van Driel & Verloop, 1999, 2002). Hinsichtlich des Modellverstehens von Lehrkräften deuten mehrere Studien darauf hin, dass angehende Lehrkräfte (Crawford & Cullin, 2005; Torres & Vasconcelos, 2015; Yenilmez Turkoglu & Oztekin, 2016), sowie praktizierende Lehrkräfte (Justi & Gilbert, 2003; Krell & Krüger, 2016; van Driel & Verloop, 1999, 2002) Modelle häufig als Repräsentationen oder direkte Kopien der Realität betrachten (vgl. naiver Realismus; Wartofsky, 1966).

Torres und Vasconcelos (2015) untersuchten beispielsweise 65 angehende portugiesische Lehrkräfte der Naturwissenschaften für die Primar- und Sekundarstufe und beobachteten sehr unterschiedliche Vorstellungen in Bezug auf Modelle und das Modellieren. Während die teilnehmenden angehenden Lehrkräfte insgesamt relativ informierte Vorstellungen zu Modellen in Bezug auf die Natur der Naturwissenschaften vorweisen konnten, gaben über ein Drittel der angehenden Lehrkräfte für die Primarstufe an, dass Modelle lediglich Repräsentationen eines Phänomens darstellen. Angehende Lehrkräfte für die Sekundarstufe gaben dagegen zumeist an, dass Modelle mit dem originalen Phänomen verglichen werden müssen, um als Referenz nutzbar zu sein. Über alle angehenden Lehrkräfte hinweg gaben zudem ein Drittel der Lehrkräfte an, dass Modelle lediglich Regelmäßigkeiten darstellen, die aus Beobachtungen der Realität abgeleitet werden, ohne einen weiteren Nutzen von Modellen zu beschreiben. Obwohl mit mehr als 80 % fast alle angehenden Lehrkräfte die Anwendung für Modelle und das Modellieren im Unterricht als gewinnbringend für das Lernen naturwissenschaftlicher Inhalte, naturwissenschaftlicher Praxis und die Natur der Naturwissenschaften einschätzten, gaben einige angehende Lehrkräfte auch an, dass Modelle ausschließlich dazu beitragen können komplexe Phänomene besser zu verstehen (Torres & Vasconcelos, 2015). Crawford und Cullin (2005) konnten im Rahmen einer Interventionsstudie mit angehenden Lehrkräften außerdem beobachten, dass das Modellverstehen der teilnehmenden angehenden Lehrkräfte sich im Rahmen ihrer Intervention zwar positiv veränderte, einige naive Vorstellungen sich jedoch hartnäckig hielten bspw. über das Verhältnis zwischen Modell und Original.

Im Hinblick auf das Modellverstehen von praktizierenden Lehrkräften konnten auch Justi und Gilbert (2003) in einer Stichprobe von 39 praktizierenden Lehrkräften der Primar- und Sekundarstufe sehr diverse Vorstellungen über Modelle feststellen. Dabei beobachteten sie vor allem inkohärente, komplexe Profile im Modellverstehen der einzelnen Lehrkräfte, die kein verallgemeinerndes Niveau ableiten lassen, da die Lehrkräfte auf einige der Fragen mit sehr informierten, auf andere Fragen mit eher naiven Vorstellungen reagierten. Darüber hinaus stellten sie eklatante Unterschiede zwischen dem Modellverstehen von Chemie- und Physiklehrkräften im Kontrast zu denen von Biologielehrkräften fest.

Biologielehrkräfte (und teilweise auch Primarstufenlehrkräfte) tendieren hierbei eher zu kohärenten uninformaten Vorstellungen über Modelle, die Modelle als einfache Repräsentation eines Phänomens verstehen. Ähnliche Ergebnisse hinsichtlich des Modellverstehens von Biologielehrkräften zeigt auch eine Studie von Krell und Krüger (2016). Hier gaben ein Großteil der 148 befragten Biologielehrkräfte zudem an, dass sie Modelle im Unterricht primär nutzen, um etwas zu zeigen oder zu erklären, wohingegen ein Vergleich mehrerer Modelle, die Evaluation eines Modells oder die Weiterentwicklung eines Modells seltener vorkommen.

Insgesamt kann damit vermutet werden, dass ein Verständnis über Modelle als epistemische Werkzeuge bei Lehrkräften in der Regel selten oder nur gering ausgeprägt ist (Gilbert und Justi 2016). Es ist somit anzunehmen, dass Lehrkräfte über ein eher eingeschränktes Modellverstehen verfügen. Diese Ergebnisse entsprechen zum größten Teil den Beobachtungen bei SchülerInnen (Grosslight et al., 1991; Grünkorn, 2014; Krell, 2013), die Modelle ebenso vorwiegend als Medien zur Veranschaulichung und Erklärung von Phänomen verstehen.

### Operationalisierung von Modellverstehen für die vorliegende Untersuchung

Modellverstehen wird in der vorliegenden Untersuchung anhand des Kompetenzmodells der Modellkompetenz (Tabelle 1) von Upmeier zu Belzen und Krüger (2010) operationalisiert, welches in der deutschsprachigen naturwissenschaftsdidaktischen Forschung als etabliert angesehen werden kann, da es bereits in verschiedenen Aufgabenformaten empirisch getestet, und zur Beschreibung der kognitiven Anteile der Modellkompetenz von SchülerInnen (z. B. Grünkorn, 2014; Krell et al., 2015; Krell et al., 2014b; Terzer et al., 2013), Studierenden (Krell & Krüger, 2017) und Lehrkräften (z. B. Krell & Krüger, 2016; Krüger & Krell, 2020) genutzt wurde.

Tabelle 1: Kompetenzmodell der Modellkompetenz (verändert nach Upmeier zu Belzen & Krüger, 2010)

Komplexität Aspekt	Niveau I	Niveau II	Niveau III
<b>Eigenschaften von Modellen</b>	Modelle sind Kopien von etwas	Modelle sind idealisierte Repräsentationen von etwas	Modelle sind theoretische Rekonstruktionen von etwas
<b>Alternative Modelle</b>	Unterschiede zwischen den Modellobjekten	Ausgangsobjekt ermöglicht Herstellung verschiedener Modelle von etwas	Modelle für verschiedene Hypothesen
<b>Zweck von Modellen</b>	Modellobjekt zur Beschreibung von etwas	Bekannte Zusammenhänge und Korrelationen von Variablen im Ausgangsobjekt erklären	Zusammenhänge von Variablen für zukünftige neue Erkenntnisse voraussagen
<b>Testen von Modellen</b>	Modellobjekt überprüfen	Parallelisieren mit dem Ausgangsobjekt, Modell von etwas testen	Überprüfen von Hypothesen bei der Anwendung, Modell für etwas testen
<b>Ändern von Modellen</b>	Mängel am Modellobjekt beheben	Modell als Modell von etwas durch neue Erkenntnisse oder zusätzliche Perspektiven revidieren	Modell für etwas aufgrund falsifizierter Hypothesen revidieren

Innerhalb des Kompetenzmodells werden fünf Aspekte des Wissens über Modelle unterschieden: „Eigenschaften von Modellen“, „Alternative Modelle“, „Zweck von Modellen“, „Testen von Modellen“ und „Ändern von Modellen“. Krell et al. (2015) heben hierbei hervor, dass diese fünf Aspekte konsensuell in den naturwissenschaftstheoretischen Überlegungen zum Wissen über Modelle berücksichtigt werden (vgl. Crawford & Cullin, 2005; Schwarz et al., 2009; Treagust et al., 2002). Zudem decken die fünf Aspekte ontologische (Eigenschaften von Modellen), epistemische (Zweck von Modellen, Alternative Modelle) und prozedurale (Testen und Ändern von Modellen) Perspektiven auf Modelle ab (vgl. Kind & Osborne, 2017). Der in der Studie von Krell und Krüger (2016) beobachtete signifikante Zusammenhang zwischen dem Modellverstehen von Biologiekraften für den Aspekt „Testen von Modellen“ und der Intensität von Lehraktivitäten mit Modellbezug, weist zudem darauf hin, dass der Aspekt „Testen von Modellen“ möglicherweise eine Schlüsselrolle im Modellverstehen einnimmt.

Für jeden dieser fünf Aspekte werden weiterhin drei Niveaustufen unterschieden. Die ersten beiden Niveaustufen werden dabei vor allem mit einer medialen Nutzung von Modellen als Anschauungs- oder Erklärobjekte verbunden (Herstellungsperspektive; Upmeier zu Belzen & Krüger, 2010), wohingegen das dritte Niveau Modelle als epistemische Werkzeuge auffasst (Anwendungsperspektive; Upmeier zu Belzen & Krüger, 2010):

- I. Modelle werden einer naiv-realistischen Perspektive folgend als direkte Kopien der Realität verstanden, wobei Eigenschaften des Modellobjekts in den Fokus rücken.
- II. Modelle werden als idealisierte Repräsentationen oder Medien verstanden, die der Visualisierung oder Erklärung dienen.
- III. Modelle werden als Werkzeug der Erkenntnisgewinnung verstanden, die zur Hypothesenbildung und Generierung von Vorhersagen über ein Phänomen eingesetzt werden.

Das Kompetenzmodell der Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) wurde inzwischen zum Kompetenzmodell der Modellierkompetenz weiterentwickelt (Krüger & Upmeier zu Belzen, 2021). In diesem weiterentwickelten Kompetenzmodell der Modellierkompetenz inkludieren Krüger und Upmeier zu Belzen (2021) zusätzlich auch das abduktive Schließen zu Niveau III bei dem anhand des Modells retrodiktiv Erklärungen über das Phänomen abgeleitet werden. Implikationen für die vorliegende Untersuchung werden in Abschnitt 6.1.4 (S. 106) diskutiert.

### 2.4.2 Modellierprozesse

Im Hinblick auf die Operationalisierung von Modellierprozessen liegen einige Vorschläge vor. Wie jede Untersuchung menschlicher Verhaltensweisen beruht auch die Operationalisierung von Modellierprozessen auf der systematischen Beobachtung von definierten Handlungselementen (Bakeman & Gottman, 1997). Eine Reduktion des menschlichen Verhaltens auf für die jeweilige Fragestellung relevante Handlungselemente, erlaubt dann die Beobachtung und Identifikation dieser Handlungselemente in ihrer sequenziellen Reihenfolge, sodass Muster erkennbar und Verhaltensweisen vergleichbar werden. Die zugrundeliegenden Handlungselemente können dabei, je nach theoretischer Konzeptualisierung, Forschungsfrage oder Zielgruppe, in unterschiedlichen Granularitäten vorliegen (Rönnebeck et al., 2016) und beispielsweise unterschiedliche Detailebenen (Chinn & Malhotra, 2002; Fischer et al., 2014; Lawson, 2004), Zeitskalen (z. B. Dalland et al., 2019) oder spezifische Arbeitsweisen (Giere et al., 2006) adressieren. Auch im Hinblick auf das Modellieren werden in der naturwissenschaftsdidaktischen Forschung bisher verschiedene Ansätze der Operationalisierung verfolgt (z. B. Löhner et al., 2005), bei denen eine subjektive Abgrenzung spezifischer Elemente des Modellierens zu unterschiedlichen Granularitäten führt (Tobin et al., 2018). Manche AutorInnen definieren vier spezifische Elemente des Modellierens z. B. *construction, use, evaluation, revision* (Cheng & Lin, 2015; Gilbert, 2004; Justi & Gilbert, 2002), andere beschränken sich auf drei z. B. *generate, evaluate, modify* (Clement, 2008; Khan, 2007; Louca & Zacharia, 2012).

Weiterhin stellen naturwissenschaftliche Modellierprozesse eine Teilmenge der verschiedenen Prozesse naturwissenschaftlicher Erkenntnisgewinnung dar (vgl. Abschnitt 2.1.2, S. 23), die wiederum dem größeren theoretischen Konstrukt der Problemlöseprozesse zugeordnet werden können (z. B. Mayer, 2007). Entsprechend können Modellierprozesse auch mit Operationalisierungen für diese beiden Konstrukte beschrieben werden. Als zentrales Rahmenwerk für die Beschreibung von naturwissenschaftlicher Erkenntnisgewinnung ist hier insbesondere die Arbeit von Klahr und Dunbar (1988) zu nennen; das anhand von Fallstudien erarbeitete Modell der *Scientific Discovery as Dual Search* beschreibt den Prozess der naturwissenschaftlichen Erkenntnisgewinnung als Suche in zwei Welten, der Hypothesenwelt und der Erfahrungswelt und findet Anklang in unzähligen weiteren Studien (z. B. Aizpurua et al., 2018; Hammann et al., 2008; Neumann et al., 2019).

#### **Empirische Forschung zu Modellierprozessen**

Reviews zur empirischen Forschungslage zu Modellen und dem Modellieren zeigen, dass die bisherige Forschung sich auf das Modellverstehen als kognitive und metakognitive Aspekte der Modellierkompetenz fokussiert (Louca & Zacharia, 2012; Nicolaou & Constantinou, 2014). Die überwiegende Anzahl von empirischen Studien im Zusammenhang mit Modellen und dem Modellieren in der Naturwissenschaftsdidaktik kann als „produktbasiert und quantitativ“ beschrieben werden (Orsenne & Upmeier zu

Belzen, 2012, S. 34). Prozessorientierte empirische Beschreibungen des Modellierens liegen dagegen bisher nur vereinzelt vor (Sins et al., 2005; Sins et al., 2009; Windschitl & Thompson, 2006). Clement und Kollegen (z. B. Clement, 2008, 2009; Clement & Rea-Ramirez, 2008; Clement & Williams, 2013) kontrastieren die modellierende Problemlösung von ExpertInnen und NovizInnen (Lehrkräften, SchülerInnen) in Bezug auf physikalische Probleme mit „*imagistic simulations*“ oder Analogien. Clement (2009) kommt zu dem Schluss, dass SchülerInnen durchaus Analogien zur Lösung naturwissenschaftlicher Probleme bilden, diese aber unsystematischer prüfen als ExpertInnen. Windschitl und Thompson (2006) beobachteten in einer Interventionsstudie allerdings auch, dass es Lehrkräften schwerfällt, selbst Modelle zu entwickeln und als Werkzeuge der Erkenntnisgewinnung zu nutzen. Während es den Lehrkräften oftmals gelang, elaboriert über gegebene wissenschaftliche Modelle zu sprechen, konnten teilnehmende Lehrkräfte theoretisches Wissen beim selbstständigen Modellieren seltener erfolgreich in ihre Modelle integrieren. Sins et al. (2005) untersuchten computerbasierte Modellierprozesse von Gruppen von SchülerInnen. Sie fanden heraus, dass erfolgreichere SchülerInnen ihre Argumentation eher auf Vorwissen stützen und Modelle holistischer betrachten als weniger erfolgreiche SchülerInnen.

Forschungsarbeiten zu den verwandten Konstrukten der naturwissenschaftlichen Erkenntnisgewinnung oder der experimentellen Problemlösung zeigen ähnliche Unterschiede zwischen ExpertInnen und NovizInnen und identifizieren darüber hinaus verschiedene Typen von Vorgehensweisen (auch Strategien). Klahr und Dunbar (1988) unterteilen beispielsweise den Typ des „*experimenters*“ und den Typ des „*theorists*“. Während Experimentierende eine gegebene naturwissenschaftliche Fragestellung untersuchen, indem sie in der Erfahrungswelt nach Regelmäßigkeiten suchen ohne Hypothesen explizit zu äußern, durchsuchen TheoretikerInnen vornehmlich die Hypothesenwelt (Klahr & Dunbar, 1988; vgl. Modellwelt, Abschnitt 2.1.2, S. 23). Sie schließen, dass ExpertInnen sich eher wie „*theorists*“ verhalten und dazu tendieren die Hypothesenwelt zu durchsuchen, was es ihnen erlaubt naturwissenschaftliche Probleme effizienter zu lösen. Klahr et al. (1993) beschreiben zudem Unterschiede zwischen den Prozessen naturwissenschaftlicher Erkenntnisgewinnung von Kindern und Erwachsenen. Hierbei konnten sie feststellen, dass es Kindern schwerfällt sich bei ihrer Suche in der Hypothesen- oder Erfahrungswelt zu beschränken, sodass sie auch irrelevante Hypothesen überprüfen oder nicht angemessene Experimente durchführen. Auch Schauble, Klopfer und Raghavan (1991) unterscheiden bei Lernenden zwei Strategien beim Experimentieren, den Ingenieurs- und den WissenschaftlerInnenmodus. Während Lernende im Ingenieursmodus so lange Variablen variieren, bis ein gewünschtes Ergebnis erzielt wird, manipulieren Lernende im fortgeschritteneren WissenschaftlerInnenmodus alle Variablen systematischer, um Zusammenhänge im gegebenen System aufzudecken. Analog zum Ingenieursmodus konnten mehrere Studien auch beim Modellieren sogenanntes *model-fitting* Verhalten beobachten. Hierbei liegt der Fokus auf dem Output eines Modells, statt auf den zu modellierenden theoretischen Beziehungen; Parameter eines Modells werden kleinschrittig angepasst, bis der Output des Modells den tatsächlichen oder erwartbaren Beobachtungen entspricht; eine tiefergehende Auseinandersetzung mit dem Modell oder

## Theoretischer Rahmen

darin enthaltenen theoretischen Elementen oder Beziehungen ist dabei nicht zwangsläufig nötig (Hogan & Thomas, 2001; Sins et al., 2005). Unterschiede in den Vorgehensweisen lassen sich möglicherweise auf Unterschiede im Vorwissen zurückführen (Rasmussen, 1981). Daneben verweisen Schauble, Klopfer und Raghavan (1991) außerdem darauf, dass erfolgreiche SchülerInnen häufiger Variablenkontrollstrategien anwenden, mehr Hypothesen generieren, systematischer Dokumentieren und zielgerichteter Arbeiten als weniger erfolgreiche SchülerInnen.

Insgesamt zeigt sich, dass zu Lehrkräften bisher kaum prozessorientierte Studien über Fähigkeiten und Vorgehensweisen beim praktischen Modellieren vorliegen (Nicolaou & Constantinou, 2014). Eine vergleichbare Typologie, wie die von Klahr und Dunbar (1988) oder Schauble, Klopfer und Raghavan (1991) liegt bisher nicht vor. Dabei können gerade solche Typologien weitere Forschungsarbeiten anleiten und die Theoriebildung fördern (z. B. Klahr & Dunbar, 1988), sowie in gezielten Interventionen zur Förderung von Modellierkompetenz eingesetzt werden (Klahr et al., 1993; Klahr & Nigam, 2004; Schauble, Glaser et al., 1991) und so die Forschung zum Modellieren maßgeblich vorantreiben.



Prozessorientierte Studien zu Fähigkeiten und Vorgehensweisen beim Modellieren von Lehrkräften fehlen.

Forschungslücke 1: Prozessorientierte Studien zum Modellieren fehlen.

**Operationalisierung von Modellierprozessen für die vorliegende Untersuchung**

Um Modellierprozesse in der vorliegenden Untersuchung möglichst genau beschreiben zu können, werden Modellierprozesse anhand des Prozessschemas naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren operationalisiert (Göhner & Krell, 2018, adaptiert nach Krell et al., 2016; Abbildung 6).

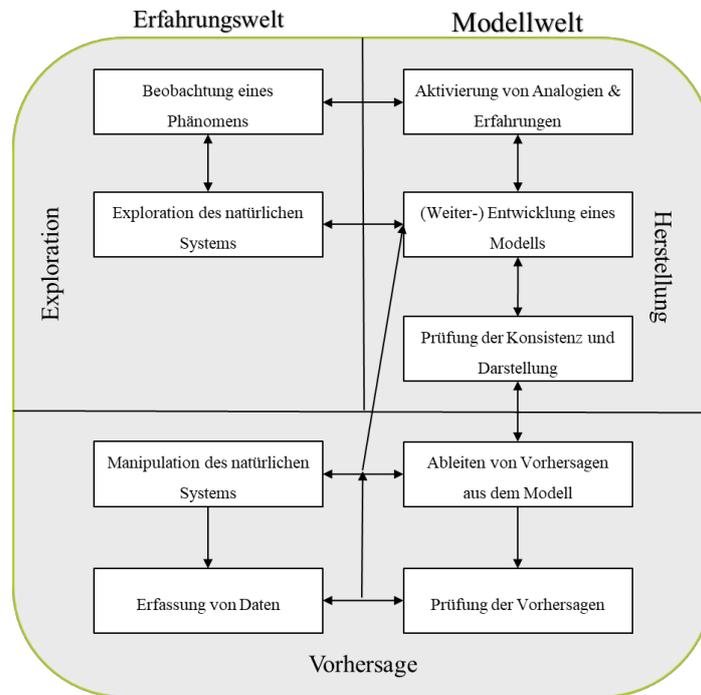


Abbildung 6: Prozessschema naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren (Göhner & Krell, 2018, adaptiert nach Krell et al., 2016). In der Abbildung sind die 3 Phasen und 9 spezifischen Handlungselemente des Modellierens dargestellt.

Das Prozessschema identifiziert auf drei Granularitätsebenen neunzehn Tätigkeiten des Modellierens, die in neun spezifische Handlungselemente und in drei größere Phasen des Modellierens unterteilt werden können (siehe auch Tabelle 8, S. 70). Modellierprozesse lassen sich mit dem Prozessschema naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren so in verschiedenen Detailgraden abbilden. Das Prozessschema naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren basiert auf verschiedenen naturwissenschaftstheoretischen und fachdidaktischen Arbeiten zum Modellieren (Clement, 1989; Giere et al., 2006) und fasst Modellierprozesse als Sequenzen von beobachtbaren Tätigkeiten des Modellierens auf. Innerhalb konkreter Modellierprozesse können Tätigkeiten des Modellierens beliebig oft und in beliebiger Reihenfolge auftreten, sodass die vorliegende Operationalisierung dem in der Literatur beschriebenen, zyklischen und dynamischen Charakter des Modellierens im Speziellen und naturwissenschaftlichen Denkens im Allgemeinen gerecht wird (Harwood, 2004; Klahr & Dunbar, 1988; Reiff et al., 2002).

### 2.4.3 Modellierprodukt

Das Ziel eines jeden Modellierprozesses ist die Entwicklung eines konkreten, sichtbaren und kommunizierbaren Artefakts (Modell bzw. Modellierprodukt). Es zeigt das Verständnis des Modellierenden über das modellierte Phänomen und kann anhand spezifischer Kriterien auf seine Qualität hin evaluiert werden. Ähnlich wie zur Frage „Was ist ein Modell?“ (vgl. Abschnitt 2.1.1, S.21) liegen auch zur Frage „Was ist ein *gutes* Modell?“ verschiedene Vorschläge vor (Hawking & Mlodinow, 2011; Jonassen, 2000; Jonassen & Ionas, 2008; Pluta et al., 2011; van der Valk et al., 2007). Dabei orientieren sich die Überlegungen zur Qualität von Modellen oft an den pragmatischen Merkmalen Zweck und Zielgruppe des Modells (Stachowiak, 1973). Werden Modelle als Repräsentationen verstanden, so sind gute Modelle dem Original möglichst ähnlich und dabei für die jeweilige Zielgruppe verständlich (Giere, 2004). Werden Modelle als epistemische Werkzeuge verstanden werden, so ergibt sich ihre Qualität aus der Nutzbarkeit für die Erreichung des jeweiligen Ziels durch eine modellierende Person. Für Hawking und Mlodinow (2011) ist ein gutes Modell beispielsweise elegant, enthält wenige beliebige oder verstellbare Elemente, bestätigt alle bisherigen Beobachtungen und lässt vor allem detaillierte Vorhersagen über zukünftige Beobachtungen zu, die das Modell falsifizieren können. Auch Pluta et al. (2011) formulieren fünf ähnliche epistemische Kriterien: Gute Modelle sind inhaltlich kohärent und klar verständlich, kompatibel mit Theorien anderer Forschungsfelder, angemessen sparsam, konsistent mit empirischen Evidenzen und wurden in der Vergangenheit für die korrekte Vorhersage neuer Beobachtungen eingesetzt.

Diese epistemischen Überlegungen zusammenfassend, sind *gute* Modelle also Modelle mit denen NaturwissenschaftlerInnen *gut Modellieren* können. In Lernkontexten sind *gute* Modelle darüber hinaus, Modelle, die die Gedanken der Modellierenden externalisieren und ihnen dabei helfen, Komponenten der theoretischen Annahmen des Modells zu visualisieren und zu prüfen (Jonassen, 2000; Jonassen & Ionas, 2008). In naturwissenschaftsdidaktischen Forschungskontexten werden Modelle daher auch oft anhand inhaltlich orientierter Kriterien evaluiert, die die Modelle auf die Integration spezifischer Komponenten und Zusammenhänge hin überprüfen (Chang et al., 2020). Chiu und Lin (2019) nutzen den Begriff des Modellierprodukts (*modeling product*) und verdeutlichen damit das Modellierprodukt als externalisiertes Ergebnis des Modellierprozesses, während sie gleichzeitig eine Abgrenzung zu den im Modellbegriff enthaltenen mentalen Modellen vornehmen.

#### **Empirische Forschung zu Modellierprodukten**

Studien die explizit Modellierprodukte thematisieren sind selten (Ausnahmen sind Cheng & Lin, 2015; Cheng et al., 2021), allerdings liegen einige Studien vor, die Modellierprodukte evaluieren, um indirekt auf Modellierverstehen oder Modellierprozesse zu schließen (Bamberger & Davis, 2013; Bielik et al., 2018; Cheng & Lin, 2015; Ergazaki et al., 2007; Schwarz et al., 2009). Darüber hinaus existieren im Bereich *model-based learning* einige Studien, die Modellierprodukte evaluieren, um beispielsweise den

Erfolg einer Intervention zu messen (Capps & Shemwell, 2020). Entsprechend merken Chiu und Lin (2019) an, dass Modellierprodukte in der naturwissenschaftsdidaktischen Forschung zumeist eine untergeordnete Rolle spielen:

*“[T]here is a lack of deep discussions on the topic of modeling products. As part of its definition of modeling competence, only a limited number of studies included modeling ‘products’[...] or even assessment of quality of the products in the modeling practices”*

*(Chiu & Lin, 2019, S. 2).*

Die Arbeiten von Cheng und Kollegen (Cheng & Lin, 2015; Cheng et al., 2021) gehören zu den wenigen Studien, die Modellierprodukte direkt adressieren und als eigene Dimension der Modellierkompetenz auffassen. Sie konnten zeigen, dass die meisten SchülerInnen Modellierprodukte zu Magnetismus entwickeln, die alleinig auf ihren Observationen beruhen und keine nicht-sichtbaren Elemente enthalten. Nur einige wenige SchülerInnen, die ihren eigenen Lernerfolg in Physik und ihr Interesse an Wissenschaft als hoch einschätzten, konnten fachlich elaboriertere Modellierprodukte zu Magnetismus entwickeln, die eine höhere inhaltliche Kohärenz aufwiesen, aber immer noch keine wissenschaftlich adäquaten Modellierprodukte darstellten. In einer weiteren Studie untersuchten Cheng et al. (2021) die Konstruktion und Evaluation von Modellierprodukten und das Modellverstehen von 95 Lehrkräften und 608 SchülerInnen der neunten Klasse. Sie konnten zeigen, dass etwa ein Drittel der Lehrkräfte in der Lage waren ein fachlich korrektes, wissenschaftliches Modell zu Magnetismus zu entwickeln und dieses auch anhand epistemischer Kriterien zu evaluieren. Innerhalb dieser Gruppe von Lehrkräften war zudem das Verstehen von Modellen im Sinne hypothetischer Artefakte, die nicht genau dem Original entsprechen müssen, am höchsten. Interessanterweise konnten sie auch beobachten, dass, entgegengesetzt zu den Lehrkräften, diejenigen SchülerInnen bessere Modellierprodukte entwickeln, die Modelle eher als Kopien eines Originals verstehen (Cheng et al., 2021). Insgesamt wird davon ausgegangen, dass Experten organisiertere und elaboriertere Modellierprodukte entwickeln als Novizen, die sich zudem durch fachliche Korrektheit und inhaltliche Kohärenz auszeichnen (Al-Balushi, 2009; Hsu et al., 2012).

### **Operationalisierung von Modellierprodukten für die vorliegende Untersuchung**

In der vorliegenden Untersuchung werden Modellierprodukte angelehnt an Chiu und Lin (2019) und Jonassen und Ionas (2008) und unter Einbezug der hier verwendeten Definitionen von Modellen und dem Modellieren als die final in einem Modellierprozess externalisierten Modelle operationalisiert. Diese können auf ihre Qualität hin untersucht werden, indem die integrierten Komponenten und Zusammenhänge geprüft werden, die nötig sind, um das zugehörige Phänomen abduktiv zu erklären oder deduktiv Vorhersagen über das Phänomen entwickeln zu können. Zudem können die Modelle auf ihre Komplexität, also die Anzahl der integrierten Komponenten und Zusammenhänge, hin geprüft werden (Bielik et al., 2018; Eidin & Bielik, 2020).

### 3. Ziele der Untersuchung

Zusammenfassend lässt sich festhalten, dass bisher keine Studien vorliegen, die die Entwicklung von Modellkompetenz bei angehenden Lehrkräften über den Studienverlauf hinweg systematisch untersuchen und feststellen, inwieweit curriculare Forderungen in Bezug auf das Modellieren (KMK, 2019) in der universitären Lehrkräftebildung erreicht werden. Die naturwissenschaftsdidaktische Forschung fokussiert sich bisher auf die Erfassung und Beschreibung der meta-kognitiven Dimension des Modellverstehens (z. B. Krell et al., 2015; Schwarz & White, 2005) sowie das Modellieren in Gruppen (z. B. Mendonça & Justi, 2011, 2013, 2014). Während für übergeordnete Konstrukte wie Problemlösen (z. B. Funke, 2011) oder auch allgemeine naturwissenschaftliche Erkenntnisgewinnung (z. B. Klahr & Dunbar, 1988), sowie das Experimentieren (Hammann et al., 2008; Schauble, Glaser et al., 1991) Untersuchungen vorliegen, die individuelle Prozesse und Strategien identifizieren und Typologien ableiten, fehlen entsprechende Untersuchungen für das Modellieren (Forschungslücke 1; Nicolaou & Constantinou, 2014). Die vorliegende Untersuchung nimmt daher eine empirische Beschreibung individueller Modellierungsstrategien in den Fokus, indem fallbasiert und prozessorientiert eine Typologie von Modellierungsstrategien Lehramtsstudierender entwickelt wird. Die Ergebnisse der Untersuchung sollen die bislang vorwiegend produktbasierten Kenntnisse naturwissenschaftsdidaktischer Forschung zu Modellen und dem Modellieren um eine Prozessfacette erweitern. Zusammen mit der gewählten und im Folgenden beschriebenen Methodik ergibt sich aus der Forschungslücke die für die vorliegende Untersuchung leitende Forschungsfrage:

**„Welche Modellierungsstrategien zeigen Lehramtsstudierende während der Problemlösung an einer Blackbox?“**

Die Möglichkeit der Erkennung individueller Modellierungsstrategien kann langfristig zur Diagnose und Förderung der Modellierkompetenz von Lehramtsstudierenden und Lehrkräften eingesetzt werden (z. B. Cheng et al., 2017; Gräsel & Trempler, 2017; Tempel, 2017) und so dazu beitragen, die Standards zu erreichen, die in den ländergemeinsamen inhaltlichen Anforderungen an die Lehrerbildung formuliert sind (KMK, 2019). Die vorliegende Untersuchung leistet so einen Beitrag zur Erweiterung des naturwissenschaftsdidaktischen Kenntnisstands (Louca & Zacharia, 2012).

Als sekundäres Ziel sollen darüber hinaus Erkenntnisse über die Dimensionalität und Extensionalität der Modellierkompetenz gewonnen werden, indem Zusammenhänge zwischen den identifizierten Modellierungsstrategien der Lehramtsstudierenden mit den beiden anderen Dimensionen der Modellierkompetenz, Modellverstehen und Modellierprodukt, und den Personenmerkmalen „Kompetenzen des naturwissenschaftlichen Denkens“ und „allgemeine kognitive Fähigkeiten“ untersucht werden.

## 4. Methodischer Zugang

Um dem Ziel einer fallbasierten und prozessorientierten Analyse der Modellierungsstrategien von Lehramtsstudierenden gerecht zu werden, wurden in der vorliegenden Untersuchung primär qualitative Methoden eingesetzt. Da qualitative Methoden im Vergleich zu quantitativen Methoden stark auf einer interpretativen Sinnkonstruktion beruhen (Patton, 2001) und sich darüber hinaus durch eine besonders hohe Vielfalt auszeichnen (Howe & Eisenhart, 1990), liegen bisher weder allgemeingültige Gütekriterien noch qualitätssichernde Maßnahmen für die qualitative fachdidaktische Forschung vor (Flick, 2010; Steinke, 2013), die als etabliert angenommen werden können. Zudem weisen Studien auf eine mangelnde Replizierbarkeit von Forschungsergebnissen in den Sozial- und Erziehungswissenschaften hin (Camerer et al., 2018; Makel & Plucker, 2014). In der naturwissenschaftsdidaktischen Forschung, die sich als Fachbereich zunehmend professionalisiert (Leuders, 2015), findet daher vermehrt eine Auseinandersetzung mit methodologischen Fragen in Bezug auf qualitative Verfahren statt (Gropengiesser, 2005; Krüger & Riemeier, 2014).

Um die Qualität der Untersuchung sicherzustellen, fand auch in der vorliegenden Untersuchung eine vertiefte methodologische Auseinandersetzung, insbesondere mit der zentralen Methode der qualitativen Inhaltsanalyse, statt. Hierzu wird zuerst die Methode der qualitativen Inhaltsanalyse beschrieben. Anschließend wird die im Rahmen der Untersuchung entstandene methodologische Publikation „Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien“ (Göhner & Krell, 2020) zusammengefasst. Das Review zeigt auf, welche methodenspezifischen Gütekriterien und Maßnahmen der Qualitätssicherung in empirischen Arbeiten, veröffentlicht in der *Zeitschrift für Didaktik der Naturwissenschaften*, beschrieben werden. Die identifizierten Gütekriterien und Maßnahmen der Qualitätssicherung werden dann entsprechend genutzt, um die Qualität der vorliegenden Untersuchung methodenspezifisch zu sichern und transparent zu machen (vgl. *methodological awareness*; Seale, 2002). Durch diese vertiefte methodologische Auseinandersetzung, kann die vorliegende Untersuchung außerdem als erprobte Forschungspraxis aufgefasst werden (Helsper et al., 2001), die zur weiteren Entwicklung und Etablierung der Naturwissenschaftsdidaktiken als empirisch orientierte Disziplinen beiträgt (Leuders, 2015).

Anschließend wird die Methode der Typologisierung, operationalisiert nach dem Modell empirisch begründeter Typenbildung (Kluge, 2000), detailliert ausgeführt, bevor die weiteren methodologischen Entscheidungen für die vorliegende Untersuchung beschrieben werden. Hierzu gehören die Stichprobe sowie die Datenaufnahme und -auswertung. Datenaufnahme und -auswertung werden dabei nach den drei Dimensionen der Modellierkompetenz gegliedert.

## 4.1 Qualitative Inhaltsanalyse

Qualitativ inhaltsanalytische Verfahren (Mayring, 2019; Schreier, 2014), die oft zusammenfassend als „die qualitative Inhaltsanalyse“ bezeichnet werden, sind in der sozialwissenschaftlichen, erziehungswissenschaftlichen und naturwissenschaftsdidaktischen Forschung etabliert (Mayring, 2015). Allen qualitativ inhaltsanalytischen Verfahren ist gemein, dass sie auf die „Systematisierung und Interpretation von manifesten und latenten Kommunikationsinhalten unter Verwendung eines Kategoriensystems“ zurückgreifen (Stamann et al., 2016, S. 13). Allerdings unterscheiden sich qualitativ inhaltsanalytische Verfahren in der praktischen Umsetzung mitunter erheblich, weswegen verschiedene Vorschläge zur Systematisierung qualitativ inhaltsanalytischer Verfahren vorliegen, die mitunter lebhaft diskutiert werden (Mayring, 2019; Mayring & Gläser-Zikuda, 2005; Schreier, 2014). Grundsätzlich lassen sich zwei Basisformen der qualitativen Inhaltsanalyse unterscheiden, die als „inhaltlich-strukturierend“ oder „extrahierend“ beschrieben werden. Zusätzlich unterscheidet sich, abhängig vom Forschungsgegenstand, das Verhältnis von deduktiv und induktiv abgeleiteten Kategorien zwischen den verschiedenen qualitativ inhaltsanalytischen Verfahren (Kuckartz, 2016; Mayring, 2015; Schreier, 2012). Stamann et al. (2016) schlagen daher vor, bei jeder qualitativen Inhaltsanalyse bestimmte Bezugspunkte zu beachten. Hierzu gehören die Basisform, die Art der Kategorien, die Art des Kategoriensystems und die Art der Kategorienbildung, sowie die weitere Vorgehensweise mit den qualitativ inhaltsanalytisch gewonnenen Daten. In der vorliegenden Untersuchung wurden insgesamt sechs qualitativ inhaltsanalytische Verfahren umgesetzt (Tabelle 2; Bezugspunkte nach Stamann et al., 2016;).

Tabelle 2: Bezugspunkte, der in dieser Untersuchung durchgeführten, qualitativ inhaltsanalytischen Verfahren nach Stamann et al. (2016)

<b>Qualitative Inhaltsanalyse</b>	<b>Basisform</b>	<b>Art der Kategorien</b>	<b>Art des Kategoriensystems</b>	<b>Modi der Kategorienbildung</b>	<b>Weitere Vorgehensweise</b>
Tätigkeiten des Modellierens	inhaltlich-strukturierend	inhaltlich	linear hierarchisch	deduktiv-induktiv	Typologisierung
Hindernisse beim Modellieren	inhaltlich-strukturierende	inhaltlich	linear hierarchisch	induktiv-deduktiv	Korrelation zu Tätigkeiten des Modellierens
Modellverstehen während des Modellierprozesses	inhaltlich-strukturierende	inhaltlich und skaliierend	statisch	deduktiv	quantitativ-statistisch
Volition beim Modellieren	inhaltlich-strukturierende	inhaltlich	linear hierarchisch	deduktiv-induktiv	Korrelation zu Tätigkeiten des Modellierens
Gütekriterien und Maßnahmen der Qualitätssicherung bei der qualitativen Inhaltsanalyse	inhaltlich-strukturierende	inhaltlich	linear hierarchisch	deduktiv-induktiv	quantitativ-statistisch
Modellierprodukte	inhaltlich-strukturierende	inhaltlich und skaliierend	statisch	induktiv	quantitativ-statistisch

## Methodischer Zugang

Ferner liegen für die Planung, Umsetzung, Rezeption, Einordnung oder externe Begutachtung (z. B. im Rahmen von *Peer-Review*-Verfahren) der qualitativen Inhaltsanalyse bisher allerdings keine eindeutig definierten Qualitätskriterien vor. Während einige AutorInnen argumentieren, die qualitative Inhaltsanalyse gehe aus einer quantitativen Tradition hervor, sodass klassische Gütekriterien wie Reliabilität und Validität bei deren Umsetzung angelegt werden sollten (Krippendorff, 1980; Mayring, 2019; Schreier, 2012), beschreiben andere AutorInnen (Mayring, 2002) methodenspezifische Gütekriterien und Maßnahmen der Qualitätssicherung, wie beispielsweise die Verfahrensdokumentation, argumentative Interpretationsabsicherung, Regelgeleitetheit, Nähe zum Gegenstand, oder kommunikative Validierung. Wieder andere AutorInnen (Lamnek, 2010) schlagen Gütekriterien wie Offenheit, Kommunikativität, Naturalistizität und Interpretativität an, die im aktuellen Diskurs zu Gütekriterien qualitativer Forschung allgemein zu verorten sind (Denzin, 2016). Um daher im Rahmen der vorliegenden Untersuchung passende Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigen zu können, wurde ein Review-Artikel erarbeitet, der Gütekriterien und Maßnahmen der Qualitätssicherung bei der Umsetzung von qualitativ inhaltsanalytischen Verfahren in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beleuchtet (Göhner & Krell, 2020). Dieser wird im folgenden Abschnitt zusammengefasst.

## 4.2 Publikation 1: Methodologischer Exkurs zu Gütekriterien und Maßnahmen der Qualitätssicherung bei qualitativ inhaltsanalytischen Verfahren

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>. Lizenziert nach [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

### **Zusammenfassung**

Um im Rahmen der vorliegenden Untersuchung passende Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigen zu können (vgl. Abschnitt 4.1 S.54), wurde in dieser Publikation ein Review erarbeitet, welches Gütekriterien und Maßnahmen der Qualitätssicherung bei der Umsetzung von qualitativ inhaltsanalytischen Verfahren in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beleuchtet. Hierzu wurden im Sinne eines systematischen Reviews (Bennett et al., 2005) 50 Artikel untersucht, die in der *Zeitschrift für Didaktik der Naturwissenschaften* veröffentlicht wurden und eine qualitative Inhaltsanalyse umsetzen. Es wurden die folgenden Fragestellungen untersucht: (1) Welche Gütekriterien werden für die qualitative Inhaltsanalyse in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschrieben? (2) Welche Maßnahmen zur Qualitätssicherung von qualitativen Inhaltsanalysen werden in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschrieben? (3) Inwieweit werden für die qualitative Inhaltsanalyse in der deutschsprachigen naturwissenschaftsdidaktischen Forschung einzelne Maßnahmen der Qualitätssicherung konsistent bestimmten Gütekriterien zugeordnet? Methodisch wurde hier ebenfalls eine qualitative Inhaltsanalyse umgesetzt, die dem „inhaltlich-strukturierenden“ Basisverfahren zugeordnet werden kann und deduktiv-induktiv gebildete, inhaltliche Kategorien in einem statischen Kategoriensystem enthält (vgl. Stamann et al., 2016). Für die deduktive Kategorienbildung wurden alle Texte als theoretische Grundlage herangezogen, die innerhalb der 50 untersuchten Artikel zur Beschreibung der qualitativen Inhaltsanalyse zitiert werden. Diese deduktiv entwickelten Kategorien wurden dann induktiv am Material erweitert und ausdifferenziert. Die angelegten Gütekriterien und umgesetzten Maßnahmen der Qualitätssicherung orientieren sich an der in Abschnitt 4.1 (S. 54) in Tabelle 22 dargestellten Zuordnung und sind in der Publikation detailliert ausgeführt.

### **Ergebnisse und Diskussion**

Im Rahmen dieser vertieften Auseinandersetzung mit der Umsetzung qualitativ inhaltsanalytischer Methoden in den untersuchten Artikeln wurden elf Gütekriterien (Forschungsfrage 1) und 16 Maßnahmen der Qualitätssicherung (Forschungsfrage 2) identifiziert (Tabelle 3). Es werden hierbei vorwiegend die klassischen Gütekriterien Validität, Reliabilität und Objektivität in den untersuchten Artikeln beschrieben, dessen häufiges Auftreten auf eine zunehmende Etablierung methodologischer Positionen in der

## Methodischer Zugang

Naturwissenschaftsdidaktik (Mayring, 2015; Schreier, 2012) oder auf die vermehrte Nutzung integrativer Ansätze hinweist, die quantitative und qualitative Methoden zusammenführen (Kelle, 2008). Hinsichtlich Maßnahmen der Qualitätssicherung werden die Modi der Kategorienbildung und die Interrater-Übereinstimmung am häufigsten beschrieben, wohingegen die Prüfung der Intrarater-Übereinstimmung in den untersuchten Artikeln keine Rolle spielt.

Tabelle 3: Übersicht über die in den untersuchten Artikeln beschriebenen Gütekriterien und Maßnahmen der Qualitätssicherung. Das vollständige Kategoriensystem inklusive Kategorienbeschreibung und Kodierhinweisen ist im Anhang der Publikation (Göhner & Krell, 2020) beschrieben.

<b>Gütekriterium</b>	<b>Anzahl der Artikel, die das Gütekriterium beschreiben</b>
Validität/Gültigkeit	15
Reliabilität	11
Objektivität/Intersubjektivität	11
Nähe zum Gegenstand/Authentizität/Naturalizität	5
Praktikabilität/Ökonomie	5
Offenheit	3
Praktische Relevanz	3
Transparenz	3
Ethische Überlegungen	2
Kommunikativität	2
Interpretativität	1
<b>Maßnahmen der Qualitätssicherung</b>	<b>Anzahl der Artikel, die die qualitätssichernde Maßnahme beschreiben</b>
Induktive Bildung des Kategoriensystems	29
Deduktive Bildung des Kategoriensystems	28
Prüfung der Interrater-Übereinstimmung	21
Dokumentation	12
Systematisches Vorgehen	12
Theoretisches Sampling	12
Kommunikative Validierung	10
Triangulation	10
Diskussion der Kodierungen	9
Diskussion des Kategoriensystems	9
Konsensbildung	7
Argumentative Interpretationsabsicherung	5
Formale Aspekte des Kategoriensystems	5
Training der BeurteilerInnen	4
Auswahl/Standardisierung des Settings	1
Prüfung der Intrarater-Übereinstimmung	0

## Methodischer Zugang

Betrachtet man das Publikationsjahr der untersuchten Artikel und die Zahl der durchschnittlich beschriebenen Maßnahmen zur Qualitätssicherung, wird ein signifikanter positiver Zusammenhang deutlich (Abbildung 7). Auch die Zuordnung der Maßnahmen zu Gütekriterien nimmt über die Jahre hinweg zu, ist in den untersuchten Artikeln allerdings inkonsistent. Diese Unterschiede hinsichtlich der Zuordnung der beschriebenen Maßnahmen der Qualitätssicherung zu den Gütekriterien findet sich allerdings auch in der Literatur wieder (vgl. "Objektivität"; Mayring, 2015; Schreier, 2012).

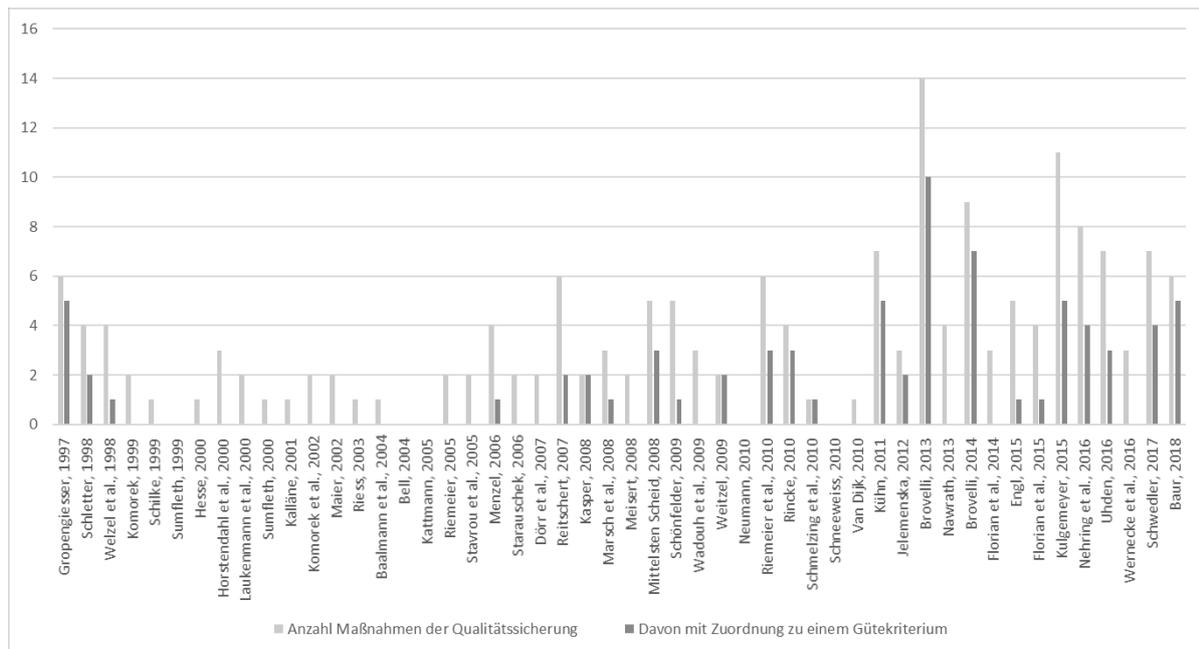


Abbildung 7: Anzahl der beschriebenen Maßnahmen der Qualitätssicherung in den untersuchten Artikeln, sowie Anteil der beschriebenen Maßnahmen, die einem Gütekriterium zugeordnet werden.

Insgesamt gibt die Publikation damit einen Überblick über etablierte Gütekriterien und Maßnahmen der Qualitätssicherung bei der Umsetzung qualitativ inhaltsanalytischer Verfahren im Kontext der deutschsprachigen naturwissenschaftsdidaktischen Forschung. Neben der Nutzung der Ergebnisse in der vorliegenden Untersuchung als Ansatzpunkt für anzulegende Gütekriterien und Maßnahmen der Qualitätssicherung, wurden auch Punkte für die weitere methodologische Diskussion in den Fachdidaktiken abgeleitet, die zur weiteren Etablierung und Systematisierung qualitativ inhaltsanalytischer Verfahren beitragen möchten: Im Sinne eines „*argument-based approach to validation*“ (Kane, 2013) sollte die Qualität von Forschungsergebnissen im Einzelfall und unter Berücksichtigung der intendierten Interpretation bewertet werden. Gerade die Umsetzung von Maßnahmen der Qualitätssicherung kann aufgrund der Vielfalt und interpretativen Komplexität qualitativer Methoden (Howe & Eisenhart, 1990; Patton, 2001) nicht routiniert nach eindeutigen Regeln erfolgen (Hartig & Frey, 2012), sondern sollte als theoriegeleiteter und komplexer Prozess verstanden werden, bei dem Maßnahmen der Qualitätssicherung abhängig vom Kontext der Untersuchung und der erwarteten Evidenz für die intendierte Interpretation der Ergebnisse ausgewählt und umgesetzt werden.

### 4.3 Typologisierung

In der empirischen Sozialforschung gilt die Typologisierung (auch Typenbildung) als etablierte Methode der Klassifikation (Collier et al., 2012; Kelle & Kluge, 2010; Kuckartz, 2010; Lamnek, 2005). Aus dem Griechischen stammend, beschreibt der Typenbegriff (*typos*) ein „Muster“ oder eine „Grundform“ einer Gruppe von Objekten, die „übrigbleibt, wenn alles Individuelle, Persönliche oder Ideografische abgezogen wird.“ (Kuckartz, 2010, S. 555). Dabei bringt eine Typologisierung von Objekten verschiedene Vorteile mit sich, beispielsweise können Objekte systematisch und komplexitätsreduziert beschrieben oder Gemeinsamkeiten und Unterschiede zwischen mehreren Objekten identifiziert werden (Bailey, 1994). Zudem können „Idealtypen“ gebildet werden, die eine bewusste Hervorhebung oder Vernachlässigung bestimmter Attribute vornehmen, die durch Forschende als relevant oder weniger relevant eingeschätzt werden, und so als theoretische Konstrukte oder epistemische Artefakte<sup>6</sup> in der weiteren Forschung Anklang finden können (Kuckartz, 2010). Trotz ihres Status als etablierte Methode, galt die Typologisierung lange als „old-fashioned“ und „unsophisticated“ (Collier et al., 2012, S. 217), was von Kuckartz (2010) unter anderem auf eine unsystematische, wenig transparente Umsetzung in empirischen Arbeiten zurückgeführt werden kann, die er als „Common-Sense-Verfahren“ bezeichnet. Inzwischen liegen zur systematischen, methodisch kontrollierten Umsetzung von Typologisierungen allerdings verschiedene Vorschläge vor (Kelle & Kluge, 2010; Kluge, 2000; Kuckartz, 2010). Diese beschreiben die Typologisierung einheitlich als die Zusammenfassung von Objekten zu Typen, die sich untereinander möglichst stark ähneln (interne Homogenität; Kluge, 2000) und sich mit Objekten, die anderen Typen zugeordnet wurden, möglichst stark unterscheiden (externe Heterogenität; Kluge, 2000). Im Gegensatz zur Taxonomisierung, die neben der Typologisierung die zweite Form der Klassifizierung darstellt, basiert eine Typologisierung dabei auf einem multidimensionalen Vergleich zweier Objekte, bei dem diese anhand mindestens zweier Attribute verglichen werden (Bailey, 1994).

---

<sup>6</sup> Damit sind „Idealtypen“ natürlich auch Modelle, was sich insbesondere in den Begriffen „Modelltyp“ und „Modellfall“ wiederfindet, z. B. beschrieben in Kelle und Kluge (2010),

### Das Modell der empirisch begründeten Typenbildung

Als methodisch flexibler und trotzdem systematischer Ansatz wurde für die vorliegende Untersuchung das Modell der empirisch begründeten Typenbildung ausgewählt (Kluge, 2000; Abbildung 8).

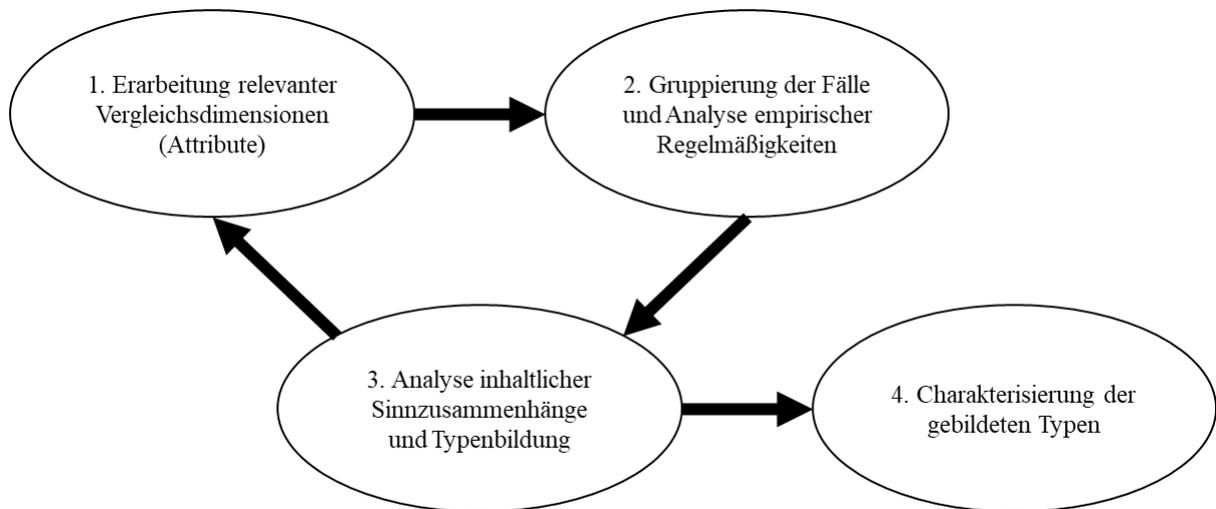


Abbildung 8: Modell empirisch begründeter Typenbildung

Das Modell beschreibt vier Schritte der Typenbildung, wobei die ersten drei Schritte beliebig oft durchlaufen werden können, um die Eignung von Attributen und deren Kombinationen zur Gruppierung der betrachteten Objekte zu prüfen. Im ersten Schritt werden relevante Attribute zur Beschreibung der untersuchten Objekte (hier Modellierprozesse) identifiziert. Im zweiten Schritt werden die Objekte nach diesen Attributen gruppiert und auf interne Homogenität und externe Heterogenität geprüft. Im dritten Schritt werden innerhalb der Gruppen inhaltliche Sinnzusammenhänge über das Attribut selbst hinaus identifiziert. Sind die Gruppen zufriedenstellend beschaffen, werden im vierten Schritt Typen konstruiert, die die identifizierten Attribute und inhaltlichen Sinnzusammenhänge berichten (vgl. „Idealtypen“; Kuckartz, 2010). Die Umsetzung der Typenbildung nach dem Modell empirisch begründeter Typenbildung in der vorliegenden Untersuchung wurde ebenfalls publiziert (Göhner & Krell, 2022b) und ist in Abschnitt 5.3 (S. 79) detailliert beschrieben.

#### 4.4 Forschungskontext

Die vorliegende Untersuchung ist in der ersten Phase der Lehrkräftebildung in Deutschland angesiedelt. In dieser Phase studieren angehende Lehrkräfte im universitär angelegten Bachelor (sechs Semester) und nachfolgenden Master (vier Semester) zwei verschiedene Fächer, bevor sie in die zweite Phase der Lehrkräftebildung übergehen, die in der Schule verortet ist. Es wird erwartet, dass Lehramtsstudierende am Ende ihres Studiums über ausreichend professionelle Kompetenzen für den Lehrberuf verfügen (Neumann et al., 2017). Zu diesen Kompetenzen gehören für Lehramtsstudierende mit mindestens einem naturwissenschaftlichen Fach auch Kompetenzen bezüglich naturwissenschaftlicher Denk- und Arbeitsweisen, wie dem Modellieren (vgl. Abschnitt 2.3.2). Empirische Studien weisen darauf hin, dass angehende Lehramtsstudierende naturwissenschaftlicher Fächer ihre Kompetenzen bezüglich naturwissenschaftlicher Denk- und Arbeitsweisen während ihres Studiums signifikant verbessern (Krüger et al., 2020b), sodass bei der Beprobung von Lehramtsstudierenden naturwissenschaftlicher Fächer mit unterschiedlichem Studienfortschritt ein diverses Spektrum an beobachtbaren Modellierprozessen zu erwarten ist.

## 4.5 Stichprobe

Die für die vorliegende Untersuchung gewählte Stichprobe besteht aus Lehramtsstudierenden der Freien Universität Berlin (im Folgenden FU) und der Humboldt Universität Berlin (im Folgenden HU) mit mindestens einem naturwissenschaftlichen Fach, die zum Zeitpunkt der Untersuchung im Bachelor oder Master studierten.

Um über die Unterschiede bezüglich des Studienfortschritts hinaus eine hohe Diversität von Modellierprozessen beobachten zu können, wird in der vorliegenden Untersuchung ein Sampling nach einem qualitativen Stichprobenplan mit zwei Samplingkriterien umgesetzt (Schreier, 2010): Als erstes Kriterium wurden Kompetenzen der naturwissenschaftlichen Erkenntnisgewinnung ausgewählt (*scientific reasoning competencies*; Krüger et al., 2020b), wobei einige AutorInnen davon ausgehen, dass Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und wissenschaftliches Modellieren in einem positiven Zusammenhang stehen (Giere et al., 2006; Passmore et al., 2014). Zur Erfassung der Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung wurde der sogenannte Ko-WADiS-Test (Hartmann et al., 2015; Krüger et al., 2020a, 2020b; Mathesius et al., 2016). Dieser papierbasierte Test umfasst 123 Multiple-Choice-Aufgaben zu den Konstrukten wissenschaftliches Untersuchen und wissenschaftliches Modellieren und wurde im Rahmen des ValiDiS-Projekts im Rahmen einer Multi-Kohorten-Längsschnittstudie erprobt (Mathesius et al., 2019).

Als zweites Kriterium wurden allgemeine kognitive Fähigkeiten ausgewählt, da diese neben verbalen und numerischen Fähigkeiten auch Fähigkeiten des räumlichen Denkens umfassen, welche als elementare Voraussetzung für Modellierprozesse gelten (Nersessian, 2002). Hierbei konnte bereits empirisch ein positiver Zusammenhang zwischen allgemeinen kognitiven Fähigkeiten und Modellverstehen (Krell et al., 2014b) sowie allgemeinen kognitiven Fähigkeiten und wissenschaftlichem Problemlösen (Nehring et al., 2015) gezeigt werden. Zur Erfassung der allgemeinen kognitiven Fähigkeiten wurden die ersten drei Teilmodule des Grundmoduls des Intelligenz-Struktur-Test 2000R eingesetzt (Liepmann et al., 2007). Die Teilmodule erfassen die Bereiche „verbale Intelligenz“, „numerische Intelligenz“ und „figurale Intelligenz“, die gemeinsam das sogenannte „schlussfolgernde Denken“ im Sinne der sog. „fluiden Intelligenz“, also Intelligenz weitgehend ohne Wissensanteile, abbilden. Aus forschungsökonomischen Gründen wurde auf die fakultativen Teilmodule „Merkfähigkeit“ und „schlussfolgerndes Denken mit Wissensanteilen“ verzichtet. Beide Kriterien wurden in einem mehrstufigen Verfahren mit den beschriebenen, etablierten paper-pencil Instrumenten erfasst (Krüger et al., 2020b; Liepmann et al., 2007). Es wurden diejenigen Studierenden zur Teilnahme an der weiteren Studie eingeladen, die für beide Kriterien signifikant bessere oder schlechtere Werte erreichten als die jeweilige Referenzstichprobe, sodass die Wahrscheinlichkeit weiter erhöht wurde, möglichst diverse Modellierprozesse beobachten zu können (heterogene Stichprobe; Patton, 1990; Schreier, 2010). Als Grenzwert für die Studienteilnahme wurde ursprünglich für beide Kriterien eine Standardabweichung besser oder schlechter

## Methodischer Zugang

als der Mittelwert der Referenzstichprobe gewählt, allerdings musste dieser Grenzwert für das Kriterium „allgemeine kognitive Fähigkeiten“ auf eine halbe Standardabweichung gesenkt werden, um die Anzahl potenziell geeigneter ProbandInnen zu erhöhen (angepeilte Stichprobengröße  $N = 40$ ).

Lehramtsstudierende wurden in Seminaren und Vorlesungen (z. B. Didaktik der Biologie, Zoologie) vom Autor dazu eingeladen an der vorliegenden Untersuchung teilzunehmen. Die Teilnahme an der vorliegenden Untersuchung erfolgte freiwillig und war kein verpflichtender Studienbestandteil. Der Autor der vorliegenden Untersuchung hatte zu den ProbandInnen kein formelles Verhältnis<sup>7</sup>. 666 Lehramtsstudierende haben im Rahmen des Samplingprozesses eines der beiden Instrumente bearbeitet; 202 Lehramtsstudierende beide Instrumente. Basierend auf den oben dargestellten Samplingkriterien konnten 57 Lehramtsstudierende persönlich oder via Email zur Teilnahme an der Studie eingeladen werden, von denen letztendlich 36 als ProbandInnen teilnahmen. Tabelle 4 zeigt die Verteilung der ProbandInnen auf die Subgruppen des qualitativen Stichprobenplans, die sich aus den beiden Samplingkriterien ergeben.

Tabelle 4: Verteilung der ProbandInnen auf die Subgruppen des qualitativen Stichprobenplan

<b>Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung</b> (Ko-WADiS-Fragebogen; Hartmann et al., 2015)				
		<b>Niedrig</b> ( $x < -1$ SD)	<b>Hoch</b> ( $x > +1$ SD)	<b>gesamt</b>
	<b>Niedrig</b> ( $x < -0,5$ SD)	10	6	16
<b>allgemeine kognitive Fähigkeiten</b> (I-S-T 2000R; Liepmann et al., 2007)	<b>Hoch</b> ( $x > +0,5$ SD)	3	17	20
	<b>gesamt</b>	13	23	36

<sup>7</sup> 4 ProbandInnen wurden von Prof. Dr. Moritz Krell und nicht vom Autor zur Teilnahme an der vorliegenden Untersuchung eingeladen. Darunter befindet sich auch der Autor selbst (Pseudonym *Floyd*).

## Methodischer Zugang

Die ProbandInnen waren zum Zeitpunkt ihrer Teilnahme zwischen 17 und 39 Jahre und im Durchschnitt 24 Jahre alt (Tabelle 5). 22 ProbandInnen befanden sich im Bachelorstudium, 14 im Masterstudium. Alle ProbandInnen studierten mindestens ein naturwissenschaftliches Fach, wobei das Fach Biologie am häufigsten vertreten ist ( $n = 35$ ). 8 ProbandInnen studierten zwei naturwissenschaftliche Fächer, hierzu zählten Biologie, Chemie, Physik, Ernährungswissenschaften, Agrarwissenschaften, Informatik.

Tabelle 5: Übersicht über die teilnehmenden Lehramtsstudierenden, alphabetisch geordnet nach Pseudonym (verändert nach Göhner & Krell, 2021)

Pseudonym	Alter	Studiengang	Semester	Universität	Studienfach 1	Studienfach 2	Untersuchungszeit
Achilles	25	Bachelor	3	FU	Biologie	Politik	01:32:02
Alexandra	20	Bachelor	3	FU	Ernährungswissenschaften	Biologie	01:38:26
Alice	22	Bachelor	3	FU	Biologie	Physik	01:38:22
Amy	21	Bachelor	1	FU	Ernährungswissenschaften	Biologie	01:22:13
Angelina	21	Bachelor	3	FU	Biologie	Mathematik	01:01:17
Ben	19	Bachelor	3	FU	Biologie	Politik	01:47:20
Boris	19	Bachelor	4	FU	Biologie	Geschichte	00:11:06
Carlo	33	Master	4	FU	Biologie	Englisch	01:15:29
Celine	17	Bachelor	1	FU	Ernährungswissenschaften	Biologie	00:50:52
Claire	21	Bachelor	4	FU	Biologie	Deutsch	01:32:42
Claudia	30	Bachelor	4	FU	Biologie	Politik	00:33:30
Cynthia	29	Master	4	FU	Sozialkunde	Biologie	01:40:43
Daphne	18	Bachelor	1	FU	Biologie	Spanisch	00:50:39
Donna	44	Bachelor	4	FU	Biologie	Philosophie	01:23:46
Floyd	24	Master	4	FU	Biologie	Informatik	00:33:55
Frida	29	Master	2	FU	Kunst	Biologie	01:49:12
Iris	28	Bachelor	3	FU	Biologie	Mathematik	01:25:16
James	39	Bachelor	3	HU	Agrarwissenschaften	Biologie	00:26:37
Jasmin	25	Master	4	FU	Biologie	Geschichte	01:28:25
Jenny	21	Bachelor	10	FU	Deutsch	Biologie	01:52:03
Jim	24	Master	2	FU	Biologie	Mathematik	01:04:14
Jonathan	25	Bachelor	3	FU	Biologie	Geschichte	00:50:13
Kara	25	Master	4	FU	Biologie	Mathematik	00:08:07
Lana	23	Bachelor	7	HU	Englisch	Biologie	00:48:33
Lauren	19	Bachelor	3	FU	Deutsche Philologie	Biologie	01:16:53
Martin	33	Master	4	FU	Biologie	Chemie	01:52:22
May	27	Master	4	FU	Biologie	Mathematik	01:27:16
Misty	24	Master	4	FU	Biologie	Mathematik	01:25:20
Natalie	21	Bachelor	3	FU	Biologie	Geschichte	01:23:55
Raphael	28	Bachelor	3	HU	Biologie	Chemie	01:05:58
Rocco	24	Master	4	FU	Grundschulpädagogik	NAWI	00:29:28
Ryan	30	Master	2	FU	Englisch	Biologie	01:05:03
Sabrina	18	Bachelor	3	FU	Biologie	Geschichte	00:34:05
Selena	24	Master	4	FU	Biologie	Deutsch	00:43:56
Susi	33	Master	6	FU	Biologie	Politik	01:34:54
Valerie	21	Bachelor	4	FU	Deutsch	Biologie	00:46:21

## 4.6 Datenaufnahme

Die ProbandInnen wurden individuell eingeladen und in einem Einführungsgespräch über das Ziel der Studie informiert und um ihr schriftliches Einverständnis zur Studienteilnahme gebeten. Anschließend wurden zuerst das Modellverstehen und danach Modellierprozess und Modellierprodukt erfasst.

### 4.6.1 Modellverstehen

Das Modellverstehen der ProbandInnen wurde mit einem etablierten Fragebogen bestehend aus einem einleitenden Satz und fünf offenen Fragen erfasst (*constructed response items*; Krell & Krüger, 2016, 2017; Krüger & Krell, 2020), der entsprechend des Modells der Modellkompetenz (Tabelle 1, S. 44) die fünf Aspekte des Modellverstehens adressiert (Tabelle 6).

Tabelle 6: Fragebogen zum Modellverstehen

Aspekt des Modellverstehens	Frage
Einleitung	Die folgenden fünf Fragen beziehen sich auf Modelle in der Biologie im Allgemeinen, d.h. nicht auf Modelle im Unterricht. Bitte beantworten Sie die Fragen daher aus allgemeiner Perspektive. Bitte antworten Sie in ganzen Sätzen. Sollte der Platz nicht ausreichen, können Sie die Rückseite dieses Bogens nutzen.
Eigenschaften von Modellen	Inwieweit entspricht ein Modell Ihrer Meinung nach seinem biologischen Ausgangsobjekt?
Alternative Modelle	Aus welchen Gründen gibt es Ihrer Meinung nach zu einem biologischen Ausgangsobjekt verschiedene Modelle? Bitte nennen Sie wenigstens zwei Gründe.
Zweck von Modellen	Welchen Zweck erfüllen Ihrer Meinung nach Modelle in der Biologie? Bitte nennen Sie wenigstens zwei unterschiedliche Zwecke.
Testen von Modellen	Wie lässt sich Ihrer Meinung nach überprüfen, ob ein biologisches Modell seinen Zweck erfüllt? Bitte nennen Sie wenigstens zwei Vorgehensweisen.
Ändern von Modellen	Nennen Sie Gründe, warum ein gegebenes biologisches Modell Ihrer Meinung nach verändert wird. Bitte nennen Sie wenigstens zwei unterschiedliche Gründe.

Da der Fragebogen sich auf „Modelle in der Biologie im Allgemeinen“ (Tabelle 6) bezieht, wird das hier erfasste Modellverstehen als *dekontextualisiertes* Modellverstehen aufgefasst (Ke & Schwarz, 2020; Krell et al., 2015). Um auch *kontextualisiertes* Modellverstehen untersuchen zu können, wurden zusätzlich auch Verbalisierungen im Rahmen der nachfolgenden Modellieraufgabe identifiziert, die dem Modellverstehen (sowie dem eingesetzten Fragebogen) zugeordnet werden können.

### 4.6.2 Modellierprozesse

Um Modellierprozesse anzuregen, wurde in der vorliegenden Studie ein Blackbox-Setting genutzt. Blackbox-Settings stellen in der naturwissenschaftsdidaktischen Forschung einen etablierten Ansatz zur Untersuchung von Prozessen der Erkenntnisgewinnung und Prozessen des wissenschaftlichen Denkens dar (Krell & Hergert, 2019; Lederman & Abd-El-Khalick, 2002; Passmore & Svoboda, 2012). Die

Blackbox repräsentiert hierbei ein natürliches Phänomen (Upmeier zu Belzen, 2014) und ist durch drei Merkmale charakterisiert: „(a) being believed to be distinct, (b) having observable (and relatable) inputs and outputs, (c) being black (that is, opaque to the observer)” (Glanville, 1982, S. 1). Ein Blackbox-Setting ist so konstruiert, dass es spontan nicht erklärbare Phänomene hervorruft (Output), die bei ProbandInnen Neugier und Motivation zu deren Untersuchung wecken (Lederman & Abd-El-Khalick, 2002; Upmeier zu Belzen, 2014). Durch eine Reduktion fachlicher Komplexität ist für die Untersuchung eines Blackbox-Settings meist kaum (fachliches) Vorwissen notwendig und es kann auf den Problemlöseprozess fokussiert werden. ProbandInnen wird oft die Aufgabe gestellt, durch Manipulation des Inputs und Beobachtung des Outputs auf die inneren Mechanismen einer Blackbox zu schließen. Dies entspricht einer etablierten Vorgehensweise zur Erfassung und Beschreibung von Problemlösestrategien (Vollmeyer & Funke, 1999). Die zielgerichtete Untersuchung eines Blackbox-Settings kann somit als könnenszentriertes, dynamisches Problemlösen verstanden werden (Friege & Lind, 2003; Funke, 2011). Blackbox-Settings wurden in der naturwissenschaftsdidaktischen Forschung bereits zur Beschreibung von experimentellen Problemlösestrategien (Klahr & Dunbar, 1988), zur Förderung des Wissenschaftsverständnisses (Nature of Science; Abd-El-Khalick, 2002; Lederman & Abd-El-Khalick, 2002), zur Analyse von Argumentationsprozessen (Passmore & Svoboda, 2012) sowie zur Förderung des Modellverstehens (Koch et al., 2015) genutzt.

### Das Blackbox-Setting der vorliegenden Untersuchung

In der vorliegenden Untersuchung wurde eine Wasser-Blackbox eingesetzt (Abbildung 9 und Abbildung 10). Diese kann durch das Einfüllen von Wasser und die anschließende Beobachtung einer austretenden Wassermenge (Output) untersucht werden. Die Wasser-Blackbox wurde im Vorfeld der vorliegenden Untersuchung bereits erfolgreich eingesetzt (Krell & Hergert, 2019).



Abbildung 9: Das Blackbox-Setting der vorliegenden Untersuchung. Die Komponenten sind wie folgt beschriftet: a) schriftliche Aufgabe, b) beispielhaft beschriftete Tafel, c) Videokamera (1 von 3), d) Blackbox, e) Eimer mit Wasser, f) verschieden große Bechergläser, g) Messkolben

## Methodischer Zugang

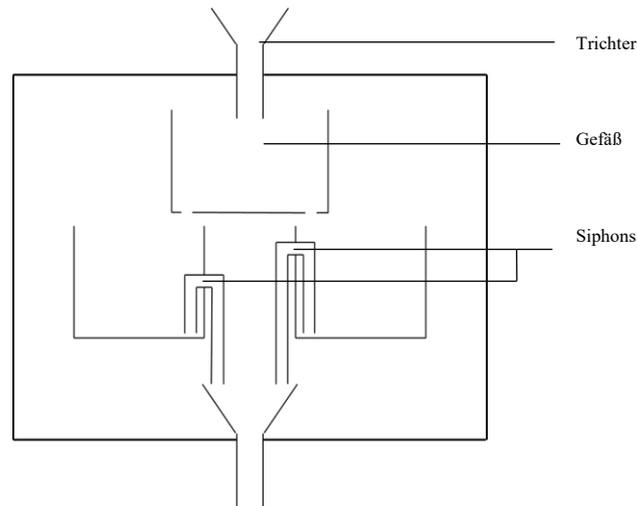


Abbildung 10: Darstellung des inneren Mechanismus der Blackbox. Wasser wird durch einen Trichter in ein erstes Gefäß geleitet, das dieses gleichmäßig auf zwei weitere Gefäße verteilt. Wird ein bestimmter Wasserstand in einem der beiden unteren Gefäße erreicht, entleert sich das jeweilige Gefäß vollständig durch den Siphon. Da die beiden Siphons in unterschiedlichen Höhen angebracht sind, unterscheidet sich der eine Entleerung auslösende Wasserstand bei beiden Gefäßen (links 350 ml, rechts 550 ml).

### Zeichnen

Als Zwischenschritte der Problemlösung und damit des Modellierprozesses können bei der Untersuchung einer Blackbox die von den ProbandInnen angefertigten externen Repräsentationen (z. B. Zeichnungen) der vermuteten Struktur des inneren Mechanismus betrachtet werden (Clement, 2009; Löhner et al., 2003). Hierbei kann angenommen werden, dass die gleichzeitig zum Modellierprozess stattfindende Anfertigung externer Repräsentationen die Analogiebildung aktiviert, die Konsistenzprüfung der angenommenen Systemstruktur fördert, die kognitive Anstrengung der ProbandInnen reduziert und daher eine Tätigkeit darstellt, die die Problemlösung unterstützt (Ainsworth et al., 2011; Blech & Funke, 2013; Heijnes et al., 2018; Leenaars et al., 2013; van Meter & Garner, 2005).

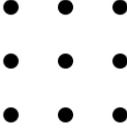
### Lautes Denken

Nach Funke (2011) sollten Problemlöseprozesse analysiert werden, indem neben den Zwischenschritten der Problemlösung parallel auch Verbalisierungen der ProbandInnen in Form von Protokollen des Lauten Denkens ausgewertet werden (Funke, 2011). Die Verbalisierungen werden hierbei als „unmittelbare Repräsentation der kognitiven Prozesse“ betrachtet (Konrad, 2010, S. 479), die zusätzliche Informationen liefern (Ericsson & Simon, 1980; Sandmann, 2014). Die Methode des Lauten Denkens gilt in der naturwissenschaftsdidaktischen Forschung als etabliert (z. B. Clement, 2009; Hung et al., 2019; Löffler & Kauertz, 2016; Ríos et al., 2019; Seah & Magana, 2019). In der vorliegenden Untersuchung wurde

## Methodischer Zugang

das laute Denken vor dem Untersuchungsbeginn geübt. Hierzu wurden die ProbandInnen gebeten drei progressiv komplexer werdende Aufgaben laut denkend zu lösen (Tabelle 7).

Tabelle 7: Übungsaufgaben zum lauten Denken

Aufgabe 1	Aufgabe 2	Aufgabe 3
Zählen Sie, wie viele Fenster in Ihrer Wohnung sind.	Beschreiben Sie den Weg von der Eingangstür dieses Institutsgebäudes bis zur Tür des Raumes, in dem wir uns momentan befinden.	Verbinden Sie die neun Punkte mit vier geraden Linien, ohne dabei den Stift abzusetzen. [Blatt & Stift vorlegen.] 

Während des eigentlichen Modellierprozesses wurden die ProbandInnen gegebenenfalls vom Autor oder einer geschulten studentischen Hilfskraft daran erinnert weiterhin laut zu denken. Alle Verbalisierungen wurden dabei audiografisch mit mehreren Mikrofonen erfasst.

### 4.6.3 Modellierprodukte

Das Ergebnis des Modellierprozesses, also die finale externe Repräsentation des Inneren der Blackbox und damit die letzte während des Modellierprozesses angefertigte Zeichnung, stellt das Modellierprodukt dar (vgl. Couso & Garrido-Espeja, 2017). Einige ProbandInnen präsentieren am Ende ihres Modellierprozesses mehrere Modellierprodukte als mögliche Lösung für die gegebene Aufgabe. Alle Modellierprodukte wurden fotografiert und digital rekonstruiert.

## 4.7 Datenauswertung

Die aufgenommenen Daten wurden unterschiedlich aufbereitet. Die handschriftlich ausgefüllten Fragebögen wurden digitalisiert und tabellarisch gespeichert. In der Software MAXQDA (VERBI Software, 1989 – 2023) wurden anhand der Video- und Audiodaten vollständige, wörtliche Transkripte erarbeitet, die auch ausgewählte Tätigkeiten der ProbandInnen enthalten (z. B. wird vermerkt, wenn ProbandInnen an der Tafel arbeiten). Die Transkripte sind dabei durch Zeitcodes mit den Videodateien verbunden. Alle Modellierprodukte wurden zudem digital rekonstruiert.

### 4.7.1 Modellverstehen

Die Antworten aller ProbandInnen auf die fünf Fragen zum Modellverstehen wurden anhand eines etablierten Kategoriensystems (Krell & Krüger, 2016; Krüger & Krell, 2020) einem Niveau von I bis III zugeordnet (vollständiges Kategoriensystem im Anhang). Antworten, die Aussagen auf mehreren Niveaus enthielten wurde immer mit dem höchsten vorgefundenen Niveau bewertet. Antworten, die keinem Niveau zugeordnet werden konnten, sowie leere Antwortfelder wurden nicht kodiert. Insgesamt wurden so 170 von 180 möglichen Antworten bewertet. Jede Antwort wurde zweimal mit mindestens einwöchigem Abstand durch den Autor dieser Arbeit und einmal durch eine geschulte studentische Hilfskraft kodiert. Cohen's Kappa ( $\kappa$ ) wurde als Maß für die Intra- und Interrater-Übereinstimmung berechnet. Hohe Intrarater- ( $\kappa = 0,68$ ) und Interrater-Übereinstimmungen ( $\kappa = 0,71$ ) deuten auf eine reliable und intersubjektive (vgl. objektive) Auswertung hin.

Zusätzlich wurde auch das Modellverstehen während des Modellierprozesses erfasst (Modellverstehen „in action“; Ke & Schwarz, 2020), indem die Transkripte der Modellierprozesse aller 36 ProbandInnen mit dem gleichen Kategoriensystem ausgewertet wurden. Um hier ebenfalls die Qualität der Analyse sicherzustellen und dabei allerdings auch forschungsökonomisch vorzugehen, wurde 20 % des transkribierten Materials (7 von 36 Transkripten) analog zum Vorgehen bei den Fragebögen zweimal mit mindestens einwöchigem Abstand durch den Autor dieser Arbeit und einmal durch eine geschulte studentische Hilfskraft kodiert. Auch hier wurden hohe Intrarater- ( $\kappa = 0,69$ ) und Interrater-Übereinstimmungen ( $\kappa = 0,72$ ) erreicht.

## Methodischer Zugang

### 4.7.2 Modellierprozesse

Die transkribierten Videodateien wurden ebenfalls in der Software MAXQDA (VERBI Software, 1989 – 2023) qualitativ inhaltsanalytisch untersucht, um einzelne Tätigkeiten des Modellierens zu identifizieren. Das eingesetzte Kategoriensystem verfügt über 19 Kategorien (=Tätigkeiten; Tabelle 8).

Tabelle 8: Kategoriensystem mit 19 Tätigkeiten des Modellierens (Krell et al., 2019) sowie deren Verortung in Phasen des Modellierprozesses (E: Exploration; H: Herstellung; V: Vorhersage; vgl. Abbildung 1)

Phase	Oberkategorie	Kategorie (=Tätigkeit)	Beschreibung
E	Beobachtung eines Phänomens	Beobachtung eines Phänomens	ProbandInnen nehmen Verhalten von Blackbox als spontan nicht erklärbar wahr oder erkennen fehlende Passung zwischen angenommenem Muster der Blackbox und Verhalten von Blackbox
E	Exploration des Systems	Input/Output nicht hypothesengeleitet	ProbandInnen nehmen Eingabe von Input vor und/ oder beobachten den Output (explorativ; nicht hypothesengeleitet)
		Zusammenfassung	ProbandInnen fassen das beobachtete Verhalten der Blackbox zusammen oder beschreiben es
		Input/Output hypothesengeleitet Daten	ProbandInnen nehmen Eingabe von Input vor und/ oder beobachten den Output (Mustererkennung)
E,H,V*	Aktivierung von Analogien und Erfahrungen	Mustererkennung	ProbandInnen bestätigen/ erkennen Muster
		Analogien und Erfahrungen	ProbandInnen formulieren Ideen über Mechanismus der Blackbox und/ oder reflektieren Vor- und Nachteile ihrer Ideen und verwerfen ggf. Ideen (ohne zeichnerische Umsetzung)
H	(Weiter-) Entwicklung eines Modells	Modellentwicklung	ProbandInnen entwickeln auf Grundlage von Beobachtung, Analogien und/ oder Erfahrungen zeichnerisch ein Modell der Blackbox
		Optimierung Modellobjekt	ProbandInnen entwickeln Modell weiter zur Optimierung von Funktionsfähigkeit, Ästhetik o.ä. (Modellobjekt)
		Optimierung Modell retrospektiv	ProbandInnen entwickeln Modell weiter aufgrund fehlerhafter Übereinstimmungen mit Beobachtungen an Blackbox (retrospektiv)
		Verwerfen Modell	ProbandInnen verwerfen Modell aufgrund fehlerhafter Übereinstimmungen mit Beobachtungen an Blackbox (retrospektiv; vgl. Kategorie 12) oder mangelnder Konsistenz des Modellobjekts (vgl. Kategorie 11).
H	Prüfung von Konsistenz und Darstellung	Prüfung Konsistenz Modellobjekt	ProbandInnen reflektieren/ überprüfen/ bewerten Konsistenz des Modells (Modellobjekt)
		Prüfung Konsistenz Modell/Daten	ProbandInnen vergleichen Eigenschaften des Modells mit Beobachtungen an Blackbox (retrospektiv)
H	Feststellung Konsistenz & Darstellung des Modells	Feststellen Konsistenz	ProbandInnen stellen Konsistenz und Angemessenheit des Modells fest
V	Ableiten von Vorhersage (Hypothese) aus Modell	Hypothese aus Modell	ProbandInnen nutzen Modell, um Vorhersage über Output bei bestimmtem Input zu treffen
V	Prüfung der Vorhersage (datenbasiert)	Input/Output hypothesengeleitet Modell	ProbandInnen nehmen Eingabe von Input vor und beobachten den Output (hypothesengeleitet aus Modell)
		Hypothese bestätigt	ProbandInnen bestätigen Hypothese aus Modell durch Output von Blackbox
		Hypothese falsifiziert	ProbandInnen falsifizieren Hypothese aus Modell durch Output von Blackbox
V	Ändern / Verwerfen des Modells (datenbasiert)	Optimierung Modell nach falsifizierter Hypothese	ProbandInnen entwickeln Modell aufgrund von falsifizierter Hypothese weiter (datenbasiert)
		Verwerfen Modell nach falsifizierter Hypothese	ProbandInnen verwerfen Modell aufgrund von falsifizierter Hypothese (datenbasiert)

\* Analogien und Erfahrungen können während des gesamten Modellierprozesses aktiviert werden. Diese Kategorie wird daher keiner spezifischen Phase zugeordnet.

Die durchgeführte qualitative Inhaltsanalyse kann entsprechend der von Stamann et al. (2016) vorgeschlagenen Merkmale wie folgt eingeordnet werden: Die qualitative Inhaltsanalyse der Modellierprozesse kann der „inhaltlich-strukturierenden“ Basisform der qualitativen Inhaltsanalyse zugeordnet

werden und orientiert sich insgesamt an dem von Schreier (2012) vorgeschlagenen Vorgehen. Inhaltliche Kategorien wurden zuerst deduktiv entwickelt und dann in Vorarbeiten induktiv ausdifferenziert (Krell et al., 2019). Während der eigentlichen Untersuchung wurden einzelne Kategorienbeschreibungen darüber hinaus leicht adaptiert, um die Trennschärfe der Kategorien zu erhöhen. Da es sich beim Forschungsgegenstand um eine Sequenz aus Elementen menschlichen Verhaltens handelt (siehe Abschnitt 2.4.2, S. 46), bilden die Kategorien diese einzelnen Elemente ab und können sich nicht überschneiden (Bakeman & Gottman, 1997). Die Kategorien sind somit in einem linear-hierarchischen Kategoriensystem angeordnet (Stamann et al., 2016).

Die Identifikation der einzelnen Tätigkeiten in den Modellierprozessen der ProbandInnen wurde vom Autoren zweimal mit zweiwöchigem Abstand sowie anschließend von einer weiteren geschulten Person (Projektleiter Prof. Dr. Moritz Krell oder studentische Hilfskraft) unabhängig voneinander vorgenommen. Dabei wurden hohe Intrarater- ( $0,67 \leq \kappa \leq 0,84$ ; Mittelwert $_{\kappa} = 0,77$ ; Landis & Koch, 1977) und Interrater-Übereinstimmungen erreicht ( $0,45 \leq \kappa \leq 0,93$ ; Mittelwert $_{\kappa} = 0,77$ ; Landis & Koch, 1977).

Die so identifizierten Sequenzen aus Tätigkeiten wurden anschließend wie folgt ausgewertet: Um die zeitliche Abfolge der Tätigkeiten in den Modellierprozessen der einzelnen ProbandInnen zu visualisieren, wurden alle identifizierten Sequenzen aus Tätigkeiten als *Codelines* dargestellt (vgl. *Time line graph*; Louca & Zacharia, 2015). Diese Codelines wurden in der statistischen Programmiersprache *R* (R Core Team, 2022) in der integrierten Entwicklungsumgebung *RStudio* (RStudio Team, 2022) mit dem Paket *ggplot2* (Wickham, 2016) erarbeitet (für Beispiele siehe Abschnitt 5.2, ab S. 75; alle Codelines im Anhang). Diese ermöglichen zusammen mit den Transkripten eine holistische fallbasierte Analyse der individuellen Vorgehensweisen. Um darüber hinaus den Zusammenhang zwischen verschiedenen Tätigkeiten deutlicher zu visualisieren, wurden zudem mit der Software *Gephi* (Bastian et al., 2009) Netzwerkdiagramme erarbeitet (Lämsä et al., 2021; Newman, 2010), die mit dem Layoutingalgorithmus *ForceAtlas2* (Jacomy et al., 2014) so ausgerichtet wurden, dass häufig nacheinander ausgeführte Tätigkeiten nah beieinander und generell häufig ausgeführte Tätigkeiten zentral ausgerichtet werden (für Beispiele siehe Abschnitt 5.3, ab S.79). Dies ermöglichte es zusätzlich verschiedene Netzwerkattribute zu quantifizieren (insbesondere Komplexität und Homogenität; Morzy et al., 2017; Porter et al., 2009; Ronqui & Travieso, 2015), die im Rahmen der Typologisierung nach dem Modell empirisch begründeter Typenbildung (Kluge, 2000; Abschnitt 4.3, ab S. 59) genutzt wurden. Die Typologisierung ist hierbei detailliert dargestellt in Abschnitt 5.3 (S. 79).

In Folgeanalysen wurden die Modellierprozesse einem ähnlichen Vorgehen folgend mit einem weiteren Kategoriensystem zu Hindernissen beim Modellieren untersucht. Dies wird in Abschnitt 5.4 (ab S. 84) zusammenfassend dargestellt.

### 4.7.3 Modellierprodukte

Um die Qualität der Modellierprodukte zu bewerten, wurde induktiv ein Kategoriensystem erarbeitet. Angelehnt an die Herangehensweise von Bielik et al. (2018) wurden die Modellierprodukte daraufhin untersucht, welche Komponenten und Beziehungen sie enthalten, die nötig sind, um das natürliche Phänomen (hier das Blackbox-Setting) zu erklären. Es wurden drei Konzepte identifiziert, die nötig sind, um das Verhalten der Wasser-Blackbox erklären zu können:

- (1) Das Modellierprodukt sollte einen oder mehrere Behälter enthalten, die mit Wasser gefüllt werden können.
- (2) Mehrere Behälter sollten in einem parallelen System angeordnet sein, sodass sie gleichzeitig mit Wasser gefüllt werden können.
- (3) Es sollte einen Mechanismus geben, der erlaubt, dass sich ein Behälter ab einem gewissen Wasserstand komplett entleert.

Die konkrete Umsetzung dieser drei Konzepte kann dabei variieren. Es sind beispielsweise verschiedene Behälter denkbar (Gläser, Zisternen, Schläuche, Schwämme). Da einige Zeichnungen nicht losgelöst evaluiert werden konnten, wurde das „laute Denken“ der ProbandInnen als zusätzliche Datenquelle herangezogen.

Für die eigentliche Kodierung der finalen Modellierprodukte wurde notiert, welche der drei Konzepte in die Zeichnung integriert wurden. Den Modellierprodukten der ProbandInnen wurde dann ein Wert von 0-3 zugeordnet, je nachdem, ob kein (=0), eins (=1), zwei (=2) oder alle drei (=3) Konzepte in das Modellierprodukt integriert oder verbal beschrieben wurden. Weitere Eigenschaften der Modellierprodukte, z. B. ihre Ästhetik oder die Passung zu den gesammelten Daten, wurden nicht bewertet. Bei ProbandInnen mit mehreren finalen Modellierprodukten wurde das Modellierprodukt mit dem höchsten Wert für alle nachfolgenden statistischen Auswertungen betrachtet. Jedes Modellierprodukt ( $N = 42$ ) wurde mit einem Abstand von zwei Wochen zwei Mal durch den Autor dieser Arbeit bewertet, sowie zusätzlich durch den Kollegen Tom Bielik. Die Berechnung von Cohen's Kappa weist auf sehr hohe Intrarater- ( $\kappa = 0,84$ ) und Interrater-Übereinstimmung hin ( $\kappa = 0,82$ ).

### 4.7.4 Quantitative Auswertung der Zusammenhänge

Die Zusammenhänge zwischen Modellierprozess und allgemeinen kognitiven Fähigkeiten sowie Fähigkeiten naturwissenschaftlicher Erkenntnisgewinnung wurden quantitativ untersucht, indem die Korrelation (Pearson) zwischen den relevanten Variablen ermittelt wurde. Die Zusammenhänge zwischen Modellverstehen, Modellierprozess und Modellierprodukt wurden ebenfalls korrelativ (Spearman's Rangkorrelation) und zusätzlich gruppenvergleichend (Wilcoxon Rangsummentest, ANOVAs, T-Tests) untersucht.

## 5. Ergebnisse und Diskussion der Einzelpublikationen

Die Ergebnisse der vorliegenden Untersuchung wurden bereits in sechs Publikationen berichtet und diskutiert (Tabelle 9), die alle in verschiedenen Publikationsorganen mit Peer-Review-Verfahren veröffentlicht wurden. Alle Publikationen außer Publikation 5 wurden *Open Access* veröffentlicht und sind so dauerhaft kostenlos verfügbar. Alle Publikationen sind angehängt. Weitere Teilergebnisse sind in den Artikeln „Pre-service biology teachers‘ responses to first-hand anomalous data during modeling processes“ (Meister et al., 2021) und „Why most Definitions of Modeling Competence in Science Education fall short: Analyzing the Relevance of Volition for Modeling“ (Ammonet et al., Manuscript submitted for publication) dargestellt, bei denen der Autor der vorliegenden Untersuchung als Mitautor wirkte. Darüber hinaus ist im Rahmen der vorliegenden Untersuchung eine Masterarbeit zu volitionalen Aspekten beim Modellieren entstanden (Nordheimer, 2019).

Tabelle 9: Übersicht über die im Rahmen der vorliegenden Untersuchung entstandenen Publikationen

Nr.	Publikation	Status
1	Göhner, M., & Krell, M. (2020). Qualitative Inhaltsanalyse in naturwissenschafts- didaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review. <i>Zeitschrift für Didaktik der Naturwissenschaften</i> , (26), 207–225. <a href="https://doi.org/10.1007/s40573-020-00111-0">https://doi.org/10.1007/s40573-020-00111-0</a>	Veröffentlicht
2	Göhner, M., & Krell, M. (2018). Modellierungsprozesse von Lehramtsstudieren- den der Biologie. <i>Erkenntnisweg Biologiedidaktik</i> , 17, 45–63.	Veröffentlicht
3	Göhner, M., & Krell, M. (2022). Preservice science teachers‘ strategies in scien- tific reasoning: The case of modeling. <i>Research in Science Education</i> , 52, 395– 414. <a href="https://doi.org/10.1007/s11165-020-09945-7">https://doi.org/10.1007/s11165-020-09945-7</a>	Veröffentlicht
4	Göhner, M., & Krell, M. (2020). Was ist schwierig am Modellieren? Identifika- tion und Beschreibung von Problemen in Modellierungsprozessen von Lehramts- studierenden naturwissenschaftlicher Fächer. <i>Zeitschrift für Didaktik der Natur- wissenschaften</i> , 27, 155-180. <a href="https://doi.org/10.1007/s40573-021-00131-4">https://doi.org/10.1007/s40573-021-00131-4</a>	Veröffentlicht
5	Göhner, M., & Krell, M. (2022). Analyzing the relationships between pre-service biology teachers' modeling processes, scientific reasoning competencies and gen- eral cognitive abilities. In M. Ergazaki & K. Kampourakis (Hrsg.), <i>Contributions from Biology Education Research</i> (S. 231–244). Heidelberg: Springer Nature.	Veröffentlicht
6	Göhner, M., Bielik, T., & Krell, M. (2022). Investigating the dimensions of the modeling competence among pre-service science teachers: Meta-modeling knowledge, modeling practice and modeling product. <i>Journal of Research in Sci- ence Teaching</i> , 59(8), 1354–1387. <a href="https://doi.org/10.1002/tea.21759">https://doi.org/10.1002/tea.21759</a>	Veröffentlicht

Publikation 1 wurde bereits in Abschnitt 4.2 (S. 56) dargestellt. Bevor die weiteren Einzelpublikationen im Hinblick auf ihre wichtigsten Ergebnisse und Interpretationen zusammengefasst und anschließend übergreifend in die vorliegende Untersuchung eingeordnet werden, stellt der folgende Abschnitt die Zusammenarbeit mit den MitautorInnen dar.

## 5.1 Zusammenarbeit mit den MitautorInnen

Tabelle 10 listet alle für diese Untersuchung relevanten Publikationen auf und macht die Zusammenarbeit mit den MitautorInnen, insbesondere dem Projektleiter Prof. Dr. Moritz Krell, für jede Publikation transparent (Brand et al., 2015).

Tabelle 10: Angabe der Rolle der Mitwirkenden Autoren Göhner (G), Krell (K) und Bielik (B) an den präsentierten Publikationen adaptiert nach Brand et al. (2015). Die Kürzel „l“, „e“ und „s“ geben dabei den Grad der Verantwortlichkeit an (lead, equal, supporting).

Publikationsschritt	Publikation 1		Publikation 2		Publikation 3		Publikation 4		Publikation 5		Publikation 6		
	G	K	G	K	G	K	G	K	G	K	G	B	K
<b>Konzeptualisierung</b> Idee; Formulierung oder Weiterentwicklung der übergreifenden Forschungsziele	e	e	s	l	e	e	l	s	s	l	e	-	e
<b>Methodologie</b> Entwicklung und Design der Methodologie; Entwicklung von Modellen	l	s	e	e	e	e	l	s	s	l	e	-	e
<b>Software</b> Programmierung, Softwareentwicklung, Tests von vorliegendem Programmcode	l	-	l	-	-	-	l	-	s	l	l	-	s
<b>Validierung</b> Prüfung von Gütekriterien wie Replizierbarkeit, Reproduzierbarkeit, Reliabilität, Objektivität und Generalisierbarkeit in Bezug auf Experimente und Ergebnisse	l	-	e	e	l	s	l	-	e	e	e	e	-
<b>Formelle Analyse</b> Anwendung statistischer, mathematischer, computerbasierter oder anderer formeller Techniken zur Datenanalyse	s	l	-	-	e	e	l	s	s	l	e	-	e
<b>Untersuchung</b> Durchführung der Untersuchung und Datenaufnahme	l	-	l	s	l	s	l	s	l	s	l	s	-
<b>Ressourcen</b> Beschaffung und Verteilung von Materialien, ProbandInnen, Arbeitsplätzen, Instrumenten, digitaler Infrastruktur	l	-	-	l	l	s	l	s	l	s	l	-	s
<b>Datenmanagement</b> Generierung von Metadaten, Pflege der Untersuchungsdaten für die vorliegende Untersuchung und ihre Archivierung	l	-	l	-	l	-	l	-	l	-	l	-	-
<b>Verfassen des Originalentwurfs</b> Entwurf oder Präsentation zur Publikation der Daten, insbesondere des Erstentwurfs	l	s	l	s	l	s	l	s	s	l	l	s	s
<b>Bearbeitung des Entwurfs</b> Bearbeitung oder Präsentation des Entwurfs, insbesondere Feedback, Kommentare und Revision vor und nach Veröffentlichung	e	e	e	e	e	e	e	e	e	e	e	s	e
<b>Visualisierung</b> Vorbereitung, Ausarbeitung und Präsentation der Untersuchung, insbesondere grafische Präsentation, Visualisierung	l	-	l	s	l	-	l	s	l	s	l	-	-

Als Projektleiter und Antragsteller bei der Deutschen Forschungsgemeinschaft (DFG) übernahm der Autor Moritz Krell über den gesamten Projektzeitraum hinweg die Rollen der Supervision, Projektadministration und Projektfinanzierung, die hier daher nicht differenzierter ausgeführt werden.

## 5.2 Publikation 2: Fallanalyse zweier Modellierprozesse

Göhner, M., & Krell, M. (2018). Modellierungsprozesse von Lehramtsstudierenden der Biologie. *Erkenntnisweg Biologiedidaktik*, 17, 45–63.

Abrufbar unter [https://www.vbio.de/fileadmin/user\\_upload/fachgesellschaften/pdf/FDdB/Publikationen/Erkenntnisweg\\_2018/B3-Erkenntnisweg\\_Goehner\\_Uebearbeitung\\_Einreichung.pdf](https://www.vbio.de/fileadmin/user_upload/fachgesellschaften/pdf/FDdB/Publikationen/Erkenntnisweg_2018/B3-Erkenntnisweg_Goehner_Uebearbeitung_Einreichung.pdf)

### **Zusammenfassung**

Die vorliegende Publikation trägt zum primären Ziel der Untersuchung bei, indem sie einen ersten qualitativen und prozessorientierten Einblick in die Modellierungsprozesse zweier ProbandInnen vornimmt. Im theoretischen Hintergrund wird das Modell des Modellierens eingeführt (Abbildung 6), das auf dem von Krell et al. (2016) beschriebenen „Prozessschema naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren“ beruht und zusätzlich eine Einteilung des Modellierprozesses in die Phasen Exploration, Herstellung und Vorhersage<sup>8</sup> vornimmt. Unter Umsetzung der oben beschriebenen Methodik, insbesondere der Untersuchung und Visualisierung der Modellierprozesse als Codelines, werden die Modellierprozesse der ProbandInnen Angelina und Raphael im Rahmen einer kontrastierenden Fallanalyse gegenübergestellt. In der Publikation werden drei grundlegende Fragestellungen untersucht:

- (1) Welche Tätigkeiten setzen Biologie-Lehramtsstudierende beim Modellieren um?
- (2) Inwiefern lassen sich in den Modellierungsprozessen von Biologie-Lehramtsstudierenden die drei Phasen Exploration, Herstellung und Anwendung zeitlich voneinander abgrenzen?
- (3) Inwiefern besteht ein Zusammenhang zwischen den gezeigten Tätigkeiten beim Modellieren und dem Modellverstehen von Biologie-Lehramtsstudierenden?

Grundsätzlich wurde erwartet, dass sich die im bereits pilotierten Kategoriensystem (Krell et al., 2019) beschriebenen Tätigkeiten beobachten lassen. Zusätzlich wurde erwartet, dass Tätigkeiten, die theoriegeleitet derselben Phase zugeordnet werden, eher zusammen auftreten (Krell et al., 2017) und dass ein ausgeprägtes Modellverstehen das Auftreten von Tätigkeiten der Vorhersage fördert (Cheng & Lin, 2015; Louca & Zacharia, 2012; Schwarz et al., 2009).

---

<sup>8</sup> In der Publikation wird hier noch der Begriff *Anwendung* verwendet, während spätere Publikationen hier den Begriff *Vorhersage* verwenden, um das abduktive Erklären von Beobachtungen am Modell, das auch eine Anwendung des Modells darstellt (vgl. Krüger und Upmeier zu Belzen, 2021), von der Vorhersage abzugrenzen und vorhersagende Tätigkeiten klarer herauszustellen.

### Der Modellierprozess von Angelina

Der Modellierprozess von Angelina ist 63 Minuten lang und besteht aus 62 Events, die sich auf 11 verschiedene Tätigkeiten der Exploration und Herstellung verteilen (vgl. Forschungsfrage 1; Abbildung 11). Die von Angelina im Rahmen ihres Modellierprozesses ausgeführten Tätigkeiten können hierbei in exakt zwei Phasen eingeteilt werden, eine lange Explorations- und eine kürzere Herstellungsphase (vgl. Forschungsfrage 2). Im beobachteten Modellierprozess gelingt es Angelina nicht, das von ihr entwickelte Modellierprodukt losgelöst von ihren Beobachtungen zu betrachten und mit dem Modellierprodukt überprüfbare Vorhersagen zu entwickeln (Louca & Zacharia, 2015). Ein deduktives Testen des Modellierprodukts zur Falsifikation findet in ihrem Modellierprozess nicht statt; allerdings evaluiert sie ihr Modellierprodukt retrospektiv anhand ihrer Beobachtungen.

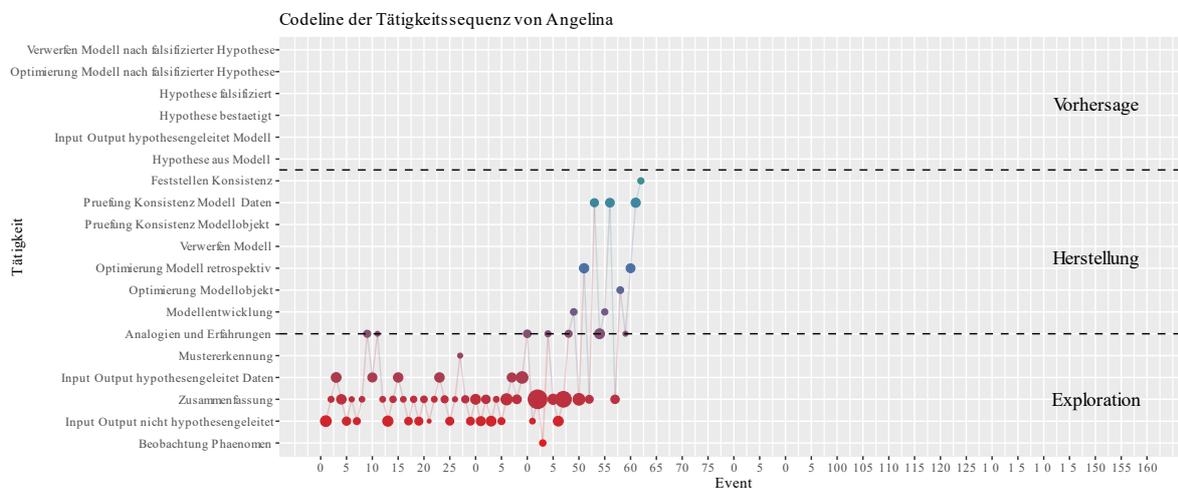


Abbildung 11: Codeline der Tätigkeitssequenz von Angelina (angepasst nach Göhner & Krell, 2018). Zeitlich länger andauernde Events sind als größere Punkte dargestellt.

### Der Modellierprozess von Raphael

Der Modellierprozess von Raphael ist 66 Minuten lang und besteht aus 77 Events, die 17 verschiedene Tätigkeiten des Modellierens umfassen. Hierbei umfasst Raphaels Modellierprozess auch wiederholt Tätigkeiten der Vorhersage (vgl. Forschungsfrage 1; Abbildung 12). Dies ermöglicht ihm die schrittweise Verbesserung seines Modellierprodukt, bis er es als ausreichend „erklärungsstark“ wahrnimmt. Raphaels Modellierprozess kann dabei in 17 deutlich kürzere Phasen unterteilt werden (vgl. Forschungsfrage 2). Dieser häufige Wechsel zwischen Phasen der Exploration, Herstellung und Vorhersage entspricht eher dem in der Literatur beschriebenen zyklischen Charakters des Modellierens (Giere et al., 2006; Gilbert & Justi, 2016; Khan, 2011; Passmore et al., 2014). Dabei kann der Modellierprozess von Raphael zwar als eher komplex und dynamisch beschrieben werden (Gilbert & Justi, 2016), ein streng linear-zyklisches Vorgehen, zum Beispiel nach dem in der Literatur oftmals beschriebenen *Generation-Evaluation-Modification* Kreislauf (Khan, 2011) ist jedoch nicht erkennbar (Louca & Zacharia, 2015).

## Ergebnisse und Diskussion der Einzelpublikationen

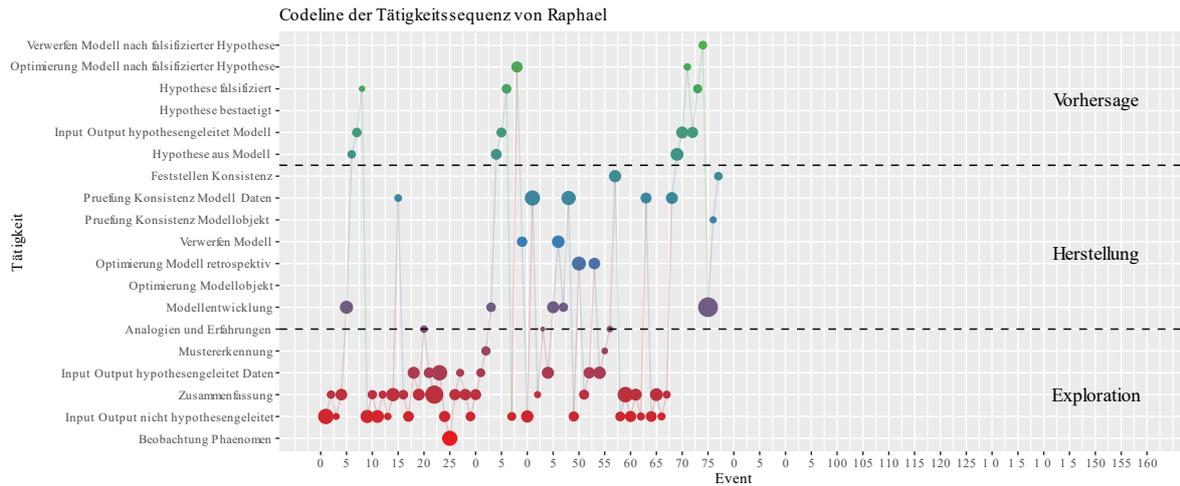


Abbildung 12: Codeline der Tätigkeitssequenz von Raphael (angepasst nach Göhner & Krell, 2018). Zeitlich länger andauernde Events sind als größere Punkte dargestellt.

### Die Modellierprozesse von Angelina und Raphael im Vergleich

Tabelle 11 zeigt eine detaillierte Übersicht über die Auftretenshäufigkeit und verbrachte Zeit mit den jeweiligen Tätigkeiten des Modellierens. Insgesamt überwiegen sowohl bei Angelina als auch bei Raphael Tätigkeiten der Exploration und Herstellung in Häufigkeit und Länge. Dies deckt sich mit Vorarbeiten (Krell et al., 2017). Analog zu den Ergebnissen von Sins et al. (2005) tritt insbesondere die Tätigkeit „Zusammenfassung“ sehr häufig und über längere Zeiträume hinweg auf, unabhängig von der vorhergehenden Tätigkeit.

Tabelle 11: Vergleich des Auftretens und des zeitlichen Anteils der in den Modellierungsprozessen von Angelina und Raphael beobachteten Tätigkeiten (Göhner & Krell, 2018)

Phase	Tätigkeit	Angelina		Raphael	
		Zeit absolut [s]	Anzahl	Zeit absolut [s]	Anzahl
Exploration	1	19,3	1	122,3	3
	2	731,1	14	682,5	16
	3	1807,2	24	946	17
	4	510,2	6	441	8
	5	8,4	1	31	2
Herstellung	6	211,1	7	20,3	3
	7	39,9	2	441,9	5
	8	23,8	1	0	0
	9	134,9	2	148,3	2
	10	0	0	105,8	2
	11	0	0	8,3	1
	12	173,9	3	338,3	5
	13	17,2	1	81,8	2
Anwendung	14	0	0	133,3	3
	15	0	0	162,7	4
	17	0	0	56,8	3
	18	0	0	58,1	2
	19	0	0	18,6	1
<b>Summe</b>		<b>3677</b>	<b>62</b>	<b>3797</b>	<b>77</b>

**Das Modellverstehen von Angelina und Raphael**

In Bezug auf Forschungsfrage 3 können Angelinas und Raphaels Antworten im Rahmen des offenen Fragebogens zum Modellverstehen (Tabelle 12) konsistent einer medialen Perspektive auf Modelle zugeordnet werden (Niveau II; Upmeier zu Belzen & Krüger, 2010). Eine Perspektive auf Modelle als Werkzeuge der Erkenntnisgewinnung geht weder aus Angelinas noch aus Raphaels Antworten hervor.

Tabelle 12: Aussagen zum Modellverstehen von Angelina und Raphael<sup>9</sup>

Aspekt	Angelina	Raphael
<b>Eigenschaften von Modellen</b>	Modelle dienen eher einer Vereinfachung des tatsächlichen Ausgangsobjekts, sodass Aspekte ggf. auch unberücksichtigt bleiben können, wenn diese nicht dem unmittelbaren Zweck des Modells dienen. Essenziell wichtige Teile sind davon nicht betroffen.	Das ist abhängig vom Modell. Ein Modell kann nur veranschaulichen, wie das biologische Ausgangsobjekt funktioniert bzw. arbeitet, dann kann das Modell in seinem Aussehen sehr weit vom Original abweichen. Die meisten anatomischen Modelle ähneln dem Original sehr stark, können aber die Funktion kaum oder nur schwer anschaulich machen.
<b>Alternative Modelle</b>	Unter Umständen werden in verschiedenen Arbeitsgruppen bei der Erstellung eines Modells unterschiedliche Schwerpunkte gesetzt, sodass Teile ausgeklammert bzw. eher fokussiert werden. Zudem können Modelle veraltet sein und neuere Modelle schließen neue Erkenntnisse gegenüber dem alten Modell ein.	Um unterschiedliche Aspekte darzustellen z. B. Funktion, Aufbau, molekulare Prozesse. Je nach Themenschwerpunkt, der betrachtet werden soll, eignen sich unterschiedliche Modellarten besser oder schlechter.
<b>Zweck von Modellen</b>	Der grundlegende Zweck liegt meiner Meinung nach in der Veranschaulichung komplexer Strukturen/Sachverhalte/Entwicklungen etc. So wird die Vorstellungskraft gestützt und das Merken wird leichter.	Modelle ermöglichen eine Veranschaulichung von Objekten, die am Original nicht so gut zu erkennen sind oder im Unterricht nicht im Original vorgeführt werden können. Schwer vorstellbare Prozesse oder Fakten können anschaulich und leichter zu verstehen dargestellt werden. Sie ermöglichen eine übersichtliche Darstellung von experimentellen Ergebnissen (Diagramm o.ä.)
<b>Testen von Modellen</b>	Indem grundlegend wichtige Aspekte, die das Modell veranschaulichen soll, auf ihr Vorhandensein geprüft werden. Andererseits indem stets neue wissenschaftliche Erkenntnisse immer noch auf das Modell anwendbar sind.	Wird ein Modell aufgestellt, um etwas darzustellen, muss es zu den wissenschaftlichen Ergebnissen passen. Alle gezeigten Strukturen müssen mit den ermittelten Daten erklärbar sein. Ein Modell, das dazu dient etwas zu veranschaulichen bzw. etwas komplexes leicht verständlich zu übermitteln, müsste durch die Zielgruppe getestet werden, wenn das Modell bei dieser den gewünschten Effekte hat, ist es geeignet.
<b>Ändern von Modellen</b>	Ein Modell könnte aufgrund einer neuen Erkenntnislage angepasst werden oder durch die Erkennung eines Fehlers im Modell auch gänzlich verworfen werden.	Neue wissenschaftliche Fakten, die das Modell widerlegen oder Teile des Modells müssen zu einer Veränderung des Modells führen. Wenn durch neuere Techniken o.ä. eine bessere Form der Darstellung möglich ist, sollte ein Modell verändert werden.

<sup>9</sup> Aus Platzgründen enthält die Publikation nur Auszüge, wohingegen hier die vollständigen Antworten von Angelina und Raphael auf den Fragebogen zum Modellverstehen dargestellt sind

### 5.3 Publikation 3: Entwicklung einer Typologie von Modellierungsstrategien

Göhner, M., & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*. 52, 395–414. <https://doi.org/10.1007/s11165-020-09945-7>. Lizenziert nach [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

#### **Zusammenfassung**

In dieser Publikation wird berichtet, wie die für die vorliegende Untersuchung zentrale Typologie von Modellierungsstrategien Lehramtsstudierender erarbeitet wurde. Methodisch orientiert sich die Publikation dabei, wie in Abschnitt 4.2 (S. 56) beschrieben, am Modell der empirisch begründeten Typenbildung (Abbildung 8; Kluge, 2000), aus dem sich die vier in dieser Publikation verfolgten Forschungsfragen direkt ableiten. Primäres Ziel ist die Identifikation der Typen (Schritt 4; Kluge, 2000), die zur folgenden Forschungsfrage führt: Welche Typen können im Rahmen einer empirisch begründeten Typenbildung identifiziert werden? Dabei richtet sich der Prozess der Typenbildung entsprechend dem Modell der empirisch begründeten Typenbildung an den vorhergehenden Schritten 1-3 aus, aus denen sich ebenfalls drei Forschungsfragen ableiten lassen:

- (1) Welche Attribute eignen sich, um empirisch begründet eine Typologie der Modellierungsstrategien Lehramtsstudierender zu entwickeln (vgl. Schritt 1; Kluge, 2000)?
- (2) Inwieweit können die Modellierprozesse – unter Maximierung der internen Homogenität und externen Heterogenität – anhand der ausgewählten Attribute gruppiert werden (vgl. Schritt 2; Kluge, 2000)?
- (3) Inwieweit bestehen zwischen den Modellierprozessen, die der gleichen Gruppe zugeordnet werden, inhaltliche Sinnzusammenhänge (vgl. Schritt 3; Kluge, 2000)?

#### **Die Entwicklung einer Typologie von Modellierungsstrategien**

Basierend auf den zwei Attributen „Auftreten von Tätigkeiten des Modellierens“ und „Übergänge zwischen den Tätigkeiten des Modellierens“ wurde die in Tabelle 13 dargestellte Typologie entwickelt (vgl. Fragestellung 1). Die Typologie beschreibt sechs Typen von Modellierungsstrategien, die drei übergeordneten Haupttypen zugeordnet werden können.

## Ergebnisse und Diskussion der Einzelpublikationen

Tabelle 13: Typen von Modellierungsstrategien, deren Komplexität und Homogenität sowie die Anzahl von ProbandInnen deren Modellierprozesse dem jeweiligen Typ zugeordnet wurden (Göhner & Krell, 2021, 2022b)<sup>10</sup>

Typ von Modellierungsstrategie	Kurzbeschreibung	Häufigkeit	Komplexität	Homogenität
1 Nur Exploration	Es werden ausschließlich Tätigkeiten der Exploration (1-5) durchgeführt. Eine Vermutung über den inneren Mechanismus der Blackbox wird lediglich verbal kommuniziert, es entsteht kein gezeichnetes Modell.	1	Niedrig	Niedrig
2a Exploration & Herstellung; Fokus Exploration	Es werden Tätigkeiten der Exploration (1-5) und der Herstellung (7-13) durchgeführt. Dabei überwiegen* Tätigkeiten der Exploration. Das gezeichnete Modell entsteht oft erst am Ende der Modellierprozesse und dient primär der abschließenden Repräsentation der Vermutungen.	12	Niedrig	Mittel
2b Exploration & Herstellung; Fokus Herstellung	Es werden Tätigkeiten der Exploration (1-5) und der Herstellung (7-13) durchgeführt. Dabei überwiegen* Tätigkeiten der Herstellung. Ein gezeichnetes Modell entsteht oft früh in Modellierprozessen und dient primär der fortlaufenden Entwicklung der Vermutungen.	2	Mittel	Mittel
2c Exploration & Herstellung	Es werden Tätigkeiten der Exploration (1-5) und der Herstellung (7-13) durchgeführt, ohne dass ein Schwerpunkt auf einer der beiden Phasen erkennbar ist. Ein gezeichnetes Modell entsteht oft früh in Modellierprozessen und dient primär der retrospektiven Kontrastierung der entwickelten Vermutung und der beobachteten Daten.	6	Mittel	Hoch
3a Exploration, Herstellung & Vorhersage	Es werden Tätigkeiten der Exploration (1-5), der Herstellung (7-13) und der Vorhersage (14-19) durchgeführt. Das gezeichnete Modell wird dabei einmalig zur Vorhersage von nachfolgend generierten Daten eingesetzt.	9	Hoch	Mittel
3b Exploration, Herstellung & mehrmalige Vorhersage	Es werden Tätigkeiten der Exploration (1-5), der Herstellung (7-13) und der Vorhersage (14-19) durchgeführt. Das gezeichnete Modell wird dabei mehrmals zur Vorhersage von nachfolgend generierten Daten eingesetzt.	6	Hoch	Hoch

\* Das überwiegende Vorkommen von Tätigkeiten setzt voraus, dass mindestens 50 % der Tätigkeiten in einem Modellierungsprozess der entsprechenden Phasen zugeordnet werden können.

Für die Typenbildung war ein zweimaliges Durchlaufen des Modells empirisch begründeter Typenbildung nötig<sup>11</sup>. Die im ersten Durchlauf vorgenommene Gruppierung der Modellierprozesse der ProbandInnen, die als Codelines visualisiert wurden (vgl. Abbildung 11, S. 76), basierte auf dem Auftreten der Tätigkeiten des Modellierens (wie z. B. in Klahr & Dunbar, 1988; Schauble, Glaser et al., 1991) und führte zu drei Gruppen, die den drei in Tabelle 13 dargestellten Haupttypen entsprechen: 1 (Nur Exploration), 2 (Exploration und Herstellung) und 3 (Exploration, Herstellung und Vorhersage). Während diese anhand des Auftretens von Tätigkeiten des Modellierens gebildeten Gruppen zueinander sehr gut abgrenzbar waren (hohe externe Heterogenität), wiesen einige Modellierprozesse, die der gleichen Gruppe zugeordnet wurden, zum Teil deutliche Unterschiede auf (niedrige interne Homogenität). Daher

<sup>10</sup> Während die Ergebnisse bereits von Göhner und Krell (2022b) beschrieben wurden, findet sich die zusammenfassende Tabelle in Göhner und Krell (2021)

<sup>11</sup> In die Typenbildung wurden die zu diesem Zeitpunkt vorliegenden Modellierprozesse von 32 ProbandInnen miteinbezogen. Die später beobachteten Modellierprozesse der verbleibenden 4 ProbandInnen wurden nachträglich in die Typologie eingeordnet.

## Ergebnisse und Diskussion der Einzelpublikationen

wurde eine erneute Gruppierung vorgenommen, die zusätzlich die Übergänge zwischen den Tätigkeiten des Modellierens als relevantes Attribut beachtet (vgl. Sins et al., 2005). Hierzu wurden die Modellierprozesse in der Softwareumgebung Gephi (Bastian et al., 2009) als Netzwerkdiagramme bzw. genauer als Zustandsübergangsdiagramme (*state transition graphs*; Andrienko & Andrienko, 2018) visualisiert und automatisch gelayoutet (*ForceAtlas2*; Jacomy et al., 2014). Abbildung 13 zeigt das kombinierte Zustandsübergangsdiagramm aller 36 ProbandInnen.

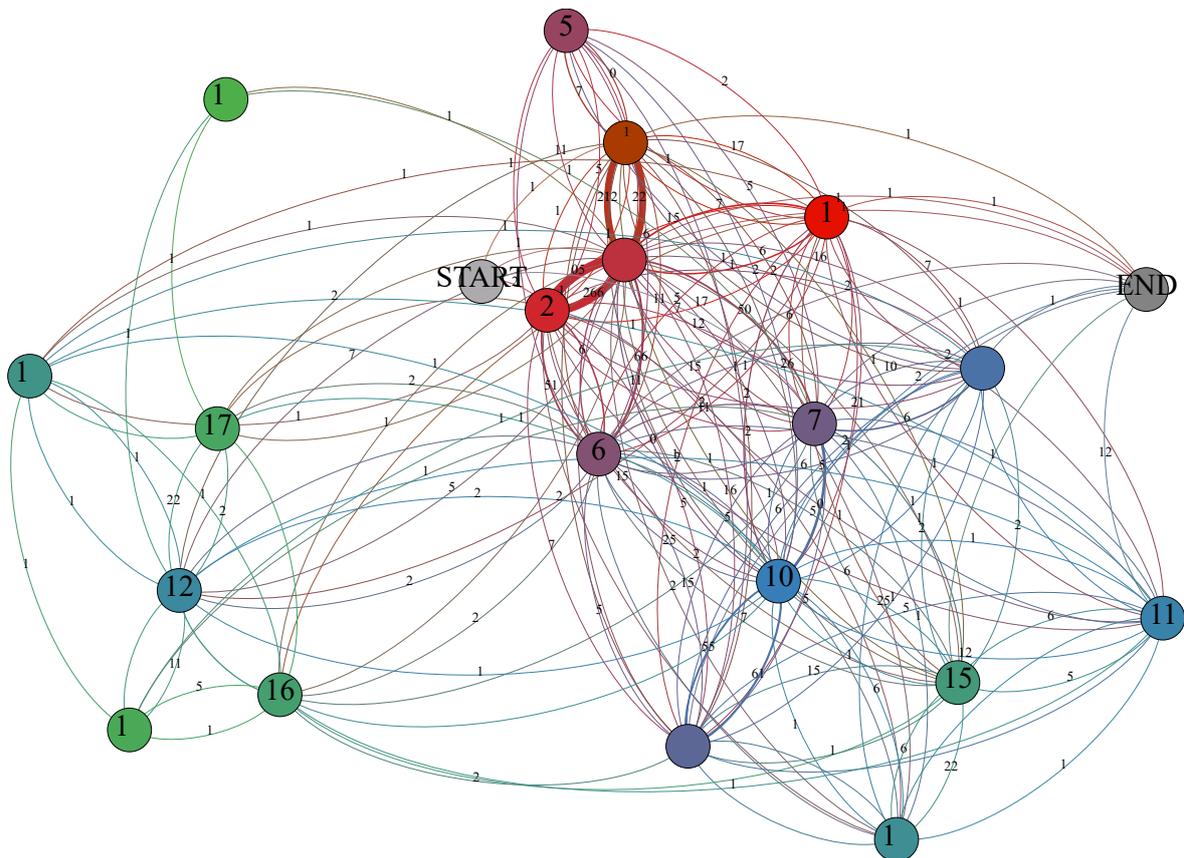


Abbildung 13: Zustandsübergangsdiagramm der kombinierten Modellierprozesse aller 36 ProbandInnen. Die 19 Tätigkeiten des Modellierens, sowie Start und Ende der Modellierprozesse, sind als Knoten dargestellt. Die Übergänge zwischen den Tätigkeiten des Modellierens sind als gerichtete Kanten dargestellt, die mit ihrer Auftretenshäufigkeit beschriftet sind. Die Kantenrichtung ist durch die Kantenkrümmung angegeben und verläuft immer im Uhrzeigersinn.

Durch die Visualisierung der Modellierprozesse aller ProbandInnen einzeln als Zustandsübergangsdiagramm sind Unterschiede bereits visuell leicht erkennbar (Abbildung 14). So können besonders häufig auftretende Übergänge identifiziert werden und Tätigkeiten des Modellierens, die oft sequenziell ausgeführt werden, gruppieren sich automatisch näher zueinander, während sich seltene, weniger häufig verknüpfte Tätigkeiten des Modellierens abkapseln. Für die visuelle Gruppierung der Zustandsübergangsdiagramme wurden dabei besonders die Komplexität, also die Größe des verbundenen Zustandsübergangsdiagramms, sowie die Homogenität, also möglichst gleichmäßige Abstände der Knoten zueinander und Auftretenshäufigkeit ihrer Übergänge, genutzt. Dieses Vorgehen erlaubte die weitere

## Ergebnisse und Diskussion der Einzelpublikationen

Ausdifferenzierung der drei Haupttypen in sechs Typen von Modellierungsstrategien und erhöhte somit die interne Homogenität.

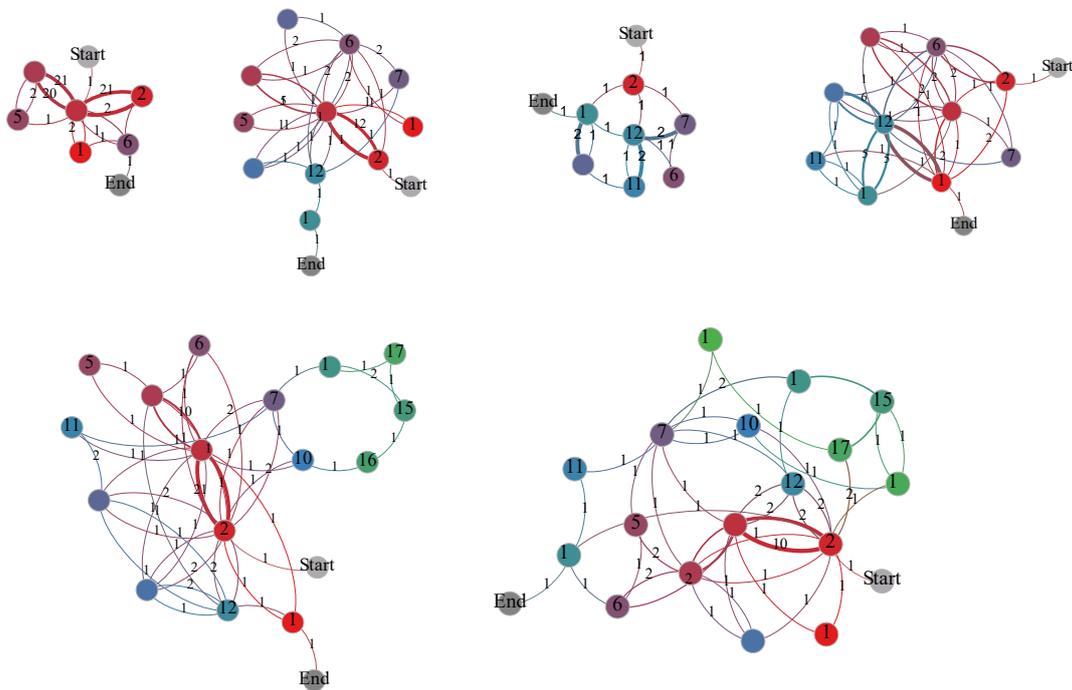


Abbildung 14: Zustandsübergangdiagramme der Modellierprozesse von einzelnen Probandinnen, die den sechs Typen von Modellierungsstrategien zugeordnet wurden. Von links oben nach rechts unten: Carlo (Typ 1), Angelina (Typ 2a), Kara (Typ 2b), Jonathan (Typ 2c), Ben (Typ 3a), Raphael (Typ 3b).

### Modellierprozesse quantifizieren

Weitergehend wurden die Zustandsübergangdiagramme anhand verschiedener Metriken quantifiziert. Die Metriken *communities* (Porter et al., 2009) und *centrality* (Ronqui & Travieso, 2015) konnten leicht angepasst (für methodische Details siehe Göhner & Krell, 2022b) als passende Metriken für die initial visuell geschätzte Komplexität und Homogenität der gebildeten Typen identifiziert werden und erlauben die weitere Evaluation der Typen (Abbildung 15).

## Ergebnisse und Diskussion der Einzelpublikationen

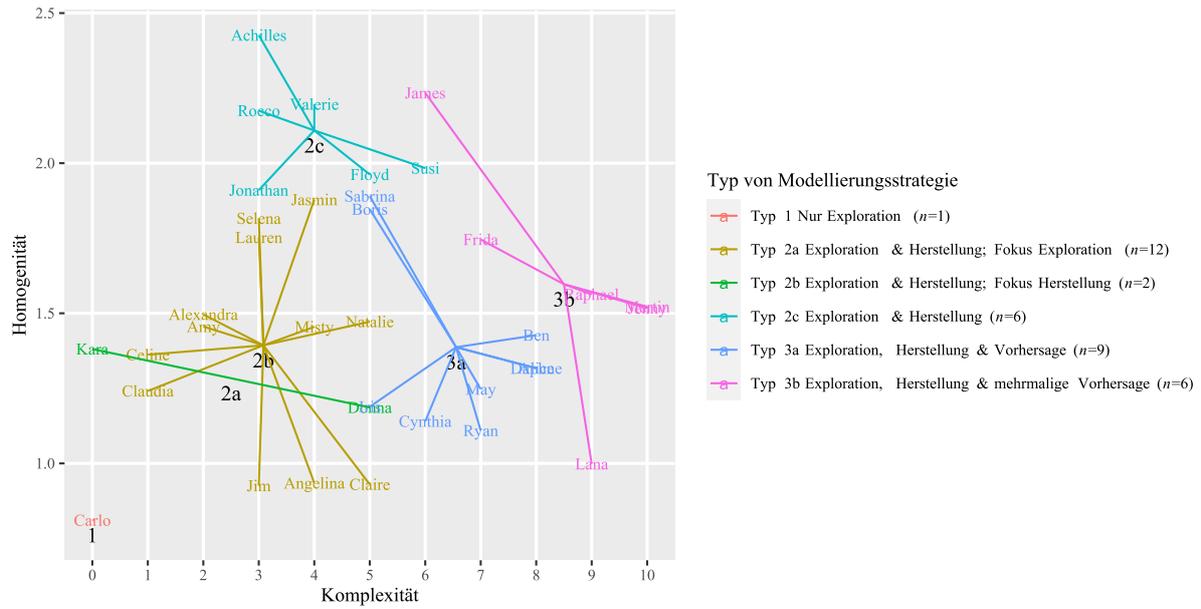


Abbildung 15: Sterndiagramm der quantifizierten Variablen Komplexität und Homogenität der Modellierprozesse. ProbandInnen, die dem gleichen Typ zugeordnet wurden, sind in derselben Farbe hervorgehoben und mit dem Durchschnittswert (Centroid) ihres Typen verbunden.

## 5.4 Publikation 4: Identifikation und Beschreibung von Hindernissen beim Modellieren

Göhner, M., & Krell, M. (2021). Was ist schwierig am Modellieren? Identifikation und Beschreibung von Hindernissen in Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer. *Zeitschrift für Didaktik der Naturwissenschaften*, 27, 155-180. <https://doi.org/10.1007/s40573-021-00131-4>. Lizenziert nach [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

### **Zusammenfassung**

In dieser Publikation werden Hindernisse in den Modellierprozessen der ProbandInnen identifiziert und beschrieben. Hiermit wurde das Ziel verfolgt, mögliche Erklärungen dafür zu finden, dass in der vorliegenden Untersuchung nur selten Modellierprozesse beobachtet wurden, die Tätigkeiten der Vorhersage umfassen und Modelle so im Sinne epistemischer Werkzeuge nutzen. Im Rahmen der theoretischen Konzeptualisierung der Modellierprozesse als Problemlöseprozesse (vgl. Abschnitt 2.1.2, S. 23; Funke, 2011; Mayer, 2007) wird ein Hindernis verstanden als eine „Stelle im Bearbeitungsprozess, in der rekonstruierbar ist, dass eine Person nichts oder etwas nicht selbstverständlich (im Sinne von nicht sicher, zweifelnd) ausführt und dabei auf nichts in der Aufgabensituation Anwendbares zugreifen möchte bzw. zugreifen kann“ (Lange, 2013, S. 32). Problemlösen wird aufgefasst als „das Beseitigen eines Hindernisses oder das Schließen einer Lücke in einem Handlungsplan durch bewusste kognitive Aktivitäten, die das Erreichen eines beabsichtigten Ziels möglich machen sollen“ (Funke, 2011, S. 138). Es werden drei Forschungsfragen verfolgt:

- (1) Welche Hindernisse lassen sich in den Modellierprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer identifizieren und beschreiben?
- (2) Inwiefern treten die identifizierten Hindernisse in einzelnen Phasen der Modellierprozesse von Lehramtsstudierenden naturwissenschaftlicher Fächer besonders häufig auf?
- (3) Inwiefern besteht ein Zusammenhang zwischen den identifizierten Hindernissen und den Modellierprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer?

Um diese Fragen beantworten zu können, wurde erneut eine strukturierende qualitative Inhaltsanalyse umgesetzt. Die Kategorien wurden initial induktiv gebildet, indem Stellen in den Modellierungsprozessen identifiziert wurden, an denen die ProbandInnen innehalten und verbalisieren, dass sie die Aufgabe als schwierig empfinden oder nicht mehr weiter wissen, sowie längere Pausen beim lauten Denken machen (vgl. „subjektive Barriere“; Dörner, 1976). Darüber hinaus wurden auch Stellen kodiert, die Verbalisierungen enthalten, welche der ausgeführten Handlung nicht entsprechen (vgl. „intersubjektive Barriere“; Dörner, 1976). Die induktiv gebildeten Kategorien wurden deduktiv erweitert, indem vergleichbare Studien zum Modellieren und anderen naturwissenschaftlichen Arbeitsweisen herangezogen

## Ergebnisse und Diskussion der Einzelpublikationen

wurden (Baur, 2018; Hammann, 2004; Hogan & Thomas, 2001; Meister et al., 2021; Sins et al., 2005; Windschitl & Thompson, 2006; Woolley et al., 2018). In Bezug auf Fragestellung 2 und 3 wurden die identifizierten Hindernisse zu den bereits identifizierten Tätigkeiten des Modellierens in Relation gesetzt.

### Hindernisse beim Modellieren

Als zentrales Ergebnis wurde das in Tabelle 14 dargestellte Kategoriensystem erarbeitet, das 13 verschiedene Hinderniskategorien identifiziert und beschreibt (vgl. Fragestellung 1).

Tabelle 14: Kategoriensystem der identifizierten Hindernisse in den Modellierungsprozessen von Biologie-Lehramtsstudierenden (N = 36) an einer Blackbox, absteigend sortiert nach Anzahl der ProbandInnen mit den Hindernissen (Göhner & Krell, 2021)

Nr.	Hindernis	Beschreibung	Beispiel	Anzahl ProbandInnen mit diesem Hindernis	Anzahl Kodierungen
1	Fehlende relevante Analogie oder Erfahrung	ProbandInnen äußern, dass sie keinerlei Ideen haben oder nennen gewünschte Wirkung, für die sie jedoch keine passende Analogie oder Erfahrung nennen können. ProbandInnen fällt es schwer sich von bereits als unzureichend erklärter Idee zu lösen.	„Ich habe schon eine Idee, ich komm nur nicht auf die, wie ich das, wie genau ich das mache, damit das so funktioniert, wie ich mir das vorstelle.“ (Achilles, 50)	32	132
2	Fehlerhafte Dokumentation	ProbandInnen äußern, dass Sie bei der Dokumentation einen Fehler gemacht oder etwas vergessen haben oder sind unsicher über die Qualität ihrer Dokumentation. ProbandInnen nehmen eine Beobachtung vor, aber dokumentieren diese gar nicht oder nicht richtig.	„Dann habe ich einen Fehler in meinen Notizen gemacht.“ (Cynthia, 1 )	30	89
3	Hinweise auf unsystematischen Input	ProbandInnen äußern (auch retrospektiv) Probleme beim Abmessen/Einfüllen der gewünschten Wassermenge oder begründen die Wassermenge willkürlich, z. B., um bestimmte runde Zahlen zu erreichen.	„D. h. ich habe jetzt eine ungenaue Menge an Wasser rein gekippt“ (Floyd, )	27	61
4	Unzufriedenheit mit den Daten	ProbandInnen gehen von einem anderen Muster aus, welches z. B. durch Messfehler und verschiedene Daten entstanden ist und verfolgen dieses weiter, ProbandInnen erkennen kein Muster, da ihre Datenreihen keine Wiederholungen der Outputs aufweisen.	„Es ist so komisch, dass sich die Datensätze verändern. Das ist doch irgendwie/ Ich kann einfach kein Muster erkennen.“ (Cynthia, 6 )	21	59
5	Erwartungen an die Aufgabe	ProbandInnen äußern die Vermutung die Lösung der Aufgabe sei leichter oder komplexer als sie zu dem Zeitpunkt annehmen und/oder nennen andere Erwartungen, die sie nicht zufriedenstellend erfüllen können (z. B. mathematische Fähigkeiten).	„Ich hoffe das ist eigentlich kein Schülerversuch für die 5 Klasse oder so und ich beiße mir daran gerade die Zähne aus.“ (Angelina, 5)	17	32
6	Limitationen bei der Exploration	ProbandInnen kritisieren Beschränkungen im erwünschten Umgang	„Darf man unten den Schlauch zuhalten und oben Wasser rein gießen? T: Nein,	15	18

## Ergebnisse und Diskussion der Einzelpublikationen

		mit der Blackbox (Öffnen, Bewegen, etc.)	bitte nicht. P: Schade. [stöhnt]" (Ben, 10 )		
7	Konkrete Analogie vorhanden, deren Funktionsweise unbekannt ist	ProbandInnen nennen konkrete Analogie, können diese in ihrer Funktionsweise jedoch nicht erklären oder zeichnen.	„So ein Überlaufschutz, der es dann komplett entleert. Wie ging sowas? Wie war denn das?“ (Jonathan, 1)	13	26
8	Zustand des Systems unklar	Der Zustand des Systems ist den ProbandInnen unklar, es werden oft Mutmaßungen über bereits vor Beginn der Untersuchung enthaltene Wassermengen aufgestellt.	„Das heißt, das muss irgendwie auf jeden Fall schon vorher Wasser drin gewesen sein.“ (Achilles, 2 )	12	15
9	Mangelnde Zeichenfähigkeiten	ProbandInnen bekunden Unzufriedenheit mit ihrer Zeichnung oder bemängeln ihre Zeichenfähigkeit im Vorhinein einer Zeichnung und führen diese infolgedessen gar nicht erst durch.	„Oh ich kann sowas nicht zeichnen.“ (Frida, 1 )	9	16
10	Randbeobachtung wird überbewertet	ProbandInnen interpretieren Beobachtungen wie die Fließgeschwindigkeit oder Geräusche im Inneren der BB als höchst relevant für das Lösen der Aufgabe und verwenden viel Zeit auf deren Untersuchung.	„Vielleicht hat das etwas mit Intensität zu tun.“ (Jim, )	8	10
11	Material unzureichend	ProbandInnen wünschen sich zusätzliche Materialien oder kritisieren vorhandenes Material.	„Um das jetzt zu testen, ohne das Gerät zu öffnen, würde ich sowas ähnliches bauen, aber ich sehe jetzt irgendwie nicht so Materialien, die sich eignen.“ (James, 25)	8	19
12	Nichterkennen/Ignorieren von Mängeln am Modell	ProbandInnen erkennen physikalische Mängel nicht oder ignorieren diese und behalten sie aufgrund fehlender Alternativideen bei.	„da müssten die Behälter ja irgendwann voll sein Mist!“ (ignoriert und macht weiter; Jenny, 49)	4	8
13	Nichterkennen/Ignorieren von fehlender Passung	Physikalische oder datenbasierte Nichtpassung werden von den ProbandInnen nicht erkannt oder bewusst aufgrund fehlender Alternative beibehalten.	„Eigentlich kann ich das werfen, weil dazu passt mein Modell nicht.“ (verwirft Modell jedoch nicht; Cynthia, 33)	4	10

Einige Kategorien (2, 3, 4, 10) replizieren die Ergebnisse der empirischen Arbeiten, die für die deduktive Kategorienbildung herangezogen wurden. Parallelen zu anderen Arbeitsweisen, z. B. dem Experimentieren, sind besonders bei explorativen Tätigkeiten erkennbar. Andere Kategorien (1, 7, 12, 13) können dagegen als spezifisch für das Modellieren oder als möglicherweise spezifisch für das hier verwendete Blackbox-Setting (5, 6, 8, 11) angesehen werden, wobei im häufigen Auftreten der Kategorie 1 die zentrale Rolle der Analogien für das Modellieren deutlich wird. Zusammen mit Kategorie 7, die zeigt, dass einige ProbandInnen es nicht schaffen richtig erkannte Zusammenhänge zu modellieren (vgl. Windschitl & Thompson, 2006), liegt die Vermutung nahe, dass Modellieren und deklaratives Fachwissen in einem korreliert-abhängigen Verhältnis stehen (Nehring & Schwichow, 2020). Es ist zu vermuten, dass ein Auftreten von Hindernissen der Kategorie 1 im Fehlen relevanten, deklarativen Fachwissens zur Blackbox begründet sein könnte. Wenn das mangelnde Fachwissen, dass sich in fehlenden relevanten Analogien oder Erfahrungen äußert, nicht mit ausreichend Kreativität ausgeglichen werden kann, ist ein abduktiver Schluss vom Modell und damit die erkenntniserweiternde Entwicklung eines erklärenden

Modells nicht möglich (Krüger & Upmeyer zu Belzen, 2021; Magnani, 1999). Kategorien 4, 12 und 13 weisen darauf hin, dass auch der Wechsel zwischen Modell- und Erfahrungswelt (Leisner-Bodenthin, 2006), im Sinne eines Abgleichs zwischen Modell und den generierten Daten (Dounas-Frazer et al., 2018), den ProbandInnen schwer fällt, wenn die beobachteten Daten ihnen unlogisch erscheinen (Hammann, 2004). Dies wird in der Studie von Meister et al. (2021) aufgegriffen: Die AutorInnen schließen, dass unerwartete, anomale Daten, wie sie vom inneren Mechanismus der Blackbox produziert werden (z. B. wenn der Output den vorgenommenen Input überschreitet), oftmals nicht anhand des Modells, sondern methodisch begründet werden. Insbesondere führen die ProbandInnen beispielsweise mangelhafte eigene Fähigkeiten an (Meister et al., 2021).

### **Der Zusammenhang zwischen auftretenden Hindernissen und Modellierprozessen**

Im Hinblick auf den Zusammenhang zwischen den identifizierten Hindernissen und ihrem Auftreten in Bezug auf Phasen und Tätigkeiten des Modellierens (Fragestellung 2), wurden die Kodierungen der Modellierprozesse auf beiden Granularitätsebenen mit den Kodierungen der Hindernisse in Bezug gesetzt.

Auf Ebene der Phasen des Modellierens (vgl. Tabelle 8) wird insgesamt erkennbar, dass Kategorien, die eher explorative und datenorientierte Hindernisse beschreiben, auch vermehrt in Phasen der Exploration beobachtbar sind (Abbildung 16). Das Gleiche gilt für Kategorien, die eher als modell-spezifisch anzusehen sind, und vermehrt in Phasen der Herstellung auftreten. Das häufige Auftreten der jeweiligen Kategorie von Hindernissen in der passenden Phase zeigt, dass die ProbandInnen Hindernisse bereits während ihres Auftretens erkennen und verbalisieren. Hiermit wird allerdings auch impliziert, dass kaum ProbandInnen den eigenen Modellierprozess im Sinne einer metakognitiven, zielgerichteten Steuerung retrospektiv reflektieren oder prospektiv planen (Chiu & Lin, 2019). Zudem konnte nicht beobachtet werden, dass ProbandInnen ein Hindernis antizipieren, das noch gar nicht aufgetreten ist. Fraglich ist, ob von der retrospektiven Reflexion eines Hindernisses auf eine größere Auswirkung des Hindernisses auf den Modellierungsprozess geschlossen werden kann. Es könnte sein, dass nur gravierende Hindernisse, die den Problemlöseprozess nachhaltig behindern, eine Verbalisierung und gegebenenfalls eine Fehlersuche nötig machen. Diese Vermutung wird dabei unterstützt von der fehlenden Beobachtung spezifischer Hindernisse für die Phase der Vorhersage. So könnten ProbandInnen, die Vorhersagen aus ihrem Modell ableiten, als ProbandInnen betrachtet werden, die bereits auf naturwissenschaftlich adäquatem Niveau modellieren und das Problem damit erfolgreich gelöst haben. Allerdings könnte es auch sein, dass sich Hindernisse, die spezifisch für die Phase der Vorhersage sind, von den ProbandInnen nicht erkannt werden, da diese die Vorhersage gar nicht als Ziel der Problemlösung wahrnehmen. Dies erscheint unter Berücksichtigung bisheriger empirischer Studien zur Modellierkompetenz von Lehrkräften (Krell & Krüger, 2016; Nielsen & Nielsen, 2021a, 2021b; Windschitl & Thompson, 2006) als

## Ergebnisse und Diskussion der Einzelpublikationen

wahrscheinlicher, insbesondere da die Modellierprozesse der ProbandInnen durchaus ein großes Spektrum abdecken (Quigley et al., 2017) und nur selten Tätigkeiten der Vorhersage enthalten.

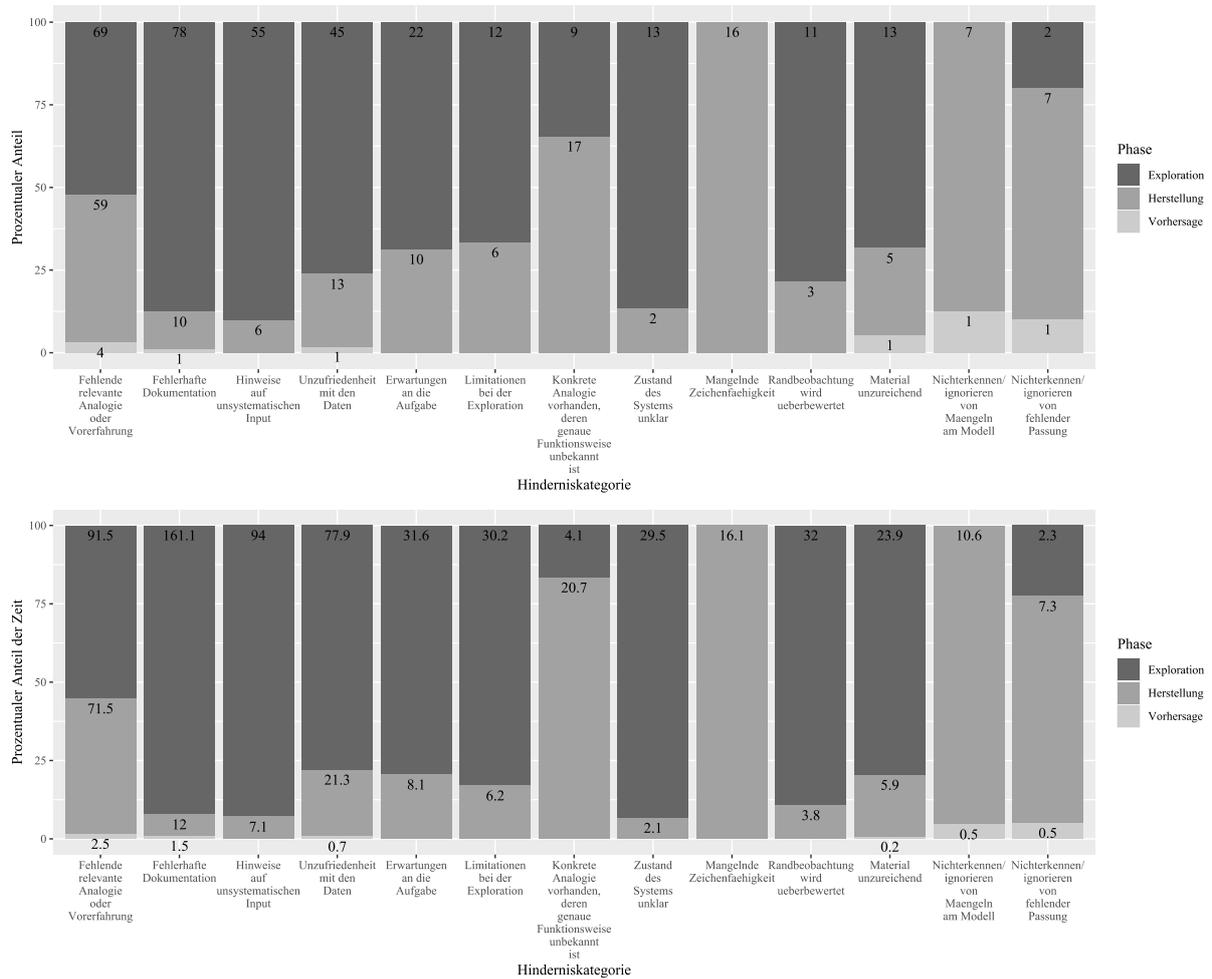


Abbildung 16: Gestapelte Säulen-Diagramme der Hindernisse unterteilt nach ihrem Auftreten in den Phasen des Modellierungsprozesses (verändert nach Göhner & Krell, 2021). Das obere Diagramm zeigt den Anteil jeder Phase am Auftreten des jeweiligen Hindernisses, das untere Diagramm den zeitlichen Anteil jeder Phase am Auftreten des jeweiligen Hindernisses. Zur besseren Sichtbarkeit der Verhältnisse wurden relative, prozentuale Werte dargestellt. Die Säulen sind zusätzlich mit den absoluten Werten beschriftet, die entweder die Häufigkeit (oben) oder die mit dem Hindernis verbrachten Minuten (unten) angeben.

Auf Ebene der einzelnen Tätigkeiten des Modellierens (Tabelle 15) zeigt sich zum einen die zentrale Rolle der Analogiebildung (Brown & Clement, 1989; Clement, 2008; Cuperman & Verner, 2019; Krell et al., 2019), da Kategorie 1 als Hindernis über alle Tätigkeiten des Modellierens hinweg beobachtet werden kann. Zum anderen wurde eine Schlüsselrolle der Tätigkeiten „Zusammenfassung“ und „Prüfung der Konsistenz zwischen dem Modell und den Daten“ deutlich: ProbandInnen verbalisieren die meisten Hindernisse während der Tätigkeit „Zusammenfassung“, die auch am häufigsten beobachtet wird (vgl. Sins et al., 2005). Darüber hinaus können alle Kategorien von Hindernissen während der Tätigkeit „Prüfung der Konsistenz zwischen dem Modell und den Daten“ beobachtet werden, was die

## Ergebnisse und Diskussion der Einzelpublikationen

Schlüsselrolle des Testens von Modellen zur Förderung der Modellierkompetenz bestärkt (Krell & Krüger, 2016; S. Lee & Kim, 2014).

Tabelle 15: Kreuztabelle der Hinderniskategorien und Tätigkeiten aller 36 ProbandInnen, die zeigt, wie viele Kodierungen je Hinderniskategorie mit den identifizierten Tätigkeiten gemeinsam auftreten (Göhner & Krell, 2021)

	1. Beobachtung eines Phänomens	2. Input/Output nicht hypothesengeleitet	3. Zusammenfassung	4. Input/Output hypothesengeleitet Daten	5. Mustererkennung	6. Analogien und Erfahrungen	7. Modellentwicklung	8. Optimierung Modellobjekt	9. Optimierung Modell retrospektiv	10. Verwerfen Modell	11. Prüfung Konsistenz Modellobjekt	12. Prüfung Konsistenz Modell/Daten	13. Feststellen Konsistenz	14. Hypothese aus Modell	15. Input/Output hypothesengeleitet Modell	16. Hypothese bestätigt	17. Hypothese falsifiziert	18. Optimierung Modell nach falsifizierter Hypothese	19. Verwerfen Modell nach falsifizierter Hypothese	Total
<b>1 Fehlende relevante Analogie oder Erfahrung</b>	30	4	19	3	1	16	8	2	4	3	12	25	2	0	0	0	3	0	0	132
<b>2 Fehlerhafte Dokumentation</b>	10	14	42	11	0	2	0	0	1	1	2	4	0	1	1	0	0	0	0	89
<b>3 Hinweise auf unsystematischen Input</b>	6	12	21	13	0	1	0	0	0	2	1	3	0	0	2	0	0	0	0	61
<b>4 Unzufriedenheit mit den Daten</b>	18	3	19	4	0	1	1	0	1	2	0	9	0	0	0	0	1	0	0	59
<b>5 Erwartungen an die Aufgabe</b>	7	3	8	3	0	4	0	0	1	3	0	2	1	0	0	0	0	0	0	32
<b>6 Limitationen bei der Exploration</b>	1	1	7	1	0	1	1	0	0	2	0	3	0	0	1	0	0	0	0	18
<b>7 Konkrete Analogie vorhanden, deren Funktionsweise unbekannt ist</b>	1	0	0	0	0	16	2	0	2	0	2	3	0	0	0	0	0	0	0	26
<b>8 Zustand des Systems unklar</b>	1	3	7	1	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	15
<b>9 Mangelnde Zeichenfähigkeiten</b>	0	0	0	0	0	2	10	1	0	0	2	1	0	0	0	0	0	0	0	16
<b>10 Randbeobachtung wird überbewertet</b>	0	3	6	1	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	14
<b>11 Material unzureichend</b>	1	4	7	1	0	1	0	0	0	0	1	3	0	0	0	1	0	0	0	19
<b>12 Nichterkennen/Ignorieren von Mängeln am Modell</b>	0	0	0	0	0	0	1	0	0	0	1	5	0	0	0	0	0	1	0	8
<b>13 Nichterkennen/Ignorieren von fehlender Passung</b>	1	0	1	0	0	1	0	0	0	0	0	4	2	0	0	0	0	1	0	10
<b>Total</b>	76	47	137	38	1	48	23	3	9	13	21	66	5	1	4	1	4	2	0	499

Im Hinblick auf dem Zusammenhang zwischen den identifizierten Hindernissen und den gesamten Modellierprozessen der ProbandInnen ergaben sich keine eindeutigen Zusammenhänge. So zeigen ProbandInnen, die auf ähnliche Hindernisse stoßen, keineswegs ähnliche Modellierprozesse und auch statistisch konnten keine signifikanten Unterschiede zwischen den sechs identifizierten Modellierungsstrategien und der Häufigkeit der einzelnen Hindernisse festgestellt werden (ANOVAs;  $p > .05$  für alle Hindernisse). Ebenso konnte ein Gruppenvergleich zwischen ProbandInnen, die Tätigkeiten der Vorhersage

## Ergebnisse und Diskussion der Einzelpublikationen

in ihrem Modellierprozess umsetzen und solchen, die keine Tätigkeiten der Vorhersage in ihrem Modellierprozess umsetzen, keine signifikanten Unterschiede im Auftreten der Hindernisse feststellen ( $t$ -Test;  $p > .05$ ). In der Publikation wurden daher die Fälle James und Jasmin qualitativ näher betrachtet (Göhner & Krell, 2021). Diese detaillierte Fallanalyse liefert Evidenz für die Relevanz der identifizierten Hindernisse in den Modellierungsprozessen (Schreier, 2010) und weist darauf hin, dass vor allem der *Umgang* mit den auftretenden Hindernissen zentral für den weiteren Modellierungsprozess sein könnte (vgl. Funke, 2011; Mayer, 2007). So entwickelt James beim Auftreten eines Hindernisses zumeist sein Modell weiter, wohingegen Jasmin die Blackbox weiter exploriert, mit dem Ziel, bessere Daten zu generieren (vgl. Angelina; Göhner et al., 2022; Göhner & Krell, 2018).

Insgesamt liefert die Publikation damit zwar nicht die gewünschte Erklärung, warum in der vorliegenden Untersuchung nur selten Modellierprozesse beobachtet wurden, die Tätigkeiten der Vorhersage umfassen. Sie bietet trotzdem einen Ausgangspunkt für die Entwicklung problemtypenbasierter Kompetenzmodellierungen. Diese können beispielsweise die Entwicklung von Diagnoseinstrumenten anleiten (Baur, 2018; Baur & Emden, 2020; Gut et al., 2014) oder zur Konstruktion von gezielten und expliziten Förderangeboten beitragen, welche Modelle und das Modellieren als Werkzeuge zur Erkenntnisgewinnung wahrnehmbar und praktisch umsetzbar machen (Clement, 2000; Lehrer & Schauble, 2006a). Allerdings erscheint weitere Forschung nötig, die untersucht, ob die Ergebnisse auch auf andere Modelleraufgaben generalisierbar sind.

## 5.5 Publikation 5: Untersuchung des Zusammenhangs zwischen Modellierprozessen und den Samplingkriterien

Göhner, M., & Krell, M. (2022). Analyzing the relationships between pre-service biology teachers' modeling processes, scientific reasoning competencies and general cognitive abilities. In M. Ergazaki & K. Kampourakis (Eds.), *Contributions from Biology Education Research* (S. 231–244). Springer Nature. [https://doi.org/10.1007/978-3-030-89480-1\\_18](https://doi.org/10.1007/978-3-030-89480-1_18).

### **Zusammenfassung**

In dieser Publikation werden die individuellen Modellierprozesse mit den Personenmerkmalen „Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung“ und „allgemeine kognitive Fähigkeiten“ in Beziehung gesetzt, welche als Samplingkriterien eingesetzt wurden. In der vorliegenden Untersuchung wurde damit die Annahme geprüft, dass diese beiden Personenmerkmale im Rahmen der Modellierkompetenz wichtige Dispositionen darstellen, die den Modellierprozess als Performanz beeinflussen (vgl. Blömeke, König et al., 2015). Es wird die folgende Forschungsfrage verfolgt: Welche Zusammenhänge gibt es zwischen den Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und den allgemeinen kognitiven Fähigkeiten der ProbandInnen und ihren jeweiligen Modellierprozessen? Es wurde erwartet, dass ProbandInnen mit höheren Werten bezüglich ihrer Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und allgemeinen kognitiven Fähigkeiten elaboriertere Modellierprozesse umsetzen, die sich durch das vermehrte Auftreten von Tätigkeiten der Vorhersage (Gouvea & Passmore, 2017; Krüger & Upmeier zu Belzen, 2021; Upmeier zu Belzen & Krüger, 2010; Upmeier zu Belzen et al., 2019), sowie eine höhere Komplexität und Homogenität auszeichnen (Schwarz et al., 2009; Wilmont et al., 2019). Methodisch wurden diese Zusammenhänge geprüft, indem die Korrelation (Pearson) zwischen den relevanten Variablen ermittelt wurde: Hierzu gehören die für alle ProbandInnen individuell ermittelten Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung (Testergebnis des Ko-WADiS; Krüger et al., 2020b), die allgemeinen kognitiven Fähigkeiten und deren Subskalen (Testergebnis des I-S-T 2000R; Liepmann et al., 2007), der zeitliche Anteil in den vier Phasen des Modellierens<sup>12</sup>, sowie die Komplexität und Homogenität der Modellierprozesse (vgl. Abschnitt zu quantifizieren, S. 71; Göhner & Krell, 2022b). Alle Daten wurden normalisiert, indem jede Variable einzeln z-standardisiert wurde.

---

<sup>12</sup> Die Analogiebildung wird hier als Sonderfall explizit aufgeführt, da sie in allen Phasen des Modellierens auftreten kann (vgl. Tabelle 8). Für die folgende Analyse wurde sie daher als eigene Phase des Modellierens behandelt.

**Der Zusammenhang zwischen Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung, allgemeinen kognitiven Fähigkeiten und Modellierprozessen**

Deskriptive Statistiken bezüglich der fünf Variablen sind in Tabelle 16 dargestellt.

Tabelle 16: Deskriptive Statistiken zu den fünf Variablen (verändert nach Göhner & Krell, 2022a). Es sind jeweils die Durchschnittswerte und (in Klammern) die Standardabweichung angegeben. Die Variablen Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und allgemeine kognitive Fähigkeiten sind entsprechend der in der englisch-sprachigen Publikation genutzten Begriffe *scientific reasoning competencies* und *general cognitive abilities* respektive als SRC und GCA abgekürzt.

	Gesamtstichprobe (N=36)	Gruppen des qualitativen Stichprobenplans			
		Hohe SRC, hohe GCA (n=17)	Hohe SRC, niedrige GCA (n=6)	Niedrige SRC, hohe GCA (n=3)	Niedrige SRC, niedrige GCA (n=10)
<b>Ø SRC (SD)</b>	0,21 (1,04)	0,97 (0,23)	0,91 (0,20)	-1,09 (0,20)	-1,13 (0,29)
<b>Ø GCA (SD)</b>	114,56 (25,24)	135,76 (9,14)	90,33 (12,47)	131,33 (6,11)	88,00 (10,81)
<b>Verbale Intelligenz</b>	38,50 (6,89)	38,24 (7,30)	39,00 (7,13)	41,00 (3,61)	37,90 (7,48)
<b>Numerische Intelligenz</b>	39,03 (13,46)	39,94 (12,44)	38,50 (12,11)	33,00 (25,00)	39,60 (13,95)
<b>Figurale Intelligenz</b>	37,03 (8,29)	36,35 (9,99)	40,83 (4,62)	34,33 (8,39)	36,70 (7,01)
<b>Ø Zeit (%) in Phase des Modellierens (SD)</b>					
<b>Exploration</b>	62 (18)	65 (16)	58 (21)	66 (25)	58 (18)
<b>Herstellung</b>	21 (12)	22 (12)	18 (7)	25 (25)	21 (12)
<b>Vorhersage</b>	3 (7)	2 (4)	6 (12)	0 (0)	4 (7)
<b>Analogiebildung</b>	5 (4)	3 (3)	7 (5)	6 (1)	6 (6)
<b>Ø Netzwerkmetriken der Modellierprozesse (SD)</b>					
<b>Komplexität</b>	4,89 (2,68)	5,53 (2,96)	4,83 (2,79)	3,33 (0,58)	4,30 (2,45)
<b>Homogenität</b>	1,52 (0,41)	1,48 (0,34)	1,55 (0,49)	1,26 (,57)	1,66 (0,45)

In Bezug auf die eigentliche Fragestellung konnte lediglich eine signifikante Korrelation ermittelt werden (Tabelle 17): Allgemeine kognitive Fähigkeiten korrelieren signifikant negativ mit mittlerer Effektstärke mit dem zeitlichen Anteil der Analogiebildung während der Modellierprozesse. Es liegt nahe, dass ProbandInnen mit niedrigen Werten in Bezug auf allgemeine kognitive Fähigkeiten mehr Zeit benötigen, um passende Analogien zu bilden, als ProbandInnen mit höheren Werten in Bezug auf allgemeine kognitive Fähigkeiten.

Tabelle 17: Pearson Korrelations-Koeffizienten der Zusammenhänge zwischen den relevanten Variablen (n. s. = p>0,05)

Variable	SRC		GCA		Verbale Intelligenz		Numerische Intelligenz		Figurale Intelligenz	
	r	p	r	p	r	p	r	p	r	p
<b>Exploration</b>	0,06	n.s.	0,20	n.s.	0,09	n.s.	-0,01	n.s.	-0,11	n.s.
<b>Herstellung</b>	0,01	n.s.	0,07	n.s.	-0,03	n.s.	-0,04	n.s.	-0,04	n.s.
<b>Vorhersage</b>	0,02	n.s.	-0,21	n.s.	-0,03	n.s.	0,03	n.s.	0,19	n.s.
<b>Analogiebildung</b>	-0,23	n.s.	-0,38	<.05	0,12	n.s.	0,20	n.s.	0,29	n.s.
<b>Komplexität</b>	0,21	n.s.	0,06	n.s.	0,25	n.s.	0,16	n.s.	0,00	n.s.
<b>Homogenität</b>	-0,08	n.s.	-0,20	n.s.	-0,06	n.s.	0,14	n.s.	-0,05	n.s.

Auch Gruppenvergleiche bezüglich der beiden Variablen Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und allgemeine kognitive Fähigkeiten mit den sechs identifizierten

Modellierungsstrategien (ANOVAs) und ProbandInnen, die Tätigkeiten der Vorhersage umsetzen und solchen, die keine Tätigkeiten der Vorhersage umsetzen (T-Tests), zeigten keine systematischen Unterschiede<sup>13</sup>.

Insgesamt zeigen diese Ergebnisse, dass die Erwartung, ProbandInnen mit höheren Werten bezüglich ihrer Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und allgemeinen kognitiven Fähigkeiten würden elaboriertere Modellierprozesse umsetzen, nicht bestätigt werden kann. Zwar legt das Fehlen eines systematischen Zusammenhangs zwischen Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und den umgesetzten Modellierprozessen auf den ersten Blick eine hohe Abgrenzbarkeit des Modellierens von anderen naturwissenschaftlichen Arbeitsweisen nahe (vgl. Modellieren als ein abgrenzbarer *style of scientific reasoning*; Kind & Osborne, 2017), das hier eingesetzte Ko-WADiS-Instrument umfasst allerdings auch zur Hälfte Fragen zum Modellieren (vgl. Abschnitt 4.5, S. 62), so dass dieser Schluss fraglich erscheint. Auch das Fehlen eines systematischen Zusammenhangs zwischen den allgemeinen kognitiven Fähigkeiten der ProbandInnen und ihren Modellierprozessen ist überraschend, da insbesondere die figurale Intelligenz und das damit einhergehende räumliche Vorstellungsvermögen als wichtige Voraussetzung für erfolgreiches Modellieren angenommen werden (Clement, 2008; Nersessian, 2002). Während letztendlich auch ganz andere Dispositionen für Modellierprozesse relevant sein könnten, deuten die vorliegenden Ergebnisse zusammen mit früheren, teils widersprüchlichen Ergebnissen hinsichtlich des Personenmerkmals Kreativität (Bailer-Jones, 1 ; Demirhan & Şahin, 2021; Mierdel & Bogner, 2021) eher auf einen großen Einfluss von situationsspezifischen Aufgabenmerkmalen auf die Modellierprozesse hin. So konnte Bailer-Jones (1999) zeigen, dass Kreativität ein wichtiger Faktor für das erfolgreiche Modellieren von NaturwissenschaftlerInnen ist, ähnliche Zusammenhänge konnten aber bei SchülerInnen und Lehrkräften nicht gefunden werden (Demirhan & Şahin, 2021; Mierdel & Bogner, 2021). Es ist zu vermuten, dass Aufgabenmerkmale, wie eine hohe Schwierigkeit der Modellieraufgabe (z. B. das Modellieren von extragalaktischen Radiowellen; Bailer-Jones, 1999) moderierend wirken und andere Dispositionen (z. B. domänenspezifisches Wissen; Ruppert et al., 2019) erforderlich machen, die für einfachere Modellieraufgaben nicht erforderlich sind (z. B. das Modellieren der DNS-Struktur; Mierdel & Bogner, 2019). Wie Kompetenz im Allgemeinen (Blömeke et al., 2013; Rychen & Salganik, 2003) erscheint damit auch Modellierkompetenz im Speziellen kontext-abhängig zu sein.

---

<sup>13</sup> In Göhner und Krell (2022a) sind aus Platzgründen nur Gruppenvergleiche zwischen den ProbandInnen, die Tätigkeiten der Vorhersage umsetzen und solchen, die keine Tätigkeiten der Vorhersage umsetzen publiziert.

## Ergebnisse und Diskussion der Einzelpublikationen

Zusammengefasst zeigen die in dieser Publikation berichteten Ergebnisse, dass die von vielen WissenschaftlerInnen angenommenen Zusammenhänge zwischen spezifischen Personenmerkmalen und der Performanz beim Modellieren, empirisch nicht unbedingt als valide aufgefasst werden können. Es erscheint immanent wichtig, auch theoretisch fundierte, logisch erscheinende Annahmen kritisch zu reflektieren und empirisch zu überprüfen. Gerade im Hinblick auf Förderungskontexte kann somit davon ausgegangen werden, dass eine Förderung anderer, theoretisch mit Modellierkompetenz zusammenhängender Fähigkeiten nicht ausreicht, um Modellierkompetenz zu fördern.

## 5.6 Publikation 6: Untersuchung des Zusammenhangs zwischen Modellverstehen, Modellierprozessen und der Qualität des Modellierprodukts

Göhner, M., Bielik, T., & Krell, M. (2022). Investigating the dimensions of the modeling competence among pre-service science teachers: Meta-modeling knowledge, modeling practice and modeling product. *Journal of Research in Science Teaching*, 59(8), 1354–1387. <https://doi.org/10.1002/tea.21759>. Lizenziert nach [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

### **Zusammenfassung**

In dieser Publikation werden die drei in der Literatur beschriebenen Dimensionen der Modellierkompetenz Modellverstehen, Modellierprozesse und Modellierprodukte (Chiu & Lin, 2019; Louca & Zacharia, 2012; Nielsen & Nielsen, 2021a) auf ihre Zusammenhänge hin überprüft. Die Publikation folgt der Fragestellung „Welche Zusammenhänge gibt es zwischen dem Modellverstehen, den Modellierprozessen und der Qualität der Modellierprodukte der Biologie-Lehramtsstudierenden?“. Dabei wurde literaturbasiert erwartet, dass alle drei Dimensionen der Modellierkompetenz positiv zusammenhängen (Gobert & Pallant, 2004; S. Lee & Kim, 2014; Schwarz & White, 2005). Zusätzlich zu den bereits vorliegenden Werten hinsichtlich des Modellverstehens und der Modellierprozesse der ProbandInnen (inkl. Komplexität und Heterogenität; Göhner & Krell, 2022b) wurde auch das Modellverstehen der ProbandInnen *während des Modellierprozesses*, sowie die *Qualität des Modellierproduktes* anhand weiterer qualitativer Inhaltsanalysen erfasst. Eine ausführliche Beschreibung der Methodik findet sich in der Publikation. Die resultierenden fünf Variablen dekontextualisiertes Modellverstehen, kontextualisiertes Modellverstehen, Komplexität des Modellierprozesses, Heterogenität des Modellierprozesses und Qualität des Modellierprodukts wurden auf Zusammenhänge hin geprüft, indem die Korrelationen zwischen den fünf Variablen (Spearman’s Rangkorrelation) ermittelt wurden. Ausgehend von den Ergebnissen der quantitativen Analyse wurden die bereits in Publikation 2 beschriebenen Fälle Angelina und Raphael als repräsentativ für die gefundenen Zusammenhänge eingestuft und erneut detailliert dargestellt, um mögliche Erklärungen für die gefundenen Zusammenhänge zu finden. Im Rahmen der vorliegenden Untersuchung trägt diese ganzheitliche Betrachtung dazu bei die Struktur des Konstrukts Modellierkompetenz zu beleuchten und insbesondere den Zusammenhang zwischen individuell umgesetztem Modellierprozess (Performanz) und konzeptuellem Modellverstehen (Metakognition) zu beleuchten, dessen Erforschung als „one of the most pressing needs for future research“ in der Naturwissenschaftsdidaktik beschrieben wird (Louca & Zacharia, 2012, S. 486).

## Ergebnisse und Diskussion der Einzelpublikationen

### Die Dimensionen der Modellierkompetenz einzeln

Betrachtet man jede Variable einzeln, stimmen die Ergebnisse weitestgehend mit den Ergebnissen früherer empirischer Studien überein. Tabelle 18 zeigt eine Übersicht über die ermittelten quantitativen Werte für alle ProbandInnen.

Tabelle 18: Übersicht aller ProbandInnen und derer quantitativen Werte hinsichtlich Modellverstehen, Modellierprozess und Modellierprodukt

ProbandIn	Modellverstehen					Modellierprozess	Modellierprodukt									
	Dekontextualisiert (Fragebogen)						Kontextualisiert (im Modellierprozess verbalisiert)			Konzepte integriert						
	Eigenschaften von Modellen	Alternative Modelle	Zweck von Modellen	Modelle testen	Modelle ändern	Eigenschaften von Modellen	Alternative Modelle	Zweck von Modellen	Modelle testen	Modelle ändern	Komplexität	Homogenität	Zeit (hh:mm:ss)	verbal	gezeichnet	verbal/gezeichnet
Carlo	1	2	3	3	2	1					6	0,81	01:15:29	2	0	2
Alexandra	2	3	2	2	2						9	1,50	01:38:26	1	1	1
Amy	2	2	2	2	2						10	1,46	01:22:13	2	2	2
Angelina	2	2	2	2	2	3		2			11	0,93	01:01:17	2	2	2
Celine	2	2	2	2	3	2		2	2	2	9	1,36	00:50:52	1	1	1
Claire		2	1		1	1					12	0,93	01:32:42	1	2	2
Claudia	2	2	1	2	2	3	3		2		7	1,24	00:33:30	1	1	1
Jasmin	3	3	3	2	3			3			11	1,88	01:28:25	1	1	1
Jim	2	2	3	2	2	3					10	0,93	01:04:14	1	1	1
Lauren	2	2	3	3	2	3		2			11	1,75	01:16:53	1	1	1
Misty	2	2	2	2	2						11	1,45	01:25:20	0	0	0
Natalie	2	3	2	2	2						12	1,47	01:23:55	1	2	2
Selena	3	3	3	3	3	3					10	1,82	00:43:56	1	1	1
Donna	2	2	1		2	3		3	3	1	12	1,19	01:23:46	1	1	1
Kara	2	1	1	2	2	1			2	2	7	1,38	00:08:07	1	1	1
Achilles	1	2	1	2	2						11	2,43	01:32:02	1	1	1
Floyd	2	2	3	3	3			2			13	1,96	00:33:55	3	2	3
Jonathan	2	2	3	2	2			2	2		10	1,91	00:50:13	2	1	2
Rocco	2	2	1		2	3					10	2,17	00:29:28	1	1	1
Susi	2	3	3	3	3		1		2	2	13	1,98	01:34:54	3	2	3
Valerie	2	3	2		2	3	3	2			12	2,20	00:46:21	1	1	1
Alice	2	2	1	2	2				3		16	1,32	01:38:22	3	2	3
Ben	2	3	3	2	2				3		16	1,43	01:47:20	1	2	2
Boris	2	2	2	3	2	2				2	13	1,85	00:11:06	1	1	1
Cynthia	2	3	3	3	3	3		3	3	3	14	1,14	01:40:43	1	1	1
Daphne	1	2	1	1	1		1	2		1	16	1,32	00:50:39	1	1	1
Iris	2	2	2	2	2	3		1			12	1,19	01:25:16	1	0	1
May	2	2	2	3	3						15	1,25	01:27:16	1	1	1
Ryan	3	3	3	3	2	3	1	2	3	3	15	1,11	01:05:03	1	1	1
Sabrina	2	2		2	2	3					13	1,89	00:34:05	0	1	1
Frida	1	3	2	1	1				3		16	1,75	01:49:12	2	2	2
James	2	2	3	2	2	3	3	2	3	3	14	2,23	00:26:37	2	1	2
Jenny	2	3				3		3	3		18	1,52	01:52:03	2	2	2
Lana	2	3	2		2	3	3		3	3	17	1,00	00:48:33	2	2	2
Martin	2	2	3	3	2	1		1	3	3	18	1,52	01:52:22	2	1	2
Raphael	2	2	2	2	2	3		2			17	1,56	01:05:58	3	2	3

Proband „Rocco“ wurde in der Publikation aus der Stichgruppe ausgeschlossen, da er als einziges im Lehramtsstudiengang für die Grundschule studiert hat. Im Folgenden wird er dagegen in alle Werte miteinbezogen.

## Modellverstehen

Das dekontextualisierte Modellverstehen der ProbandInnen impliziert eine mediale Perspektive auf Modelle (Justi & Gilbert, 2003; Krell & Krüger, 2016; Torres & Vasconcelos, 2015; Upmeier zu Belzen et al., 2019). Sowohl elaborierte Antworten, in denen Modelle als epistemische Werkzeuge beschrieben werden (vgl. Niveau III; Krell & Krüger, 2016), als auch naiv-realistische Antworten, in denen Modelle als Kopien der Realität beschrieben werden (vgl. Niveau I; Krell & Krüger, 2016), sind in der vorliegenden Untersuchung eher selten, obwohl naiv-realistische Antworten in manchen Studien durchaus häufig beobachtet werden (Torres & Vasconcelos, 2015).

Tabelle 19: Verteilung der Antwortniveaus hinsichtlich der fünf Aspekte des Modellverstehens aller 36 ProbandInnen (verändert nach Göhner et al., 2022)

Art des Modellverstehens (Quelle)	Aspekt	N	n Antworten			Ø (± SD)
			Niveau I	Niveau II	Niveau III	
Dekontextualisiert (Antworten im Fragebogen)	Eigenschaften von Modellen	35	4	28	3	1,97 (±0,45)
	Alternative Modelle	36	1	23	12	2,31 (±0,53)
	Zweck von Modellen	34	8	13	13	2,15 (±0,78)
	Testen von Modellen	30	2	18	10	2,27 (±0,58)
	Ändern von Modellen	35	3	25	7	2,11 (±0,53)
Kontextualisiert (Verbalisierungen während des Modellierprozesses)	Eigenschaften von Modellen	22	4	2	16	2,55 (±0,80)
	Alternative Modelle	7	3	0	4	2,14 (±1,07)
	Zweck von Modellen	16	2	10	4	2,13 (±0,62)
	Testen von Modellen	15	0	5	10	2,67 (±0,48)
	Ändern von Modellen	11	2	4	5	2,27 (±0,79)

Innerhalb der Modellierprozesse konnten 126 Aussagen identifiziert werden, die als kontextualisiertes Modellverstehen interpretiert wurden. Entsprechend der Bewertung von Antworten auf den Fragebogen wurde auch in den Modellierprozessen der ProbandInnen immer das elaborierteste Niveau für jeden Aspekt berücksichtigt. Obwohl in Fallstudien beobachtet werden konnte, dass spezifische Kontexte, wie die hier verwendete Blackbox, besonders elaborierte Perspektiven auf Modelle provozieren (Ke & Schwarz, 2020), konnte dies in der vorliegenden Untersuchung lediglich für den Aspekt *Eigenschaften von Modellen* gezeigt werden. Das durchschnittliche Niveau der Aussagen zum Modellverstehen ist für diesen Aspekt in den Verbalisierungen während des Modellierprozesses (kontextualisiertes Modellverstehen) als einziges signifikant höher (Wilcoxon-Vorzeichen-Rang-Test:  $z = -3,00$ ,  $p = 0,004$ ;  $d = 1,72$ , große Effektstärke) als in den Antworten auf den Fragebogen (dekontextualisiertes Modellverstehen). Als mögliche Erklärung kommt in Frage, dass das Modellieren der Blackbox dazu anregt den hypothetischen Charakter von Modellen zu reflektieren. Alle Aspekte betrachtet, ergibt sich allerdings kein signifikanter Unterschied zwischen dekontextualisiertem und kontextualisiertem Modellverstehen ( $p > 0,05$ ). Zudem variieren die Niveaus der Aussagen einzelner ProbandInnen über die fünf Aspekte hinweg zum Teil stark (Krell et al., 2014b).

## Modellierprozesse

Auf eine erneute Darstellung der Modellierprozesse wurde in der vorliegenden Untersuchung verzichtet, da die hier relevanten Ergebnisse bereits in Publikation 2 (Göhner & Krell, 2018; S. 75) und insbesondere Publikation 3 (Göhner & Krell, 2022b; S. 79) dargestellt sind. Für die vorliegende Publikation und damit einhergehende Untersuchung der Zusammenhänge zwischen Modellverstehen, Modellierprozessen und Modellierprodukten wurden primär die quantitativen Werte der Komplexität und Homogenität genutzt.

## Modellierprodukte

32 ProbandInnen präsentierten ein einziges Modellierprodukt als finale Lösung. Zwei ProbandInnen, Claudia und Kara, präsentierten mehrere Modellierprodukte als mögliche finale Lösungen (Claudia fünf, Kara respektive zwei). Die beiden verbleibenden ProbandInnen, Carlo und Iris, präsentieren gar kein Modellierprodukt als finale Lösung; Carlo stellt hierbei überhaupt kein Modellierprodukt her und Iris verwirft ihr Modellierprodukt wieder, da sie es nicht als mögliche Lösung ansieht. Die Modellierprodukte der ProbandInnen (alle dargestellt im Anhang, Abschnitt 9.3.1, S.IICXXXIV) wurden inhaltlich bewertet, indem jedes Modell im Hinblick auf das Vorhandensein dreier Zielkonzepte untersucht wurde. Diese Zielkonzepte umfassen die folgenden Ideen: 1. Ein beliebiges Gefäß füllt sich mit Wasser (z. B. Abbildung 17a); 2. Mehrere Gefäße werden in einem parallelen System gleichzeitig mit Wasser gefüllt (z. B. Abbildung 17b); 3. Ein beliebiger Mechanismus entleert ein Gefäß ab einem spezifischen Grenzwert vollständig (z. B. Abbildung 17c). Diese Zielkonzepte können entweder gezeichnet vorliegen oder während des lauten Denkens verbalisiert werden. Alternative Konzepte, wie zum Beispiel eine zufällig gesteuerte Öffnung, werden zwar verbalisiert, aber von den ProbandInnen oft direkt wieder verworfen und wurden deshalb hier nicht näher beleuchtet. Im Mittel integrieren die ProbandInnen ein bis zwei der Zielkonzepte ( $MW = 1,56$ ,  $SD = 0,74$ ). Lediglich vier ProbandInnen integrieren alle drei Konzepte, wobei auch von diesen niemand alle drei Konzepte *zeichnet*.

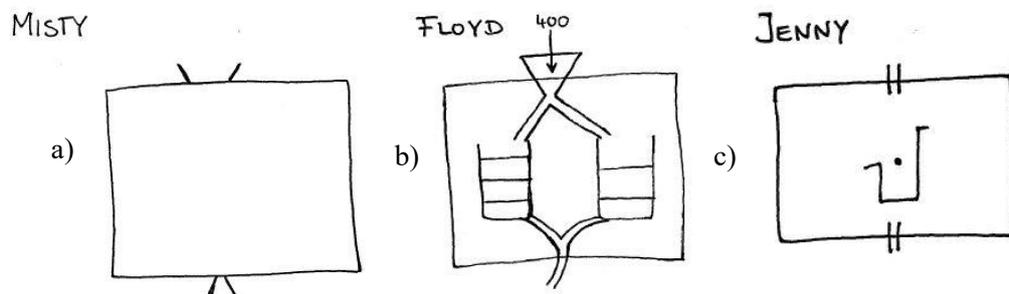


Abbildung 17: Beispiele für Modellierprodukte. Misty's Modellierprodukt integriert Konzept 1. Floyd's Modellierprodukt integriert die Konzepte 1 und 2. Jenny's Modellierprodukt integriert die Konzepte 1 und 3.

## Ergebnisse und Diskussion der Einzelpublikationen

Tabelle 20 zeigt eine Übersicht über die Integration der drei Konzepte durch die ProbandInnen.

Tabelle 20: Übersicht über die Integration der drei Konzepte durch die ProbandInnen in den Modellierprodukten

<b>ProbandInnen haben das Konzept</b>	<b>Konzept 1</b> (beliebiges Gefäß füllt sich mit Wasser)	<b>Konzept 2</b> (paralleles System verteilt Wasser)	<b>Konzept 3</b> (Mechanismus entleert Gefäß bei Grenzwert vollständig)
verbalisiert	33 (91,67 %)	7 (19,44 %)	11 (30,56 %)
gezeichnet	33 (91,67 %)	9 (25,00 %)	3 (8,33 %)
verbalisiert oder gezeichnet	35 (97,22 %)	9 (25,00 %)	12 (33,33 %)

### Zusammenhänge zwischen den drei Dimensionen der Modellierkompetenz

Korrelationsanalysen (Spearman's Rangkorrelation) der fünf Variablen dekontextualisiertes Modellverstehen, kontextualisiertes Modellverstehen, Komplexität der Modellierprozesse, Homogenität der Modellierprozesse und Anzahl der integrierten Konzepte in das Modellierprodukt zeigten, bis auf eine Ausnahme, keine signifikanten Zusammenhänge an (Tabelle 21); lediglich die Komplexität der Modellierprozesse und die Anzahl integrierter Konzepte in das Modellierprodukt korrelieren signifikant miteinander (schwache Effektstärke).

Tabelle 21: Korrelationstabelle der Spearman'schen Rangkorrelation zwischen den Variablen dekontextualisiertes Modellverstehen, kontextualisiertes Modellverstehen, Komplexität der Modellierprozesse, Homogenität der Modellierprozesse und Anzahl der integrierten Konzepte in das Modellierprodukt

<b>Variable</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
1. Dekontextualisiertes Modellverstehen	0,22 (n.s.)	0,19 (n.s.)	0,15 (n.s.)	0,03 (n.s.)
2. Kontextualisiertes Modellverstehen		0,20 (n.s.)	0,15 (n.s.)	-0,05 (n.s.)
3. Komplexität der Modellierprozesse			-0,03 (n.s.)	0,38 (<.05)
4. Homogenität der Modellierprozesse				0,13 (n.s.)
5. Anzahl der integrierten Konzepte in das Modellierprodukt				

Die Stichprobengröße für das kontextualisierte Modellverstehen umfasst nicht alle ProbandInnen ( $N=30$ ), für alle anderen Variablen wurden die Werte aller 36 ProbandInnen berücksichtigt

Auch ein Gruppenvergleich (Wilcoxon Rangsummentest) zwischen ProbandInnen, die Tätigkeiten der Vorhersage umsetzen und ProbandInnen, die keine Tätigkeiten der Vorhersage umsetzen, zeigte keine signifikanten Unterschiede hinsichtlich der anderen drei Variablen dekontextualisiertes Modellverstehen ( $p = 0,61$ ), kontextualisiertes Modellverstehen ( $p = 0,19$ ) oder Anzahl der integrierten Konzepte in das Modellierprodukt ( $p = 0,51$ ).

Zusammenfassend lässt sich damit festhalten, dass entgegen der theoretischen Annahmen (Gobert & Pallant, 2004; S. Lee & Kim, 2014; Schwarz & White, 2005) kaum Hinweise auf signifikante quantitative Zusammenhänge zwischen den verschiedenen Dimensionen der Modellierkompetenz in der vorliegenden Untersuchung gefunden werden konnten. In der Publikation werden deshalb die beiden Fälle

## Ergebnisse und Diskussion der Einzelpublikationen

Angelina und Raphael, die als repräsentativ für die Ergebnisse der quantitativen Analyse aufgefasst werden können (vgl. Tabelle 18), erneut aufgegriffen. Während die Fallanalysen in der veröffentlichten Publikation detailliert dargestellt sind, wird im Folgenden die Synthese der quantitativen und qualitativen Analyse dargestellt. Diese zielt darauf ab, mögliche qualitative Erklärungen in den qualitativen Daten für das Vorhandensein oder Fehlen der jeweiligen quantitativen Zusammenhänge zu finden.

Für den einzigen signifikant positiven Zusammenhang zwischen der Komplexität des Modellierprozesses und der Qualität der Modellierprozesse sind mehrere Erklärungen denkbar. Theoretisch kann angenommen werden, dass Modellierprodukte von niedriger Qualität Modellierende in ihrem Prozess limitieren, da von diesen Modellierprodukten keine sinnvollen Hypothesen abgeleitet werden können. In den qualitativen Daten lassen sich jedoch keinerlei solcher Hinweise erkennen, ganz im Gegenteil, der Fall von Raphael zeigt, dass manche ProbandInnen auch vom denkbar einfachsten Modellierprodukt, einem einzelnen Gefäß, erfolgreich Hypothesen ableiten und testen können, die zur Erkenntnisgewinnung und Weiterentwicklung des Modellierprodukts genutzt werden. Gleichzeitig zeigt der Fall von Raphael aber auch, dass ein gutes Modellierprodukt nicht zwangsläufig zu Modellierprozessen führt, die Tätigkeiten der Vorhersage umsetzen. Sein finales Modellierprodukt integriert alle drei Konzepte, wird von Raphael aber nicht mehr für Vorhersagen genutzt oder getestet. Genau diesem Testen eines Modells, mit dem Ziel es zu falsifizieren, wird in der Literatur allerdings zentrale Relevanz für das Modellieren zugesprochen (Giere et al., 2006). Der Vergleich mit dem Modellierprozess von Angelina könnte eventuell darauf hindeuten, dass inhaltliche Qualitätskriterien, wie hier die Anzahl der integrierten Konzepte, nicht unbedingt erklären können, wann ein Modellierprodukt für Modellierprozesse genutzt wird, die Tätigkeiten der Vorhersage umsetzen. Vielmehr könnte die Wahrnehmung des Modellierprodukts durch den Modellierenden eine wichtige Rolle spielen (in der Publikation wird dies mit dem Begriff *perceived soundness* umschrieben). Während Raphael seine Modellierprodukte von Anfang an als fruchtbar für das Testen seiner Ideen wahrnimmt, traut sich Angelina erst ganz am Schluss ihres Modellierprozesses überhaupt ein Modellierprodukt zu zeichnen und verbalisiert fortwährend ihre eigene Unsicherheit und Zweifel, ob dieses Modellierprodukt richtig sein könnte. Eine Analyse der von ProbandInnen während des Modellierprozesses verbalisierten epistemischen Qualitätskriterien (vgl. Pluta et al., 2011) könnte in weiteren Untersuchungen zielführend sein, um diese Erklärung zu prüfen. Gleichzeitig weist der hier gefundene quantitative Zusammenhang durchaus auf eine gewisse Validität der Analyse inhaltlicher Qualitätskriterien von Modellierprodukten hin, die in der wissenschaftlichen Praxis oft als Ersatz für die zeitintensive Analyse von Modellierprozessen, herangezogen werden (Bamberger & Davis, 2013; Cheng & Lin, 2015; Ergazaki et al., 2007; Schwarz et al., 2009).

Das Fehlen eines Zusammenhangs zwischen dem Modellverstehen und den Modellierprozessen konnte in der vorliegenden Publikation nun auch quantitativ gezeigt werden (vgl. 5.2, S. 75). Dies widerspricht gängigen theoretischen Annahmen, nach denen das Modellverstehen Modellierprozesse lenkt (Louca &

Zacharia, 2012; Schwarz et al., 2009) oder eine Umsetzung von Modellierprozessen elaboriertes Modellverstehen aktiviert (Gobert & Pallant, 2004) und deutet darauf hin, dass Modellverstehen insgesamt kein guter Prädiktor für die Umsetzung von Modellierprozessen ist. In der vorliegenden Untersuchung wurden dabei sowohl Fälle beobachtet, bei denen eine elaborierte Perspektive auf Modelle als epistemische Werkzeuge als Modellverstehen vorliegt, aber nur sehr basale Modellierprozesse umgesetzt werden (z. B. Carlo), als auch Fälle bei denen ein eher medial orientiertes Modellverstehen vorliegt, bei denen dann aber konsistent Tätigkeiten der Vorhersage in den Modellierprozessen umgesetzt werden und das Modell damit klar als epistemisches Werkzeug eingesetzt wird (z. B. Raphael). Entgegen der Annahme von Ke und Schwarz (2020), das Modellverstehen stehe im Zusammenhang mit dem vorliegenden Kontext (vgl. Fragebogen und Verbalisierung im Modellierprozess), konnte auch zwischen dem kontextualisierten Modellverstehen und den Modellierprozessen kein quantitativer Zusammenhang beobachtet werden. Generell ergibt sich für das Modellverstehen innerhalb der Aussagen einzelner ProbandInnen eine hohe Varianz über die fünf Aspekte des Modellverstehen und den Erhebungskontext hinweg, was einen weiteren Hinweis auf die Kontextabhängigkeit von Modellverstehen und Modellierkompetenz allgemein darstellen könnte (Gobert & Pallant, 2004; Krell et al., 2012, 2014a; Sikorski, 2019; Sins et al., 2009).

## 6. Übergeordnete Diskussion

In diesem Abschnitt erfolgt eine zusammenfassende Bewertung der Publikationen im Hinblick auf ihre methodologische Qualität und die in Abschnitt 3 (S. 52) formulierten inhaltlichen Ziele. Hierbei wird vor allem die Generalisierbarkeit der Ergebnisse diskutiert und es werden offene Forschungsfragen, sowie Implikationen für die Lehrkräftebildung abgeleitet.

### 6.1 Methodologische Diskussion

Um im Sinne der weiteren Entwicklung und Etablierung der Naturwissenschaftsdidaktiken als empirisch orientierte Disziplinen (Leuders, 2015) die Qualität der vorliegenden Untersuchung methodenspezifisch zu sichern und transparent zu machen (vgl. „*methodological awareness*“; Seale, 2002) und die vorliegende Untersuchung damit als erprobte Forschungspraxis auffassen zu können (Helsper et al., 2001), wird die vorliegende Untersuchung anhand von Gütekriterien und angewandten qualitätssichernden Maßnahmen diskutiert.

Da bisher keine etablierten allgemeingültigen Gütekriterien oder qualitätssichernden Maßnahmen für die qualitative fachdidaktische Forschung vorliegen (Flick, 2010; Steinke, 2013), zeigt Tabelle 22 auf Basis der in Publikation 1 (Göhner & Krell, 2020) identifizierten Maßnahmen der Qualitätssicherung bei der Umsetzung qualitativ inhaltsanalytischer Verfahren einen Vorschlag zur Zuordnung der in der vorliegenden Untersuchung umgesetzten Maßnahmen der Qualitätssicherung zu den klassischen Gütekriterien Objektivität, Reliabilität und Validität.

Tabelle 22: Vorschlag zur Zuordnung von Maßnahmen der Qualitätssicherung zu den klassischen Gütekriterien Objektivität, Reliabilität und Validität bei der Umsetzung qualitativ inhaltsanalytischer Verfahren

Gütekriterium	Unterkriterium	Maßnahme
Objektivität	Durchführungsobjektivität	Standardisierung der Testsituation
	Auswertungsobjektivität (Intersubjektivität)	Kodierung durch geschulte Personen und Überprüfung der Interrater-Übereinstimmung
Reliabilität	---	Zweimalige Kodierung durch dieselbe Person und Überprüfung der Intrarater-Übereinstimmung
Validität	Konstruktvalidität	Deduktive Kategorienentwicklung
	Inhaltsvalidität	Induktive Kategorienentwicklung heterogene Fallauswahl Expertendiskussionen*
	Augenscheinvalidität	Überprüfung des Kategoriensystems auf Verständlichkeit und logische Konsistenz durch Personen mit unterschiedlicher Expertise**
	Kriteriumsvalidität	Vergleich mit Konstrukten zur theoretischen Fallauswahl

\* Expertendiskussionen fanden regelmäßig im Rahmen der Kolloquien der AG Krüger statt

\*\* Im Rahmen der Expertendiskussionen, auf Konferenzen und bei Schulung der studentischen Hilfskräfte

### 6.1.1 Stichprobe und Forschungskontext

Bei der Stichprobenziehung wurde mit dem qualitativen Stichprobenplan nach dem Prinzip einer bewusst heterogenen Fallauswahl (Kelle & Kluge, 2010; Schreier, 2010) das Ziel verfolgt, eine möglichst große Diversität an Modellierungsstrategien beobachten zu können (vgl. Inhaltsvalidität, Tabelle 22). Durch das Sampling von Biologie-Lehramtsstudierenden als ProbandInnen im Rahmen regelmäßig stattfindender obligatorischer Seminare an der Freien Universität Berlin und Humboldt Universität Berlin konnte der Zugang zum Feld gewährleistet werden. Die Stichprobe deckt damit aber natürlich nur einen kleinen Teil der deutschen Lehrkräftebildung ab.

Da die bewusst heterogene Fallauswahl anhand der als relevant identifizierten Kriterien „Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung“ und „allgemeine kognitive Fähigkeiten“ keinen systematischen Zusammenhang mit den beobachteten Modellierprozessen gezeigt hat, sind weitere Untersuchungen nötig, die Evidenz für die Validität der vorliegenden Ergebnisinterpretation über Berlin hinaus sammeln. Gleichzeitig haben die für die bewusst heterogene Fallauswahl eingesetzten Instrumente (KoWADiS-Fragebogen: Hartmann et al., 2015; I-S-T 2000R: Liepmann et al., 2007) die Stichprobe von 666 teilnehmenden Studierenden auf 57 Studierende reduziert, die alle Anforderungen des Stichprobenplans erfüllten, obwohl eine Verringerung des Grenzwerts für die Selektion der Stichprobe durch das Kriterium „allgemeine kognitive Fähigkeiten“ zu Projektbeginn vorgenommen wurde (von 1SD auf 0.5SD besser oder schlechter als die Normstichprobe). Von diesen 57 Studierenden nahmen 36 an der vorliegenden Untersuchung teil. Die angepeilte Stichprobengröße von mindestens 40 ProbandInnen wurde so insgesamt knapp verfehlt, was durchaus mit den eingesetzten Testinstrumenten zusammenhängen könnte, die von den ProbandInnen, im Gegensatz zur eigentlichen Erhebung, als sehr zeitaufwändig und anstrengend wahrgenommen wurden; besonders bei „schlechten“ Testergebnissen. Eine Auswirkung auf die Motivation der ProbandInnen im Verlauf der weiteren Untersuchung ist daher ebenfalls nicht auszuschließen. Allerdings zeigen sich keine Zusammenhänge zwischen den Samplingkriterien und den beobachteten Modellierprozessen (Publikation 5, Abschnitt 5.5, S. 91). Zukünftige Untersuchungen sollten hier stärker forschungsökonomische Perspektiven – insbesondere auch die der Studierenden – beachten.

In Bezug auf die Stichprobe muss zudem transparent gemacht werden, dass der Autor selbst vor Beginn seiner Mitarbeit im Projekt als Proband teilnahm (Pseudonym *Floyd*), da er zu diesem Zeitpunkt ebenfalls Lehramtsstudent in den Fächern Biologie und Informatik an der Freien Universität Berlin war. Auf der einen Seite erleichterte dies den Umgang mit ProbandInnen und die Auswertung der Daten sehr, da der Autor mit dem Ablauf der Untersuchung vertraut war und diese selbst durchlebt hat. Zum anderen ist ein gewisser Bias bei der Bewertung der eigenen Performanz natürlich nicht auszuschließen, auch wenn die Berechnung von Cohen's Kappa auf eine hohe Interrater-Übereinstimmung hinweist ( $\kappa=0,77$ ; vgl. Auswertungsobjektivität bzw. Intersubjektivität, Tabelle 22). Letztendlich könnte ein genereller

Einbezug der ProbandInnen in die Datenauswertung in zukünftigen Projekten im Sinne einer kommunikativen Validierung (vgl. Mayring, 2002) auch weitere Evidenz für die Validität der Ergebnisinterpretation generieren.

### 6.1.2 Das Blackbox-Setting

Für Modellierkompetenz liegen bereits eine Reihe von Untersuchungen vor, die zeigen, dass die von einer Person in einer Testsituation gezeigte Performanz stark durch Eigenschaften der jeweiligen Aufgabe beeinflusst werden kann („Kontextabhängigkeit“; Krell et al., 2014a; Löffler, 2016). Entsprechend ist es nicht verwunderlich, dass innerhalb der Peer-Review-Verfahren der einzelnen Publikationen wiederholt die Generalisierbarkeit des Blackbox-Settings hinterfragt wurde. Die Blackbox als standardisiertes Setting (vgl. Objektivität; Tabelle 22) wurde in der vorliegenden Studie authentischeren naturwissenschaftlichen Kontexten (z. B. Daten aus echten biologischen Feldstudien) vorgezogen, um die inhaltliche und temporale Komplexität des Problems zu reduzieren. Hierdurch ist für die Untersuchung der Blackbox kein ausgeprägtes (naturwissenschaftliches<sup>14</sup> und insbesondere biologisches) Vorwissen notwendig und es kann auf den Problemlöseprozess fokussiert werden („könnenszentriertes Problemlösen“; Friege & Lind, 2003). Die Generalisierbarkeit der vorliegenden Ergebnisse, einschließlich der erarbeiteten Typologie, auf Modellierprozesse in anderen Kontexten kann daher limitiert sein, da domänenspezifisches Wissen als Voraussetzung für die erfolgreiche Umsetzung von Modellierprozessen angesehen wird (Ruppert et al., 2019). Gleichzeitig wird die situationsspezifische Natur von Kompetenzen und deren Erfassung betont (Blömeke, König et al., 2015; Blömeke et al., 2013). Weitere Evidenz für eine möglicherweise limitierte Generalisierbarkeit der Ergebnisse stellen einige der in Publikation 4 identifizierten Hindernisse dar, die ebenfalls dem konkreten Blackbox-Setting zugeordnet werden können (Hindernisse 5, 6, 8, 11). Hieraus ergibt sich die folgende Forschungslücke, die im Fazit (Abschnitt 7, S. 112) zudem weiter ausgeführt wird.



Inwieweit können die identifizierten Modellierungsstrategien in anderen (realistisch biologischen) Settings repliziert werden?

Forschungslücke 2: Inwieweit sind die identifizierten Modellierungsstrategien übertragbar?

---

<sup>14</sup> Für die Modellierung der hier verwendeten Blackbox kann physikalisches Wissen hilfreich sein (z. B. aus dem Bereich Hydraulik), allerdings wurde die Anwendung relevanten physikalischen Vorwissens bei den teilnehmenden ProbandInnen selten beobachtet; Ausnahmen sind ggf. James und Jonathan.

Zuletzt begünstigt das Blackbox-Setting als eher offene und abstrakte Aufgabenstellung (vgl. Leden et al., 2020) möglicherweise den hohen Anteil explorativer Tätigkeiten innerhalb der Modellierprozesse über alle ProbandInnen hinweg.

### 6.1.3 Modellieren durch Zeichnen und Lautes Denken

Weitere mögliche Limitationen der Validität der Ergebnisinterpretation der vorliegenden Untersuchung betreffen den eigentlichen Modellierprozess und das Modellierprodukt. Den ProbandInnen war es in der vorliegenden Untersuchung nur möglich Modelle zeichnend zu externalisieren. Dies entspricht zwar einem weit verbreiteten Vorgehen in der naturwissenschaftsdidaktischen Forschung (Chang et al., 2020), es kann allerdings nicht ausgeschlossen werden, dass das Zeichnen die Modellierprozesse der ProbandInnen eingeschränkt hat. Allerdings kommt Orsenne (2016) in einem Vergleich zwischen zeichnerischen und gegenständlichen (d. h. materialgestützten) Modellentwicklungen zu dem Schluss, dass eine zeichnerische Modellentwicklung kreative Analogiebildungen unterstützt, während eine gegenständliche Modellentwicklung die Kreativität von ProbandInnen durch eine Fokussierung auf vorgegebene Materialien einschränkt. Trotzdem ergeben sich Hinweise auf eine mögliche Limitation der ProbandInnen durch das Zeichnen. Die Modellierprozesse der zwei ProbandInnen sind detailliert dargestellt in Publikation 4 (Göhner & Krell, 2021). Während ein Proband kein Modell zeichnet, verwirft eine andere Probandin ihr – aus ihrer Sicht falsches – Modellierprodukt vor Beendigung ihres Modellierprozesses.

Studien aus der Repräsentationsforschung bieten hier zwei mögliche Erklärungen: ProbandInnen sind keine „VisualisiererInnen“ (Koč-Januchta et al., 2017) und Zeichnen erscheint ihnen nicht als hilfreich zur Lösung der Aufgabe. Zum anderen könnte das Zeichnen von ProbandInnen aktiv vermieden werden, um kein „fehlerhaftes“ Modell als Modellierprodukt vorweisen zu müssen. Hier wird besonders sichtbar, dass volitionale Aspekte (z. B. Selbstregulation, Selbstwirksamkeitserwartung; Bandura, 1994) bei Modellierprozessen von zentraler Wichtigkeit sind (Hogan & Thomas, 2001; Nehring et al., 2015; Nordheimer, 2019). Obwohl sie Teil gängiger Kompetenzdefinitionen sind (Blömeke et al., 2013; Weinert, 2001a), werden motivationale und volitionale Aspekte jedoch in der Kompetenzforschung oft explizit exkludiert (Klieme et al., 2007). Hieraus ergibt sich die folgende Forschungslücke, die demnächst in Ammoneit et al. (Manuscript submitted for publication) adressiert wird.



Welchen Einfluss haben motivationale und volitionale Aspekte auf Modellierprozesse von Lehramtsstudierenden.

Forschungslücke 3: Welchen Einfluss haben Motivation und Volition auf Modellierkompetenz?

Das Laute Denken liefert weitere Evidenz für die Validität der Ergebnisinterpretation der vorliegenden Untersuchung, indem es einen Einblick in die gedankliche Modellbildung gibt. Allerdings kann hier ebenfalls nicht sichergestellt werden, dass alle Denkprozesse der ProbandInnen im Rahmen des Lauten Denkens geäußert wurden, da Lautes Denken eine hohe kognitive Zusatzbelastung mit sich bringt (Konrad, 2010; Sandmann, 2014). Einige ProbandInnen weisen entsprechend in ihren Modellierprozessen längere Schweigephasen auf.

### 6.1.4 Datenauswertung

Um die Qualität der Datenauswertung sicherzustellen, wurden in der vorliegenden Untersuchung und der vorangegangenen Pilotierung (Krell et al., 2019) wie in Tabelle 22 dargestellt viele verschiedene Maßnahmen der Qualitätssicherung umgesetzt. Insbesondere die mehrfache Kodierung durch den Autor, sowie die Kodierung durch eine weitere geschulte Person haben den Prozess der Datenauswertung dabei maßgeblich unterstützt, wobei die hohen Übereinstimmungen als wichtige Evidenz für die Validität (und Objektivität) der Ergebnisinterpretation aufgefasst werden können. Einige Limitationen ergeben sich allerdings trotzdem in Bezug auf die Auswertung der Daten:

Hinsichtlich der qualitativen Datenauswertung ergeben sich Fragen der Validität und Generalisierbarkeit im Hinblick auf das Modellverstehen und die Modellierprodukte. Das Modellverstehen wurde wie in Abschnitt 2.4.1 (S. 42) bereits beschrieben als Modellverstehen (Wissen über Modelle) erfasst, da bislang keine Operationalisierung für Modellierverstehen existieren. Hieraus ergibt sich die folgende Forschungslücke:



Wie kann Modellierverstehen, im Sinne des Wissens über den Modellierprozess, erfasst werden?

Forschungslücke 4: Wie kann Modellierverstehen erfasst werden?

Das beobachtete Modellverstehen während des Modellierprozesses (Ke & Schwarz, 2020) kann hier eventuell eine Annäherung darstellen; durch das Fehlen einer expliziten Aufforderung solches Modellverstehen *in-action* zu äußern, sind die Datensätze hier allerdings unvollständig und ihre Generalisierbarkeit erscheint limitiert. Da zusätzlich das erweiterte Kompetenzmodell der Modellierkompetenz (Krüger & Upmeyer zu Belzen, 2021) eine Zuordnung des abduktiven Schließens zu Niveau III der Modellierkompetenz vornimmt, wenn anhand von Modellen retrodiktiv Erklärungen über das Phänomen abgeleitet werden, könnte eine sorgfältige Reanalyse aller Aussagen zum Modellverstehen weitere Erkenntnisse liefern. Aus forschungsökonomischen Gründen wurde diese Analyse jedoch nicht in der vorliegenden Untersuchung umgesetzt. Hinsichtlich der Modellierprodukte war eine Auswertung

## Übergeordnete Diskussion

anhand epistemischer Kriterien (Pluta et al., 2011) durch die geringe Komplexität und Varianz der Modellierprodukte ebenfalls nicht möglich.

Die Ergebnisse der quantitativen Analysen in den Publikationen 5 und 6 sollten mit weiteren Untersuchungen überprüft werden, da die Stichprobengröße mit 36 ProbandInnen für quantitative Analysen insgesamt recht gering ausfällt. Für einige Subgruppen konnten nicht ausreichend ( $n = 10$ ) Studierende als ProbandInnen gewonnen werden konnten (Tabelle 4, S. 63) und insbesondere in den Subgruppen mit entgegengesetzten Ausprägungen der Selektionskriterien war die Gewinnung von ausreichend geeigneten ProbandInnen nicht möglich.

### 6.2 Inhaltliche Diskussion

Primäres Ziel der vorliegenden Untersuchung war die fallbasierte und prozessorientierte empirische Beschreibung individueller Modellierungsstrategien, um daraus eine Typologie von Modellierungsstrategien Lehramtsstudierender abzuleiten. Hierbei war die Frage „Welche Modellierungsstrategien zeigen Lehramtsstudierende während der Problemlösung an einer Blackbox?“ leitend, die im Rahmen der Publikationen 2 (Göhner & Krell, 2018) auf individueller Prozessebene und im Rahmen der Publikation 3 (Göhner & Krell, 2022b) auf übergeordneter Ebene einer Typologie beantwortet werden konnte. Zudem konnten Hindernisse beim Modellieren (Publikation 4; Göhner & Krell, 2021) identifiziert werden.

Auf Ebene der individuellen Modellierprozesse und Ebene der Typologie von Modellierungsstrategien ergeben sich einige Bezüge zu vorangegangenen Untersuchungen, die die Generalisierbarkeit der Ergebnisse aus Publikation 2 und 3 unterstützen.

#### 6.2.1 Individuelle Modellierprozesse

Durch die gewählte Operationalisierung anhand des Prozessschemas naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren (Göhner & Krell, 2018, adaptiert nach Krell et al., 2016; Abbildung 6) konnten individuelle Modellierprozesse in einem hohen und gleichzeitig variablen Detailgrad abgebildet werden. Insbesondere die Tätigkeit der Zusammenfassung der beobachteten Daten (Tätigkeit 3) über die Modellierprozesse der ProbandInnen hinweg stellte eine zentrale Tätigkeit dar. Dies deckt sich mit den Beobachtungen von Sins et al. (2005), die bei modellierenden SchülerInnengruppen ähnliches Verhalten feststellen konnten. Auch die Aktivierung von Analogien und Vorerfahrungen (Tätigkeit 6) nimmt eine zentrale Position in den individuellen Modellierprozessen ein, da sie über den gesamten Modellierprozess hinweg zu jeder Zeit auftreten kann (vgl. Clement, 2008).

#### 6.2.2 Typologie von Modellierungsstrategien

Durch die in Abschnitt 4.3 (S. 59) und Publikation 3 (S. 79) dargestellte Methode nach dem Modell empirisch begründeter Typenbildung konnten sechs verschiedene Typen von Modellierungsstrategien erfolgreich identifiziert werden, die als spezifisch für das Modellieren angesehen werden können. Auf dieser Ebene der Typologie von Modellierungsstrategien zeigen Vergleiche Parallelen mit vorherigen Arbeiten zu Strategien der naturwissenschaftlichen Erkenntnisgewinnung, sowie insbesondere dem Experimentieren auf. ProbandInnen, die Modellierprozesse vom Typ 2b umsetzen, die also Tätigkeiten der Exploration und der Herstellung umsetzen und sich dabei auf die Exploration fokussieren, gehen ähnlich vor, wie ProbandInnen in der Studie von Schauble, Glaser et al. (1991) die dem Ingenieursmodus zugeordnet werden können. Ähnlich zu den Beobachtungen von Schauble, Glaser et al. (1991) untersuchen die ProbandInnen das natürliche Phänomen der Blackbox, bis sich ein spezifischer Effekt einstellt (z. B. ein wiederkehrendes Muster von Outputs) und bearbeiten dann die vorgegebene Aufgabe, indem sie ein

Modell des Inneren der Blackbox zeichnen und aufhören. Einige ProbandInnen setzten dagegen vermehrt Tätigkeiten der Herstellung um (Typ 2b/c), was dem von Sins et al. (2009) beschriebenen *model-fitting* nahe kommt. Obwohl andere empirische Studien zu verschiedenen Populationen darauf hindeuten, dass epistemische Perspektiven auf Modelle und das Modellieren, die die Vorhersage umfassen, kaum eine Rolle spielen (Krell & Krüger, 2016; Passmore et al., 2014) wurden in der Hälfte der beobachteten Modellierprozesse Tätigkeiten der Vorhersage umgesetzt (Typen 3a/b). Diese Modellierprozesse, die als eher komplex und dynamisch beschrieben werden können (Gilbert & Justi, 2016), entsprechen jedoch nicht dem streng linear-zyklischen Vorgehen, das in der Literatur oftmals beschrieben wird (z. B. *Generation-Evaluation-Modification* Kreislauf, Khan, 2011; Louca & Zacharia, 2015). Vielmehr setzen nur wenige ProbandInnen mehrmals Tätigkeiten der Vorhersage in ihrem Modellierprozess um (Typ 3b) und selbst diese ProbandInnen beenden ihren Modellierprozess, wenn sie ein Modellierprodukt erarbeitet haben, das sie für eine mögliche Lösung der Aufgabe halten. Es finden keine Versuche statt das Modellierprodukt aktiv zu falsifizieren, was von manchen AutorInnen als besonders relevant hervorgehoben wird (Grosslight et al., 1991; Louca & Zacharia, 2015).

### 6.2.3 Dimensionalität und Extensionalität von Modellierkompetenz

Als sekundäres Ziel der Untersuchung wurden Erkenntnisse über die Dimensionalität und Extensionalität der Modellierkompetenz angestrebt. Um die Dimensionalität der Modellierkompetenz näher zu beleuchten, wurde der Zusammenhang zwischen den beobachteten Modellierprozessen und den weiteren Dimensionen der Modellierkompetenz, Modellverstehen und Modellierprodukt, untersucht (Publikationen 6; Göhner et al., 2022). Um die Extensionalität der Modellierkompetenz näher zu beleuchten, wurden Zusammenhänge zwischen den beobachteten Modellierprozessen und den erfassten Personenmerkmalen, insbesondere den Variablen der heterogenen Fallauswahl, untersucht (Publikation 5, Göhner & Krell, 2022a).

#### **Dimensionalität**

Im Hinblick auf die Dimensionalität der Modellierkompetenz konnte in der vorliegenden Untersuchung, entgegen der in der naturwissenschaftsdidaktischen Forschung weit verbreiteten Annahme, ein ausgeprägtes Modellverstehen unterstütze die Umsetzung von Modellierprozessen (Cheng & Lin, 2015; Louca & Zacharia, 2012; Schwarz et al., 2009), kein signifikanter quantitativer Zusammenhang zwischen dem Modellverstehen und den beobachteten Modellierprozessen der ProbandInnen gefunden werden. Allerdings zeigen einige ProbandInnen (vgl. *Angelina*; Publikation 2) durchaus Modellverstehen (insbesondere in den Aspekten „Testen von Modellen“ und „Ändern von Modellen“), das zu den durchgeführten Modellierprozessen passt. Allerdings ist weiterhin unklar, ob Modellverstehen die Modellierprozesse anleitet oder der durchgeführte Modellierprozess das Modellverstehen fördert (Gobert & Pallant, 2004; Schwarz et al., 2009). Zwischen Modellierprozess und Modellierprodukt wurde zwar ein

signifikanter quantitativer Zusammenhang mit schwacher Effektstärke gefunden, insgesamt betonen die Ergebnisse der vorliegenden Untersuchung aber, dass für eine möglichst valide Diagnose von Modellierkompetenz alle drei Dimensionen berücksichtigt werden sollten. Modellverstehen als alleiniger Indikator für Modellierkompetenz erscheint wenig valide. Weitere Forschung ist außerdem nötig, um den Zusammenhang zwischen Modellierverstehen (als Wissen über den Modellierprozess) und dem tatsächlichen Modellierprozess zu untersuchen (vgl. Forschungslücke 4).

### **Extensionalität**

Im Hinblick auf die Extensionalität der Modellierkompetenz identifizieren theoretische Arbeiten und andere empirische Studien viele Personenmerkmale, die möglicherweise mit dem Modellieren zusammenhängen. So wurde erwartet, dass allgemeine kognitive Fähigkeiten (Nehring et al., 2015), fachliches bzw. domänenspezifisches Vorwissen (Ruppert et al., 2017, 2019), Kreativität (Mierdel & Bogner, 2019) sowie Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung (Krell et al., 2014b) die Umsetzung von Modellierprozessen unterstützen. Zudem wurden deskriptive Personenmerkmale wie das Alter der ProbandInnen oder die Studienfachkombination untersucht. Wie in Publikation 5 beschrieben, wurden keine nennenswerten signifikanten quantitativen Zusammenhänge zwischen „allgemeine kognitive Fähigkeiten“, „Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung“ und den beobachteten Modellierprozessen gefunden. Dagegen wurde ein potenziell hoher Einfluss volitionaler Aspekte beobachtet (Ammonet et al., Manuscript submitted for publication; Nordheimer, 2019), der weiterer Forschung bedarf. Wie bei der Dimensionalität der Modellierkompetenz konnten vorliegende theoretische Annahmen über Personenmerkmale mit potenziell großem Einfluss auf den Modellierprozess (allgemeine kognitive Fähigkeiten, räumliches Vorstellungsvermögen, Kreativität; Bailer-Jones, 1999; Clement, 2008; Nersessian, 2002) hier nicht empirisch bestätigt werden.

### **Modellierkompetenz als Kontinuum**

Die Ergebnisse zur Dimensionalität und Extensionalität der Modellierkompetenz machen deutlich, dass eine valide Erfassung und trennscharfe Abgrenzung von Modellierkompetenz schwierig ist. Viele Teilergebnisse (z. B. setting-spezifische Hindernisse; Publikation 4) betonen, dass Kompetenz im Allgemeinen (Blömeke et al., 2013; Rychen & Salganik, 2003) und damit auch Modellierkompetenz im Speziellen kontextabhängig sind. Generell ist nicht davon auszugehen, dass die ProbandInnen über mehrere Problemlöseaufgaben hinweg konsistente Modellierungsstrategien zeigen (vgl. inkonsistentes Modellverstehen von Lehrkräften; Justi & Gilbert, 2003). Folgt man dem Verständnis von Kompetenzen als (situationsspezifisches) Kontinuum (Blömeke, Gustafsson & Shavelson, 2015; Blömeke, König et al., 2015; Blömeke et al., 2013), so erfordert die tatsächliche Diagnose von Kompetenz die Erfassung von Performanzen einer Person über viele Problemlöseaufgaben hinweg, die unterschiedlich kontextualisiert sind. Explanative Studien, die diese Kontextabhängigkeit systematisch untersuchen, liegen für

Modellierkompetenz bislang kaum vor. Darüber hinaus sind viele Studien zur Kontextabhängigkeit von Modellierkompetenz quantitativ ausgerichtet und produktorientiert; prozessorientierte Studien, in denen die Interaktion zwischen Personen und Eigenschaften einer vorliegenden Testsituation auch qualitativ untersucht werden, sind ebenso selten.



Inwiefern unterscheiden sich die Modellierprozesse von Lehramtstudierenden in Abhängigkeit von situationsspezifischen Merkmalen?

Forschungslücke 5: Inwiefern sind Modellierprozesse situationsspezifisch?

Im Hinblick auf Modellierkompetenz könnten bspw. der Einfluss epistemischer Überlegungen und spezifischer Aufgabenmerkmale auf das Modellieren untersucht werden.

### 6.2.4 Modellieren in der schulischen Praxis

Durch den Fokus der vorliegenden Untersuchung auf die Modellierkompetenz von angehenden Biologie-Lehrkräften selbst (*content knowledge*; Baumert & Kunter, 2013), tragen die vorliegenden Ergebnisse wenig dazu bei, Erkenntnisse über die Kompetenz von Lehrkräften im Hinblick auf ihren Einsatz von Modellen und dem Modellieren in der schulischen Praxis (*pedagogical content knowledge*) zu gewinnen. Es kann vermutet werden, dass SchülerInnen und Lehrkräfte in individuellen Modellierprozessen prinzipiell ähnlichen Herausforderungen begegnen, allerdings stellt die Integration von Modellen und dem Modellieren im Unterricht potenziell noch eine weitaus schwierigere Herausforderung für Lehrkräfte dar (Shi et al., 2021). Lehrkräfte müssen hier zusätzlich in der Lage sein, die Vorstellungen und Fähigkeiten von SchülerInnen zu Modellen und dem Modellieren im Unterricht zu diagnostizieren, sowie entsprechende Lerngelegenheiten planen und umsetzen können (Elby & Hammer, 2010; Sung & Oh, 2018). Da diese Kompetenzen oft über lange Zeiträume erworben werden (Vo et al., 2019), sind hier Längsschnittstudien nötig, die zeigen, wie die Modellierkompetenz von Lehrkräften, die Planung und Umsetzung von Lernangeboten zu Modellen und dem Modellieren und die Modellkompetenz der teilnehmenden SchülerInnen sich über größere Zeiträume hinweg verändern (vgl. Blömeke et al., 2022).



Wie verändern sich die Modellierkompetenz von Lehrkräften, ihr Umgang mit Modellen und dem Modellieren in schulischer Praxis und die Modellkompetenz ihrer SchülerInnen über längere Zeiträume hinweg?

Forschungslücke 6: Wie verändern sich Modellierkompetenz und schulische Praxis über die Zeit?

## 7. Fazit

Die Erfassung und Förderung der komplexen, naturwissenschaftlichen Denkprozesse wird von Osborne (2013) als *die* Herausforderung der naturwissenschaftsdidaktischen Forschung für das 21. Jahrhundert proklamiert. Die vorliegende Arbeit nahm sich dieser Herausforderung an, indem sie das Modellieren als eine zentrale Denk- und Arbeitsweise der Naturwissenschaften untersucht hat (Giere et al., 2006; Halloun, 2004). Im Folgenden werden die zentralen Implikationen der vorliegenden Arbeit für die Forschung und Lehrkräftebildung kurz zusammengefasst.

### 7.1 Implikationen für die Forschung

Zusammenfassend wird deutlich, dass die Erfassung von Modellierkompetenz weiterhin eine Herausforderung bleibt (vgl. Osborne, 2013), die eine umfassende, holistische Perspektive auf den Kompetenzbegriff erfordert. Die vorliegende Untersuchung hat hierbei eine Typologie erarbeitet, die dabei helfen kann, eine Dimension der Modellierkompetenz, den Modellierprozess, zu erfassen. Die vorliegende Typologie erweitert das von Krell et al. (2019) erarbeitete Kategoriensystem und ermöglicht eine schnelle, potenziell automatisierte Zuordnung von individuellen Modellierprozessen zu einer der Modellierungsstrategien, was die Komplexität zukünftiger Analysen reduziert und Unterschiede zwischen individuellen Modellierprozessen hervorheben kann. Die Typologie kann damit potenziell weitere Forschungsarbeiten anleiten und die Theoriebildung fördern (vgl. Klahr & Dunbar, 1988), womit sie einen Beitrag zur Erweiterung des naturwissenschaftsdidaktischen Kenntnisstands darstellt (Louca & Zacharia, 2012).

Darüber hinaus können die in der vorliegenden Untersuchung identifizierten Hindernisse im Rahmen der Kompetenzmodellierung und -erfassung langfristig zur Unterscheidung von Problemtypen des Modellierens führen (Gut et al., 2014; Pitt Hild, 2020) und so die Entwicklung weiterer Messinstrumente oder Forschungsdesigns anleiten.

Die im Rahmen der vorliegenden Untersuchung erarbeiteten methodologischen Ergebnisse zu Gütekriterien und Maßnahmen der Qualitätssicherung bei qualitativ inhaltsanalytischen Verfahren tragen zudem zur Qualität und methodischen Transparenz zukünftiger Forschungsarbeiten bei und unterstützen so die Professionalisierung der Naturwissenschaftsdidaktik als Fachbereich (Leuders, 2015), wobei die vorliegende Untersuchung als erprobte Forschungspraxis aufgefasst werden kann (Helsper et al., 2001).

Die im vorangegangenen Abschnitt formulierten Forschungslücken können zudem in zukünftigen Forschungsprojekten als zentrale Fragestellungen aufgegriffen werden.

## 7.2 Implikationen für die Lehrkräftebildung

Von den 36 teilnehmenden angehenden Biologielehrkräften konnten lediglich sechs der elaboriertesten Modellierungsstrategie zugeordnet werden. Diese zeichnet sich dadurch aus, dass – entsprechend dem in der Literatur beschriebenen Vorgehen beim naturwissenschaftlichen Modellieren – wiederholt Modelle entwickelt und als Werkzeuge zur Lösung der Aufgabe genutzt werden. Die Ergebnisse der vorliegenden Untersuchung legen damit nahe, dass im Zuge der Biologie-Lehrkräftebildung die Fähigkeiten zur Umsetzung naturwissenschaftlicher Modellierprozesse stärker gefördert werden sollten.

Das entwickelte Instrumentarium bestehend aus den Kategoriensystemen zu Tätigkeiten des Modellierens und Hindernissen in Modellierprozessen, sowie der entwickelten Typologie können zukünftig im Rahmen von Programmen zur Förderung der Modellkompetenz von (angehenden) Lehrkräften genutzt werden (Cheng et al., 2017; Gräsel & Trempler, 2017; Tempel, 2017). Zum einen können Vermittlungssituationen identifiziert werden, in denen das individuelle Vorgehen (angehender) Lehrkräfte potenziell entwickelt werden kann, insbesondere in Bezug auf Tätigkeiten der Vorhersage, indem typenspezifische Reflexionsimpulse und Interventionsmaßnahmen zur Förderung der Modellkompetenz von (angehenden) Biologie-Lehrkräften entwickelt werden (vgl. Günther et al., 2019). Gleichzeitig können auch die identifizierten Hindernisse genutzt werden, um angemessene Hilfestellungen und Impulse zu planen, die Lehrkräften bei Hindernissen im Modellierprozess spontan präsentiert werden, sodass eine nachhaltige Behinderung des Modellierprozesses vermieden wird (Baur, 2018; Sung & Oh, 2018) oder Teilziele bei der Problemlösung vorzugeben (Sins, Savelsbergh, & van Joolingen, 2005). Die in dieser Untersuchung erfolgte holistische Charakterisierung einzelner Fälle, kann zudem genutzt werden, um das Vorgehen der Lehramtsstudierenden (als NovizInnen) mit dem von ExpertenInnen zu kontrastieren. Somit kann ein Schwerpunkt auf die Vermittlung von denjenigen (kognitiven) Prozessen gelegt werden, die bei ExpertenInnen beobachtet, von den Lehramtsstudierenden aber nicht oder nur in Ansätzen umgesetzt werden (Clement & Williams, 2013). Die vorliegende Untersuchung kann somit dazu beitragen, die von der KMK (2015) definierten inhaltlichen Anforderungen für die Lehrerbildung zu erreichen (Modellkompetenz als Bestandteil des *content knowledge*; vgl. Günther et al., 2019; Oh & Oh, 2011). Lehrkräfte werden dazu befähigt, im eigenen Unterricht naturwissenschaftliche Modellierprozesse zu initiieren, sowie typische Vorgehensweisen beim Modellieren erkennen und gezielt unterstützen zu können (*pedagogical content knowledge*; vgl. Günther et al., 2019; Henze et al., 2007, 2008; Justi & van Driel, 2005), indem die angehenden Lehrkräfte lernen ihre eigenen Perspektiven auf Modelle zu reflektieren und ihre eigenen Fähigkeiten im Umgang mit Modellen und dem Modellieren einschätzen zu können (Harlow et al., 2013; Vo et al., 2015, 2019).

## 8. Literaturverzeichnis

- Abd-El-Khalick, F. (2002). Rutherford's enlarged: A content-embedded activity to teach about nature of science. *Physics Education*, 37(1), 64.
- ACARA. (2015). *The Australian Curriculum: Science*. Australian Curriculum, Assessment and Reporting Authority. <http://www.australiancurriculum.edu.au/>
- Acher, A., Arcá, M. & Sanmartí, N. (2007). Modeling as a teaching learning process for understanding materials: A case study in primary education. *Science Education*, 91(3), 398–418.
- Ainsworth, S., Prain, V. & Tytler, R. (2011). Drawing to learn in science. *Science*, 333(6046), 1096–1097.
- Aizpurua, A., Lizaso, I. & Iturbe, I. (2018). Learning strategies and reasoning skills of university students. *Revista de Psicodidáctica (English ed.)*, 23(2), 110–116. <https://doi.org/10.1016/j.psicoe.2018.02.002>
- Al-Balushi, S. M. (2009). Factors influencing pre-service science teachers' imagination at the microscopic level in chemistry. *International Journal of Science and Mathematics Education*, 7(6), 1089–1110. <https://doi.org/10.1007/s10763-009-9155-1>
- Al-Balushi, S. M. (2011). Students' evaluation of the credibility of scientific models that represent natural entities and phenomena. *International Journal of Science and Mathematics Education*, 9, 571–601.
- Ammoneit, R., Göhner, M. F., Bielik, T. & Krell, M. (Manuscript submitted for publication). Why most Definitions of Modeling Competence in Science Education fall short: Analyzing the Relevance of Volition for Modeling. *Science Education*.
- Andrienko, N. & Andrienko, G. (2018). State transition graphs for semantic analysis of movement behaviours. *Information Visualization*, 17(1), 41–65. <https://doi.org/10.1177/1473871617692841>
- Bailer-Jones, D. (1999). Creative Strategies Employed in Modelling: A Case Study. *Foundations of Science*, 4, 375–388.
- Bailer-Jones, D. (2002). Scientists' thoughts on scientific models. *Perspectives on Science*, 10(3), 275–301. <https://doi.org/10.1162/106361402321899069>
- Bailer-Jones, D. (2003). When scientific models represent. *International Studies in the Philosophy of Science*, 17(1), 59–74.
- Bailey, K. (1994). *Typologies and taxonomies: An introduction to classification techniques*. Sage.
- Bakeman, R. & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press.

- Bamberger, Y. M. & Davis, E. A. (2013). Middle-school science students' scientific modelling performances across content areas and within a learning progression. *International Journal of Science Education*, 35(2), 213–238. <https://doi.org/10.1080/09500693.2011.624133>
- Bandura, A. (1994). Self-Efficacy. In V. S. Ramachaudran (Hrsg.), *Encyclopedia of human behavior* (S. 71–81). Academic Press.
- Bastian, M., Heymann, S. & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 8(2009), 361–362.
- Baumert, J. & Kunter, M. (2011). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project* (S. 25–48). Springer US. [https://doi.org/10.1007/978-1-4614-5149-5\\_2](https://doi.org/10.1007/978-1-4614-5149-5_2)
- Baumfalk, B., Bhattacharya, D., Vo, T., Forbes, C. T., Zangori, L. & Schwarz, C. V. (2018). Impact of model-based science curriculum and instruction on elementary students' explanations for the hydrosphere. *Journal of Research in Science Teaching*. Vorab-Onlinepublikation. <https://doi.org/10.1002/tea.21514>
- Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 24, 115–129. <https://doi.org/10.1007/s40573-018-0078-7>
- Baur, A. & Emden, M. (2020). How to open inquiry teaching? An alternative teaching scaffold to foster students' inquiry skills. *Chemistry Teacher International*. Vorab-Onlinepublikation. <https://doi.org/10.1515/cti-2019-0013>
- BCMOE. (2019). *Curriculum redesign*. British Columbia Ministry of Education. <https://curriculum.gov.bc.ca/rethinking-curriculum>
- Bennett, Judith, Lubben, F., Hogarth, S. & Campbell, B. (2005). Systematic reviews of research in science education: rigour or rigidity? *International Journal of Science Education*, 27(4), 387–406. <https://doi.org/10.1080/0950069042000323719>
- Bielik, T., Opitz, S. & Novak, A. (2018). Supporting students in building and using models: Development on the quality and complexity dimensions. *Education Sciences*, 8(3), 149.
- Billingsley, B. & Fraser, S. (2018). Towards an Understanding of Epistemic Insight: the Nature of Science in Real World Contexts and a Multidisciplinary Arena [Editorial]. *Research in Science Education*, 48, 1107–1113. <https://doi.org/10.1007/s11165-018-9776-x>

- Blech, C. & Funke, J. (2013). Zur Reaktivität von Kausaldiagramm-Analysen beim komplexen Problemlösen. *Zeitschrift für Psychologie*, 214(4), 185–195. <https://doi.org/10.11588/heidok.00015565>
- 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>
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G. & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Blömeke, S., König, J., Suhl, U., Hoth, J. & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift für Pädagogik*, 61(3), 310–327.
- Blömeke, S., Zlatkin-Troitschanskaia, O., Kuhn, C. & Fege, J. (2013). Modeling and measuring competencies in higher education. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Hrsg.), *Modeling and measuring competencies in higher education* (S. 1–10). Springer.
- BMBF. (2007). *Zur Entwicklung nationaler Bildungsstandards*.
- Böschl, F., Forbes, C. T. & Lange-Schubert, K. (2023). Investigating scientific modeling practices in U.S. and German elementary science classrooms: A comparative, cross-national video study. *Science Education*(107), 368–400.
- Böttcher, F. & Meister, A. (2011). Argumentation in science education: A model-based framework. *Science & Education*(20), 103–140. <https://doi.org/10.1007/s11191-010-9304-5>
- Box, G. E. P. & Draper, N. R. (1987). *Empirical model-building and response surfaces*. *Wiley series in probability and mathematical statistics*. John Wiley & Sons.
- Brand, A., Allen, L., Altman, M., Hlava, M. & Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, 28(2), 151–155. <https://doi.org/10.1087/20150211>
- Broman, K., Bernholt, S. & Parchmann, I. (2018). Using model-based scaffolds to support students solving context-based chemistry problems. *International Journal of Science Education*, 1–22. <https://doi.org/10.1080/09500693.2018.1470350>
- Brown, D. & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional Science*, 18(4), 237–261. <https://doi.org/10.1007/bf00118013>

- C. J. Suckling, K. E. Suckling & C. W. Suckling. (1978). *Chemistry through models: Concepts and applications of modelling in chemical science, technology and industry*. Cambridge University Press.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2, 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Campbell, T. & Neilson, D. (2012). Modeling electricity: Model-based inquiry with demonstrations and investigations. *The Physics Teacher*, 50(6), 347–350. <https://doi.org/10.1119/1.4745686>
- Campbell, T., Oh, P. S., Maughn, M., Kiriazis, N. & Zuwallack, R. (2015). A review of modeling pedagogies: Pedagogical functions, discursive acts, and technology in modeling instruction. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(1), 159–176. <https://doi.org/10.12973/eurasia.2015.1314a>
- Campbell, T., Oh, P. S. & Neilson, D. (2013). Reification of five types of modeling pedagogies with model-based inquiry (MBI) modules for high school science classrooms. In Information Resources Management Association (Hrsg.), *K-12 Education: Concepts, methodologies, tools, and applications: Development and design methodologies* (Bd. 2, S. 401–421). IGI Global. <https://doi.org/10.4018/978-1-4666-4502-8>
- Capps, D. K. & Shemwell, J. T. (2020). Moving beyond the model as a copy problem: investigating the utility of teaching about structure-preserving transformations in the model-referent relationship. *International Journal of Science Education*, 42(12), 2008–2031. <https://doi.org/10.1080/09500693.2020.1810354>
- Chang, H.-Y. (201 ). Students' representational competence with drawing technology across two domains of science. *Science Education*, 1–21. <https://doi.org/10.1002/sc.21457>
- Chang, H.-Y., Lin, T.-J., Lee, M.-H., Lee, S. W.-Y., Lin, T.-C., Tan, A.-L. & Tsai, C.-C. (2020). A systematic review of trends and findings in research employing drawing assessment in science education. *Studies in Science Education*, 56(1), 77–110. <https://doi.org/10.1080/03057267.2020.1735822>
- Cheng, M.-F. & Lin, J.-L. (2015). Investigating the relationship between students' views of scientific models and their development of models. *International Journal of Science Education*, 37(15), 2453–2475. <https://doi.org/10.1080/09500693.2015.1082671>

- Cheng, M.-F., Lin, J.-L., Lin, S.-Y. & Cheng, C.-H. (2017). Scaffolding Middle High School Students' Modeling Processes. *Journal of Baltic Science Education*, 16(2), 207–217. <http://oaji.net/articles/2017/987-1497156343.pdf>
- Cheng, M.-F., Wu, T.-Y. & Lin, S.-F. (2021). Investigating the relationship between views of scientific models and modeling practice. *Research in Science Education*, 51, 307–323. <https://doi.org/10.1007/s11165-019-09880-2>
- Chinn, C. A. & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. <https://doi.org/10.1002/sce.10001>
- Chiu, M.-H. & Lin, J.-W. (201 ). Modelling competence in science education. *Disciplinary and Interdisciplinary Science Education Research*, 1, Artikel 12. <https://doi.org/10.1186/s43031-019-0012-y>
- Clement, J. (1989). Learning via model construction and criticism. In J. A. Glover, R. R. Ronning & C. R. Reynolds (Hrsg.), *Handbook of Creativity* (S. 341–381). Springer US. [https://doi.org/10.1007/978-1-4757-5356-1\\_20](https://doi.org/10.1007/978-1-4757-5356-1_20)
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9), 1041–1053.
- Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Springer Science & Business Media.
- Clement, J. (2009). The role of imagistic simulation in scientific thought experiments. *Topics in cognitive science*, 1(4), 686–710. <https://doi.org/10.1111/j.1756-8765.2009.01031.x>
- Clement, J. & Rea-Ramirez, M. A. (Hrsg.). (2008). *Model based learning and instruction in science*. Springer.
- Clement, J. & Williams, G. (2013). Parallel roles for nonformal reasoning in expert scientific model construction and classroom discussions in science. In J. Clement & E. G. Williams (Vorsitz), *National Association for Research in Science Teaching*.
- Collier, D., LaPorte, J. & Seawright, J. (2012). Putting typologies to work: Concept formation, measurement, and analytic rigor. *Political Research Quarterly*, 65(1), 217–232.
- Couso, D. & Garrido-Espeja, A. (2017). Models and Modelling in Pre-service Teacher Education: Why We Need Both. In K. Hahl, K. Juuti, J. Lampiselkä, A. Uitto & J. Lavonen (Hrsg.), *Cognitive and Affective Aspects in Science Education Research: Selected Papers from the ESERA 2015 Conference* (S. 245–261). Springer International Publishing. [https://doi.org/10.1007/978-3-319-58685-4\\_19](https://doi.org/10.1007/978-3-319-58685-4_19)

- Crawford, B. A. & Cullin, M. J. (200 ). Supporting prospective teachers' conceptions of modelling in science. *International Journal of Science Education*, 26(11), 1379–1401. doi:10.1080/09500690410001673775.
- Crawford, B. A. & Cullin, M. J. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. Boersma, M. Goedhart, O. de Jong & H. Eijkelhof (Hrsg.), *Research and the quality of science education* (S. 309–323). Springer.
- Cuperman, D. & Verner, I. M. (2019). Fostering Analogical Reasoning Through Creating Robotic Models of Biological Systems. *Journal of Science Education and Technology*, 28, 90–103. <https://doi.org/10.1007/s10956-018-9750-4>
- Dalland, C., Klette, K. & Svenkerud, S. (2019). Video studies and the challenge of selecting time scales AU - Dalland, Cecilie P. *International Journal of Research & Method in Education*, 1–14. <https://doi.org/10.1080/1743727X.2018.1563062>
- Demirhan, E. & Şahin, F. (2021). The Effects of Different Kinds of Hands-on Modeling Activities on the Academic Achievement, Problem-Solving Skills, and Scientific Creativity of Prospective Science Teachers. *Research in Science Education*, 51, 1015–1033. <https://doi.org/10.1007/s11165-019-09874-0>
- Denzin, N. (2016). *The qualitative manifesto: A call to arms*. Routledge.
- Department of Basic Education Republic of South Africa. (2011). *National Curriculum Statement Life Sciences*.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung* (1. Aufl.). Kohlhammer-Standards Psychologie Studententext. Kohlhammer.
- Dounas-Frazer, D. R., Ríos, L., Pollard, B., Stanley, J. T. & Lewandowski, H. (2018). Characterizing lab instructors' self-reported learning goals to inform development of an experimental modeling skills assessment. *Physical Review Physics Education Research*, 14(2), 20118.
- Driver, R., Leach, J. & Millar, R. (1996). *Young people's images of science*. McGraw-Hill Education (UK).
- Dubs, R. (2004). Bildungsstandards - ein erfolgversprechender Paradigmawechsel? Ein Umsetzungsversuch als Diskussionsgrundlage im Fach Volkswirtschaftslehre. In M. Wosnitza (Hrsg.), *Erziehungswissenschaften: Bd. 16. Lernprozess, Lernumgebung und Lerndiagnostik: Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (S. 38–55). Verlag Empirische Pädagogik. <https://www.alexandria.unisg.ch/8950/>

- Eidin, E. & Bielik, T. (2020). *Characterizing computational thinking in the context of technology-enhanced multilevel system modeling*. National Association for Research in Science Teaching (NARST) Conference, Portland, OR.
- Elby, A. & Hammer, D. (2010). Epistemological resources and framing: A cognitive framework for helping teachers interpret and respond to their students' epistemologies. *L. D. Bendixen & F. C. Feucht (Eds.), Personal epistemology in the classroom: theory, research, and implications for practice*, 409–434. <https://doi.org/10.1017/CBO9780511691904.013>
- Ergazaki, M., Zogza, V. & Komis, V. (2007). Analysing students' shared activity while modeling a biological process in a computer-supported educational environment. *Journal of Computer Assisted Learning*, 23(2), 158-168. <https://www.semanticscholar.org/paper/Analysing-students'-shared-activity-while-modeling-Ergazaki-Zogza/66a085bcbdc28f535ed398a62849aa976eaedfec>
- Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, 87(3), 215.
- Finkelstein, N. (2005). Learning Physics in Context: A study of student learning about electricity and magnetism. *International Journal of Science Education*, 27(10), 1187–1209. <https://doi.org/10.1080/09500690500069491>
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S. & Fischer, M. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 2(3), 28–45.
- Flick, U. (2010). Gütekriterien qualitativer Forschung. In G. May & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 395–407). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-92052-8\\_28](https://doi.org/10.1007/978-3-531-92052-8_28)
- Forbes, C. T., Lange-Schubert, K., Vo, T., Gogolin, S. & Böschl, F. (2017). *Model-based learning in primary science: A collaborative approach to exploring strategies for assessing scientific modelling*. Dublin City University. European Science Education Research Association, Dublin, Ireland.
- Friege, G. & Lind, G. (2003). Allgemeine und fachspezifische Problemlösekompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 63–74. [ftp://ftp.rz.uni-kiel.de/pub/ipn/zfdn/2003/4.friege\\_lind\\_63-74.pdf](ftp://ftp.rz.uni-kiel.de/pub/ipn/zfdn/2003/4.friege_lind_63-74.pdf)
- Frigg, R. & Hartmann, S. (2020). Models in science. In Edward N. Zalta (Hrsg.), *In Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/models-science/>
- Funke, J. (2011). Problemlösen. In T. Betsch, J. Funke & H. Plessner (Hrsg.), *Denken–Urteilen, Entscheiden, Problemlösen*. Springer Nature.

- Giere, R. N. (1999). *Science without laws*. Chicago, IL: The University of Chicago Press.
- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, 71(5), 742–752. <https://doi.org/10.1086/425063>
- Giere, R. N., Bickle, J. & Mauldin, R. (2006). *Understanding scientific reasoning* (5. Aufl.). Thomson Wadsworth.
- Gilbert, J. K. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education*, 2(2), 115–130. <https://doi.org/10.1007/s10763-004-3186-4>
- Gilbert, J. K. (2006). On the Nature of “Context” in Chemical Education. *International Journal of Science Education*, 28(9), 957–976. <https://doi.org/10.1080/09500690600702470>
- Gilbert, J. K., Boulter, C. J. & Rutherford, M. (1998). Models in explanations, Part 1: Horses for courses? *International Journal of Science Education*, 20(1), 83–97. <https://doi.org/10.1080/0950069980200106>
- Gilbert, J. K. & Justi, R. (2016). *Modelling-based teaching in science education*. Springer.
- Gilbert, J. K. & Osborne, R. J. (1980). The use of models in science and science teaching. *European Journal of Science Education*, 2(13), Artikel 3. <https://doi.org/10.1080/0140528800020103>
- Glanville, R. (1982). Inside every white box there are two black boxes trying to get out. *Behavioral Science*, 27(1), 1–11. <https://doi.org/10.1002/bs.3830270102>
- Glaser, R., Schauble, L., Raghavan, K. & Zeitz, C. (1992). Scientific reasoning across different domains. In E. de Corte, M. C. Linn, H. Mandl & L. Verschaffel (Hrsg.), *Computer-based learning environments and problem solving* (S. 345–371). Springer.
- Gobert, J. D. & Pallant, A. (200 ). Fostering students’ epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology*, 13(1), 7–22. <https://doi.org/10.1023/B:JOST.0000019635.70068.6f>
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and philosophy*, 21(5), 725–740.
- Göhner, M., Bielik, T. & Krell, M. (2022). Investigating the dimensions of the modeling competence among pre-service science teachers: Meta-modeling knowledge, modeling practice and modeling product. *Journal of Research in Science Teaching*, 59(8), 1354–1387. <https://doi.org/10.1002/tea.21759>
- Göhner, M. & Krell, M. (2018). Modellierungsprozesse von Lehramtsstudierenden der Biologie. *Erkenntnisweg Biologiedidaktik*, 17, 45–63.

- 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>
- Göhner, M. & Krell, M. (2021). Was ist schwierig am Modellieren? Identifikation und Beschreibung von Hindernissen in Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer. *Zeitschrift für Didaktik der Naturwissenschaften*, 27, 155–180. <https://doi.org/10.1007/s40573-021-00131-4>
- Göhner, M. & Krell, M. (2022a). Analyzing the relationships between pre-service biology teachers' modeling processes, scientific reasoning competencies and general cognitive abilities. In M. Ergazaki & K. Kampourakis (Hrsg.), *Contributions from Biology Education Research* (S. 231–244). Springer Nature.
- Göhner, M. & Krell, M. (2022b). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52, 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
- Gouvea, J. & Passmore, C. M. (2017). 'Models of' versus 'Models for'. *Science & Education*, 26(1-2), 49–63. <https://doi.org/10.1007/s11191-017-9884-4>
- Gräber, W., Nentwig, P., Koballa, T. R. & Evans, R. (Hrsg.). (2002). *Scientific literacy*. Springer VS.
- Gräsel, C. & Trempler, K. (2017). *Entwicklung von Professionalität pädagogischen Personals*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-07274-2>
- Gropengiesser, H. (2005). Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lernforschung. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (2. Aufl., S. 172–189). Beltz.
- Grosslight, L., Unger, C., Jay, E. & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Grünkorn, J. (2014). *Modellkompetenz im Biologieunterricht: Empirische Analyse von Modellkompetenz bei Schülerinnen und Schülern der Sekundarstufe I mit Aufgaben im offenen Antwortformat* [Dissertation, Freie Universität Berlin, Berlin]. EndNote.
- Grünkorn, J. & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells zur Modellkompetenz. In U. Harms & F. X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik: Didaktik der Biologie - Standortbestimmung und Perspektiven* (Bd. 5, S. 9–27). Studienverlag.

## Literaturverzeichnis

- Günther, S. L., Fleige, J., Upmeier zu Belzen, A. & Krüger, D. (2017). Interventionsstudie mit angehenden Lehrkräften zur Förderung von Modellkompetenz im Unterrichtsfach Biologie. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals: Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 215–236). Springer VS.
- Günther, S. L., Fleige, J., Upmeier zu Belzen, A. & Krüger, D. (2019). Using the case method to foster preservice biology teachers' content knowledge and pedagogical content knowledge related to models and modeling. *Journal of Science Teacher Education*, 30(4), 321–343. <https://doi.org/10.1080/1046560X.2018.1560208>
- Gut, C., Metzger, S., Hild, P. & Tardent, J. (2014). *Problemtypenbasierte Modellierung und Messung experimenteller Kompetenzen von 12- bis 15-jährigen Jugendlichen*. Didaktik der Physik Frühjahrstagung. [https://phzh.ch/MAPortrait\\_Data/158541/11/Gut\\_etal\\_2014\\_Problemtypenbasierte\\_Modellierung.pdf](https://phzh.ch/MAPortrait_Data/158541/11/Gut_etal_2014_Problemtypenbasierte_Modellierung.pdf)
- Halloun, I. A. (2004). *Modeling theory in science education*. Kluwer. <https://doi.org/10.1007/978-94-010-0876-1>
- Hammann, M. (2004). Kompetenzentwicklungsmodelle: Merkmale und ihre Bedeutung - dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 57(4), 196–203.
- Hammann, M., Phan, T. H. & Bayrhuber, H. (2008). Experimentieren als Problemlösen: Lässt sich das SDDS- Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaft* (S. 33–49). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-90865-6\\_3](https://doi.org/10.1007/978-3-531-90865-6_3)
- Harlow, D. B., Bianchini, J. A., Swanson, L. H. & Dwyer, H. A. (2013). Potential teachers' appropriate and inappropriate application of pedagogical resources in a model-based physics course: A “knowledge in pieces” perspective on teacher learning. *Journal of Research in Science Teaching*, 50(9), 1098–1126. <https://doi.org/10.1002/tea.21108>
- Harré, R. (1970). *The principles of scientific thinking*. Macmillan and Co Ltd.
- Harrison, A. (2001). How do Teachers and Textbook Writers Model Scientific Ideas for Students? *Research in Science Education*, 31(3), 401–435. <https://doi.org/10.1023/A:1013120312331>
- Harrison, A. & Treagust, D. (2000). A typology of school science models. *International Journal of Science Education*, 22(9), 1011–1026.

- Hartig, J. & Frey, A. (2012). Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten. *Psychologische Rundschau*, 63(1), 43–49. <https://doi.org/10.1026/0033-3042/a000109>
- Hartmann, S., Upmeyer zu Belzen, A., Krüger, D. & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of pre-service science teachers' competencies. *Zeitschrift für Psychologie*, 223(1), 47–53. <https://doi.org/10.1027/2151-2604/a000199>
- Harwood, W. S. (2004). A new model for inquiry: Is the scientific method dead? *Journal of College Science Teaching*, 33(7), 29–33.
- Hawking, S. W. & Mlodinow, L. (2011). *The grand design: New answers to the ultimate questions of life*. Bantam.
- Heijnes, D., van Joolingen, W. & Leenaars, F. (2018). Stimulating Scientific Reasoning with Drawing-Based Modeling. *Journal of Science Education and Technology*, 27(1), 45–56.
- Helsper, W., Herwartz-Emden, L. & Terhart, E. (2001). Qualität qualitativer Forschung in der Erziehungswissenschaft. *Zeitschrift für Pädagogik*, 14(2), 251–269.
- Henze, I., van Driel, J. H. & Verloop, N. (2007). Science Teachers' Knowledge about Teaching Models and Modelling in the Context of a New Syllabus on Public Understanding of Science. *Research in Science Education*, 37(2), 99–122. <https://doi.org/10.1007/s11165-006-9017-6>
- Henze, I., van Driel, J. H. & Verloop, N. (200 ). Development of Experienced Science Teachers' Pedagogical Content Knowledge of Models of the Solar System and the Universe. *International Journal of Science Education*, 30(10), 1321–1342. <https://doi.org/10.1080/09500690802187017>
- Hetmanek, A., Engelmann, K., Opitz, A. & Fischer, F. (2018). Beyond intelligence and domain knowledge: Scientific reasoning and argumentation as a set of cross-domain skills. *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge*. New York: Routledge.
- 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>
- Hogan, K. & Thomas, D. (2001). Cognitive Comparisons of Students' Systems Modeling in Ecology. *Journal of Science Education and Technology*, 10(4), 319–345.
- Höttecke, D. & Allchin, D. (2020). Reconceptualizing nature-of-science education in the age of social media. *Science Education*, 104(4), 641–666. <https://doi.org/10.1002/scs.21575>

- Howe, K. & Eisenhart, M. (1990). Standards for Qualitative (and Quantitative) Research: A Prolegomenon. *Educational researcher*, 19(4), 2–9. <https://doi.org/10.3102/0013189x019004002>
- Hsu, Y.-S., Lin, L.-F., Wu, H.-K., Lee, D.-Y. & Hwang, F.-K. (2012). A Novice-Expert Study of Modeling Skills and Knowledge Structures about Air Quality. *Journal of Science Education and Technology*, 21(5), 588–606. <https://doi.org/10.1007/s10956-011-9349-5>
- Hung, J.-Y., Chang, H.-Y. & Hung, J.-F. (2011). An Experienced Science Teacher's Metavisualization in the Case of the Complex System of Carbon Cycling. *Research in Science Education*, 51, 493–521. <https://doi.org/10.1007/s11165-018-9804-x>
- Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6).
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85.
- Jonassen, D. H. & Ionas, I. G. (2008). Designing effective supports for causal reasoning. *Educational Technology Research and Development*, 56(3), 287–308. <https://doi.org/10.1007/s11423-006-9021-6>
- Justi, R. & Gilbert, J. K. (2002). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. *International Journal of Science Education*, 24(12), 1273–1292. <https://doi.org/10.1080/09500690210163198>
- Justi, R. & Gilbert, J. K. (2004). Teachers' views on the nature of models. *International Journal of Science Education*, 25(11), 1369–1386. <https://doi.org/10.1080/0950069032000070324>
- Justi, R. & van Driel, J. H. (2005). The development of science teachers' knowledge on models and modelling: promoting, characterizing, and understanding the process. *International Journal of Science Education*, 27(5), 549–573.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Ke, L. & Schwarz, C. V. (2020). Supporting students' meaningful engagement in scientific modeling through epistemological messages: A case study of contrasting teaching approaches. *Journal of Research in Science Teaching*, 58(3), 335–365. <https://doi.org/10.1002/tea.21662>
- Kelle, U. (Hrsg.). (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: Theoretische Grundlagen und methodologische Konzepte* (2. Aufl.). VS Verlag für Sozialwissenschaften.

## Literaturverzeichnis

- Kelle, U. & Kluge, S. (2010). *Vom Einzelfall zum Typus: Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung* (2. Aufl.). *Qualitative Sozialforschung*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92366-6>
- Khan, S. (2007). Model-based inquiries in chemistry. *Science Education*, 91(6), 877–905. <https://doi.org/10.1002/sce.20226>
- Khan, S. (2011). What's missing in model-based teaching. *Journal of Science Teacher Education*, 22(6), 535–560.
- Kind, P. & Osborne, J. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, 101(1), 8–31. <https://doi.org/10.1002/sce.21251>
- Kiroğlu, K., Türk, C. & Erdoğan, İ. (2021). Which One Is More Effective in Teaching the Phases of the Moon and Eclipses: Hands-On or Computer Simulation? *Research in Science Education*, 51, 733–760. <https://doi.org/10.1007/s11165-018-9815-7>
- Klahr, D. (2002). *Exploring science: The cognition and development of discovery processes*. MIT press.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive science*, 12(1), 1–48.
- Klahr, D., Fay, A. L. & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive psychology*, 25(1), 111–146.
- Klahr, D. & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661–667. <https://doi.org/10.1111/j.0956-7976.2004.00737.x>
- 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: Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung* (S. 5–15).
- Kluge, S. (2000). Empirisch begründete Typenbildung in der qualitativen Sozialforschung, 1(1), Artikel 14.
- KMK. (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Wolters Kluwer.
- KMK. (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_16-Fachprofile-Lehrerbildung.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf)
- KMK (Hrsg.). (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der BRD.

- [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2020/2020\\_06\\_18-BildungsstandardsAHR\\_Biologie.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf)
- Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of Science*, 72(5), 1260–1271.
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, 42(2), 262–271. <https://doi.org/10.1016/j.shpsa.2010.11.034>
- Koch, S., Krell, M. & Krüger, D. (2015). Förderung von Modellkompetenz durch den Einsatz einer Blackbox. *Erkenntnisweg Biologiedidaktik*, 14, 93–108.
- Koć-Januchta, M., Höffler, T., Thoma, G.-B., Prectl, H. & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures – An eye-tracking study. *Computers in Human Behavior*, 68, 170–179. <https://doi.org/10.1016/j.chb.2016.11.028>
- Konrad, K. (2010). Lautes Denken. In G. May & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 476–490). VS Verlag für Sozialwissenschaften.
- Krell, M. (2013). *Wie Schülerinnen und Schüler biologische Modelle verstehen: Erfassung und Beschreibung des Modellverstehens von Schülerinnen und Schülern der Sekundarstufe I*. Logos.
- Krell, M. (2019). Assessment of meta-modeling knowledge: Learning from triadic concepts of models in the philosophy of science. *Science Education Review Letters*, 1–7. <https://doi.org/10.18452/19813>
- Krell, M. & Hergert, S. (2019). The blackbox approach: Analyzing modeling strategies. In A. Upmeyer zu Belzen, D. Krüger & J. H. van Driel (Hrsg.), *Towards a competence-based view on models and modeling in science education* (S. 147–160). Springer.
- Krell, M., Hergert, S. & Krüger, D. (2017). Pre-service Science Teachers' Modelling-strategies.
- Krell, M. & Krüger, D. (2013). Wie werden Modelle im Biologieunterricht eingesetzt? Ergebnisse einer Fragebogenstudie. *Erkenntnisweg Biologiedidaktik*, 12, 9–26.
- Krell, M. & Krüger, D. (2016). Testing models: A key aspect to promote teaching activities related to models and modelling in biology lessons? *Journal of Biological Education*, 50(2), 160–173. <https://doi.org/10.1080/00219266.2015.1028570>
- Krell, M. & Krüger, D. (2017). University students' meta-modelling knowledge. *Research in Science & Technological Education*, 35(3), 261–273. <https://doi.org/10.1080/02635143.2016.1274724>
- Krell, M., Reinisch, B. & Krüger, D. (2015). Analyzing Students' Understanding of Models and Modeling Referring to the Disciplines Biology, Chemistry, and Physics. *Research in Science Education*, 45(3), 367–393. <https://doi.org/10.1007/s11165-014-9427-9>

- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2012). Students' understanding of the purpose of models in different biological contexts. *International Journal of Biology Education*, 2, Artikel 2, 1–34.
- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2014a). Context-specificities in students' understanding of models and modelling. In C. P. Constantinou, N. Papadouris & A. Hadjigeorgiou (Hrsg.), *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning* (S. 1024–1035).
- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2014b). Students' levels of understanding models and modelling in biology: Global or aspect-dependent? *Research in Science Education*, 44(1), 109–132. <https://doi.org/10.1007/s11165-013-9365-y>
- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2016). Modellkompetenz im Biologieunterricht. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologiedidaktische Forschung: Schwerpunkte und Forschungsgegenstände* (S. 83–102). Logos.
- Krell, M., Walzer, C., Hergert, S. & Krüger, D. (2019). Development and application of a category system to describe pre-service science teachers' activities in the process of scientific modelling. *Research in Science Education*, 49(5), 1319–1345. <https://doi.org/10.1007/s11165-017-9657-8>
- Krippendorff, K. (1980). Validity in content analysis. In E. Mochmann (Hrsg.), *Computerstrategien für die Kommunikationsanalyse* (S. 69–112). Campus.
- Krüger, D., Hartmann, S., Nordmeier, V. & Upmeier zu Belzen, A. (2020a). Interventionsstudie zur Wirkung von schülerzentrierten Methoden auf Modellkompetenz und Fachwissen. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper & C. Lautenbach (Hrsg.), *Student Learning in German Higher Education*. Springer.
- Krüger, D., Hartmann, S., Nordmeier, V. & Upmeier zu Belzen, A. (2020b). Measuring scientific reasoning competencies: Multiple aspects of validity. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper & C. Lautenbach (Hrsg.), *Student Learning in German Higher Education* (S. 261–280). Springer.
- Krüger, D. & Krell, M. (2020). Maschinelles Lernen mit Aussagen zur Modellkompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 26, 157–172. <https://doi.org/10.1007/s40573-020-00118-7>
- Krüger, D. & Riemeier, T. (2014). Die qualitative Inhaltsanalyse – eine Methode zur Auswertung von Interviews. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *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*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s40573-021-00129-y>
- Kuckartz, U. (2010). Typenbildung. In G. May & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 553–568). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-92052-8\\_39](https://doi.org/10.1007/978-3-531-92052-8_39)
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (3 Aufl.). *Grundlagentexte Methoden*. Beltz.
- Lamnek, S. (2005). *Qualitative Sozialforschung* (4. Aufl.). PVU.
- Lamnek, S. (2010). *Qualitative Sozialforschung* (5. Aufl.). Beltz.
- Lämsä, J., Uribe, P., Jiménez, A., Caballero, D., Härmäläinen, R. & Araya, R. (2021). Deep Networks for Collaboration Analytics: Promoting Automatic Analysis of Face-to-Face Interaction in the Context of Inquiry-Based Learning. *Journal of Learning Analytics*, 8(1), 113–125. <https://doi.org/10.18608/jla.2021.7118>
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 33(1), 159–174.
- Lange, D. (2013). *Inhaltsanalytische Untersuchung zur Kooperation beim Bearbeiten mathematischer Problemaufgaben* (1. Aufl.). Waxmann Verlag GmbH. [http://www.content-select.com/index.php?id=bib\\_view&ean=9783830978824](http://www.content-select.com/index.php?id=bib_view&ean=9783830978824)
- Laubichler, M. D. & Müller, G. B. (2007). *Modeling biology: Structures, behavior, evolution*. MIT press.
- Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, 2(3), 307.
- Lazenby, K., Stricker, A., Brandriet, A., Rupp, C., Mauger-Sonnek, K. & Becker, N. (2020). Mapping undergraduate chemistry students' epistemic ideas about models and modeling. *Journal of Research in Science Teaching*, 57(5), 794–824. <https://doi.org/10.1002/tea.21614>
- Leden, L., Hansson, L. & Ideland, M. (2020). The mangle of school science practice: Teachers' negotiations of two nature of science activities at different levels of contextualization. *Science Education*, 104(1), 5–26. <https://doi.org/10.1002/sc.21553>
- Lederman, N. G. & Abd-El-Khalick, F. (2002). Avoiding de-natured science: Activities that promote understandings of the nature of science. In W. McComas (Hrsg.), *The Nature of Science in Science Education* (Bd. 5, S. 83–126). Kluwer.

- Lee, S. & Kim, H.-B. (2011). Exploring secondary students' epistemological features depending on the evaluation levels of the group model on blood circulation. *Science & Education*, 23(5), 1075–1099. <https://doi.org/10.1007/s11191-013-9639-9>
- Lee, T. D. & Jones, M. G. (2011). Elementary Teachers' Selection and Use of Visual Models. *Journal of Science Education and Technology*, 27(1), 1–29.
- Leenaars, F. A. J., van Joolingen, W. R. & Bollen, L. (2011). Using self-made drawings to support modelling in science education. *British Journal of Educational Technology*, 44(1), 82–94. <https://doi.org/10.1111/j.1467-8535.2011.01272.x>
- Lehrer, R. & Schauble, L. (2006a). *Cultivating Model-Based Reasoning in Science Education*. Cambridge University Press.
- Lehrer, R. & Schauble, L. (2006b). Scientific thinking and science literacy. *Handbook of child psychology*, 6, 153–196.
- Leisner-Bodenthin, A. (2006). Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 91–109.
- Leuders, T. (2015). Empirische Forschung in der Fachdidaktik. Eine Herausforderung für die Professionalisierung und die Nachwuchsqualifizierung. *Beiträge zur Lehrerinnen-und Lehrerbildung*, 33(2), 215–234.
- Levy, A. & Currie, A. (2015). Model Organisms are Not (Theoretical) Models. *The British Journal for the Philosophy of Science*, 66(2), 327–348. <https://doi.org/10.1093/bjps/axt055>
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R: IST 2000 R* (2. Aufl.). Hogrefe.
- Löffler, P. (2016). *Modellanwendung in Problemlöseaufgaben: Wie wirkt Kontext? Studien zum Physik- und Chemielernen*. Logos.
- Löffler, P. & Kauertz, A. (2016). Modellanwendung in Problemlöseaufgaben: Wie wirkt Kontext? *Gesellschaft für Didaktik der Chemie und Physik (GDGP)*(36), 41–43. <https://doi.org/10.1002/sce.20186>
- Löhner, S., van Joolingen, W. R. & Savelsbergh, E. R. (2003). The effect of external representation on constructing computer models of complex phenomena. *Instructional Science*, 31(6), 395–418. <https://doi.org/10.1023/a:1025746813683>
- Löhner, S., van Joolingen, W. R., Savelsbergh, E. R. & van Hout-Wolters, B. (2005). Students' reasoning during modeling in an inquiry learning environment. *Computers in Human Behavior*, 21(3), 441–461. <https://doi.org/10.1016/j.chb.2004.10.037>

- Louca, L. T. & Zacharia, Z. C. (2012). Modeling-based learning in science education: Cognitive, meta-cognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471–492. <https://doi.org/10.1080/00131911.2011.628748>
- Louca, L. T. & Zacharia, Z. C. (2015). Examining learning through modeling in k-6 science education. *Journal of Science Education and Technology*, 24(2), 192–215. <https://doi.org/10.1007/s10956-014-9533-5>
- Magnani, L. (1999). Model-Based Creative Abduction. In L. Magnani, N. J. Nersessian & P. Thagard (Hrsg.), *Model-Based Reasoning in Scientific Discovery*. Springer.
- Mahr, B. (2008). Ein Modell des Modellseins. Ein Beitrag zur Aufklärung des Modellbegriffs. *Modelle*, 187–218.
- Mahr, B. (2011). On the epistemology of models. In G. Abel & J. Conant (Hrsg.), *Berlin Studies in Knowledge Research. Rethinking Epistemology: Volume 1* (1. Aufl., S. 301–352). De Gruyter. <https://www.degruyter.com/view/book/9783110253573/10.1515/9783110253573.301.xml>
- Maia, P. F. & Justi, R. (2009). Learning of chemical equilibrium through modelling-based teaching. *International Journal of Science Education*, 31(5), 603–630. doi:10.1080/09500690802538045.
- Makel, M. C. & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational researcher*, 43(6), 304–316. <https://doi.org/10.3102/0013189x14545513>
- Mathesius, S., Hartmann, S., Upmeier zu Belzen, A. & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education: Assessing competencies using a paper-pencil test. In T. Tal & A. Yarden (Hrsg.), *The future of biology education research* (S. 93–110).
- Mathesius, S., Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2019). Überprüfung eines Tests zum wissenschaftlichen Denken unter Berücksichtigung des Validitätskriteriums relations-to-other-variables. Vorab-Onlinepublikation. <https://doi.org/10.25656/01:23991>
- Max, C. (1999). *Entwicklung von Kompetenz: Ein neues Paradigma für das Lernen in Schule und Arbeitswelt ; Ertrag und Perspektiven der Französischsprachigen Kompetenzforschung und ihre Bedeutung als Gestaltungsprinzip von Bildung. Europäische Hochschulschriften. Reihe 11: Bd. 796*. Peter Lang.
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung* (S. 177–186). Springer.
- Mayring, P. (2002). *Einführung in die Qualitative Sozialforschung* (5. Aufl.). Beltz.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Beltz.

- Mayring, P. (2019). Qualitative Inhaltsanalyse - Abgrenzungen, Spielarten, Weiterentwicklungen. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3). <http://www.qualitative-research.net/index.php/fqs/article/download/3343/4558>
- Mayring, P. & Gläser-Zikuda, M. (Hrsg.). (2005). *Die Praxis der Qualitativen Inhaltsanalyse* (2. Aufl.). Beltz.
- McComas, W. (Hrsg.). (2002). *The Nature of Science in Science Education*. Kluwer.
- Meister, S., Krell, M., Göhner, M. & Upmeyer zu Belzen, A. (2021). Pre-service biology teachers' responses to first-hand anomalous data during modelling Processes. *Research in Science Education*, 51, 1459–1479. <https://doi.org/10.1007/s11165-020-09929-7>
- Mendonça, P. C. C. & Justi, R. (2011). Contributions of the Model of Modelling Diagram to the Learning of Ionic Bonding: Analysis of A Case Study. *Research in Science Education*, 41(4), 479–503. <https://doi.org/10.1007/s11165-010-9176-3>
- Mendonça, P. C. C. & Justi, R. (2013). The Relationships Between Modelling and Argumentation from the Perspective of the Model of Modelling Diagram. *International Journal of Science Education*, 35(14), 2407–2434. <https://doi.org/10.1080/09500693.2013.811615>
- Mendonça, P. C. C. & Justi, R. (2014). An instrument for analyzing arguments produced in modeling-based chemistry lessons. *Journal of Research in Science Teaching*, 51(2), 192–218. <https://doi.org/10.1002/tea.21133>
- Mierdel, J. & Bogner, F. X. (2019). Is creativity, hands-on modeling and cognitive learning gender-dependent? *Thinking skills and creativity*, 31, 91–102. <https://doi.org/10.1016/j.tsc.2018.11.001>
- Mierdel, J. & Bogner, F. X. (2021). Investigations of Modellers and Model Viewers in an Out-of-School Gene Technology Laboratory. *Research in Science Education*, 51, 801–822. <https://doi.org/10.1007/s11165-019-09871-3>
- Millar, R. (1996). Towards a Science Curriculum for Public Understanding. *School Science Review*, 77(280), 7–18. <https://eric.ed.gov/?id=EJ522129>
- Miller, A. R. & Kastens, K. A. (2018). Investigating the impacts of targeted professional development around models and modeling on teachers' instructional practice and student learning. *Journal of Research in Science Teaching*, 55(5), 641–663. <https://doi.org/10.1002/tea.21434>
- Morrison, M. & Morgan, M. S. (1999). Models as Mediators: Perspectives on Natural and Social Science. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511660108.003, 10–37.
- Morzy, M., Kajdanowicz, T. & Kaziemko, P. (2017). On measuring the complexity of networks: Kolmogorov complexity versus entropy. *Complexity*, 2017.

- Nachtigall, C., Kröhne, U., Enders, U. & Steyer, R. (2008). Causal effects and fair comparison: Considering the influence of context variables on student competencies. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of Competencies in Educational Contexts* (1st ed., S. 315–335). Hogrefe Publishing.
- National Academy of Sciences. (1996). *National Science Education Standards: Observe, interact, change, learn*. National Academy Press. <https://doi.org/10.17226/4962>
- Nehring, A., Nowak, K. H., Upmeyer zu Belzen, A. & Tiemann, R. (2015). Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic variables. *International Journal of Science Education*, 37(9), 1343–1363. <https://doi.org/10.1080/09500693.2015.1035358>
- Nehring, A. & Schwichow, M. (2020). Was ist Wissen, was ist Können? Deutungen des Kompetenzbegriffs und deren psychometrische Konsequenzen im Kontext von Fachwissen und Variablenkontrollstrategie. *Zeitschrift für Didaktik der Naturwissenschaften*, 26, 73–87. <https://doi.org/10.1007/s40573-020-00113-y>
- Neilson, D., Campbell, T. & Allred, B. (2010). Model-based inquiry: A buoyant force module for high school physics classes. *The Science Teacher*, 77(8), 38–43.
- Nelson, M. M. & Davis, E. A. (2012). Preservice Elementary Teachers' Evaluations of Elementary Students' Scientific Models: An aspect of pedagogical content knowledge for scientific modeling. *International Journal of Science Education*, 34(12), 1931–1959. <https://doi.org/10.1080/09500693.2011.594103>
- Nersessian, N. J. (1999). Model-based reasoning in conceptual change. In L. Magnani, N. J. Nersessian & P. Thagard (Hrsg.), *Model-Based Reasoning in Scientific Discovery*. Kluwer Academic/Plenum Publishers.
- Nersessian, N. J. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich & M. Siegal (Hrsg.), *The cognitive basis of science* (S. 133–153). Cambridge University Press.
- Neumann, K., Härtig, H., Harms, U. & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki & Hirano T. (Hrsg.), *Model science teacher preparation programs: An international comparison of what works*. Information Age.
- Neumann, K., Schecker, H. & Theyßen, H. (2019). Assessing complex patterns of student resources and behavior in the large scale. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 233–249. <https://doi.org/10.1177/0002716219844963>
- Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.

- NGSS Lead States (Hrsg.). (2013a). *Next Generation Science Standards: For States, By States*. The National Academies Press.
- NGSS Lead States. (2013b). Next generation science standards: For states, by states: Appendix f - science and engineering practices in the ngss. In NGSS Lead States (Hrsg.), *Next Generation Science Standards: For States, By States*. The National Academies Press.
- NGSS Lead States. (2017). *Next Generation Science Standards: For States, By States*. The National Academies Press.
- Nicolaou, C. T. & Constantinou, C. P. (2014). Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13, 52–73. <https://doi.org/10.1016/j.edurev.2014.10.001>
- Nielsen, S. S. & Nielsen, J. A. (2021a). A competence-oriented approach to models and modelling in lower secondary science education: Practices and rationales among danish teachers. *Research in Science Education*, 51, 565–593. <https://doi.org/10.1007/s11165-019-09900-1>
- Nielsen, S. S. & Nielsen, J. A. (2021b). Models and Modelling: Science Teachers' Perceived Practice and Rationales in Lower Secondary School in the Context of a Revised Competence-Oriented Curriculum. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(4), Artikel 1954. <https://doi.org/10.29333/ejmste/10790>
- Nordheimer, R. (2019). *Volition in Modellierungsprozessen von Biologie-Lehramtsstudierenden: Entwicklung eines Kategoriensystems zur Erfassung der Selbststeuerung* [unveröffentlichte Master Thesis]. Freie Universität Berlin, Berlin.
- Oh, P. S. & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education*, 33(8), 1109–1130.
- Orsenne, J. (2016). *Aktivierung von Schülervorstellungen zu Modellen durch praktische Tätigkeiten der Modellbildung* [Dissertation]. Humboldt Universität, Berlin.
- Orsenne, J. & Upmeier zu Belzen, A. (2012). Hands-On Aufgaben zur Erfassung und Förderung von Modellkompetenz im Biologieunterricht. In U. Harms & F. X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik: Didaktik der Biologie - Standortbestimmung und Perspektiven* (S. 29–44). Studienverlag.
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking skills and creativity*, 10, 265–279.
- Pasley, J., Trygstad, P. & Banilower, E. (2016). *What does "Implementing the NGSS" mean? Operationalizing the science practices for K-12 classrooms*. Horizon.

- Passmore, C. M., Gouvea, J. S. & Giere, R. (2014). Models in science and in learning science: Focusing scientific practice on sense-making. In M. R. Matthews (Hrsg.), *International handbook of research in history, philosophy and science teaching* (S. 1171–1202). Springer.
- Passmore, C. M., Stewart, J. & Cartier, J. (200 ). Model-based inquiry and school science: Creating connections. *School Science and Mathematics*, 109(7), 394–402.
- Passmore, C. M. & Svoboda, J. (2012). Exploring opportunities for argumentation in modeling classrooms. *International Journal of Science Education*, 34(10), 1535–1554.
- Patton, M. (1990). Purposeful sampling. *Qualitative evaluation and research methods*, 2, 169–186.
- Patton, M. (2001). *Qualitative research & evaluation methods* (3. Aufl.). SAGE Publications.
- Philips, B. U. (2020). *Flattening the Curve*. Rural Health Quarterly. <https://ruralhealthquarterly.com/home/2020/06/09/flattening-the-curve/>
- Pitt Hild. (2020). *Problemtypenbasierte Kompetenzmodellierung beim praktisch-naturwissenschaftlichen Arbeiten: Design, Validierung und Einsatz von Aufgaben zum effektbasierten Vergleichen* [Dissertation], Zürich.
- Pluta, W. J., Chinn, C. A. & Duncan, R. G. (2011). Learners’ epistemic criteria for good scientific models. *Journal of Research in Science Teaching*.
- Porter, M. A., Onnela, J.-P. & Mucha, P. J. (2009). Communities in Networks. *Notices of the American Mathematical Society*, 56(9), 1082–1097.
- Quigley, D., Ostwald, J. & Sumner, T. (2017). Scientific modeling: Using learning analytics to examine student practices and classroom variation. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 329–338.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In Rasmussen J. & R. W. B (Hrsg.), *NATO Conference Series: Bd. 15. Human Detection and Diagnosis of System Failures* (S. 241–258). Springer.
- Reiff, R., Harwood, W. S. & Phillipson, T. (Hrsg.) (2002). *A scientific method based upon research scientists' conceptions of scientific inquiry*.
- Ríos, L., Pollard, B., Dounas-Frazer, D. R. & Lewandowski, H. J. (2019). Using think-aloud interviews to characterize model-based reasoning in electronics for a laboratory course assessment. *Physical Review Physics Education Research*, 15(1), Artikel 010140.

- 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>
- Ronqui, J. R. F. & Travieso, G. (2015). Analyzing complex networks through correlations in centrality measurements. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(5), Artikel P05030.
- Rowbottom, D. P. (200 ). Models in Biology and Physics: What’s the Difference? *Foundations of Science*, 14(4), 281–294. <https://doi.org/10.1007/s10699-009-9160-4>
- RStudio Team. (2022). *RStudio: Integrated Development for R*. RStudio, PBC. <http://www.rstudio.com/>
- Ruppert, J., Duncan, R. G. & Chinn, C. A. (2017). Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s11165-017-9656-9>
- Ruppert, J., Duncan, R. G. & Chinn, C. A. (2019). Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education*, 49(3), 921–948. <https://doi.org/10.1007/s11165-017-9656-9>
- Rychen, D. S. & Salganik, L. H. (2003). A holistic model of competence. In D. S. Rychen & L. H. Salganik (Hrsg.), *Key Competencies for a Successful Life and Well-Functioning Society* (1st ed., S. 41–62). Hogrefe Publishing.
- Sagan, C. (1990). Why we need to understand science. *Skeptical Inquirer*, 14(3), 263–269.
- Sandmann, A. (201 ). Lautes Denken - die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *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)
- Schauble, L., Glaser, R., Raghavan, K. & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, 1(2), 201–238. [https://doi.org/10.1207/s15327809jls0102\\_3](https://doi.org/10.1207/s15327809jls0102_3)
- Schauble, L., Klopfer, L. E. & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28(9), 859–882.
- Schreier, M. (2010). Fallauswahl. In G. May & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 238–251). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-92052-8\\_16](https://doi.org/10.1007/978-3-531-92052-8_16)
- Schreier, M. (2012). *Qualitative Content Analysis in Practice*. SAGE Publications.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum qualitative Sozialforschung*, 15(1). <https://doi.org/10.17169/fqs-15.1.2043>

- Schwartz, R. S. & Lederman, N. G. (2005). What scientists say: Scientists' views of models. *AERA*.
- Schwarz, C. V., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B. & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.
- Schwarz, C. V. & White, B. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165–205.
- Schwichow, M. & Nehring, A. (2018). Variablenkontrolle beim Experimentieren in Biologie, Chemie und Physik: Höhere Kompetenzausprägungen bei der Anwendung der Variablenkontrollstrategie durch höheres Fachwissen? Empirische Belege aus zwei Studien. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 217–233. <https://doi.org/10.1007/s40573-018-0085-8>
- Seah, Y. Y. & Magana, A. J. (201 ). Exploring Students' Experimentation Strategies in Engineering Design Using an Educational CAD Tool. *Journal of Science Education Technology*, 28, 195–208. <https://doi.org/10.1007/s10956-018-9757-x>
- Seale, C. (2002). Quality Issues in Qualitative Inquiry. *Qualitative Social Work*, 1(1), 97–110. <https://doi.org/10.1177/147332500200100107>
- Shavelson, R. (2013). On an Approach to Testing and Modeling Competence. *Educational Psychologist*, 48(2), 73–86. <https://doi.org/10.1080/00461520.2013.779483>
- Shi, F., Wang, L., Liu, X. & Chiu, M.-H. (2021). Development and validation of an observation protocol for measuring science teachers' modeling-based teaching performance. *Journal of Research in Science Teaching*. Vorab-Onlinepublikation. <https://doi.org/10.1002/tea.21712>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, 15(2), 4–14.
- Sikorski, T.-R. (201 ). Context-Dependent “Upper Anchors” for Learning Progressions. *Science & Education*, 28(8), 957–981. <https://doi.org/10.1007/s11191-019-00074-w>
- Sins, P. H., Savelsbergh, E. R. & van Joolingen, W. R. (2005). The difficult process of scientific modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education*, 27(14), 1695–1721.
- Sins, P. H., Savelsbergh, E. R., van Joolingen, W. R. & van Hout-Wolters, B. H. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education*, 31(9), 1205–1229.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Springer.

## Literaturverzeichnis

- Stamann, C., Janssen, M. & Schreier, M. (2016). *Qualitative Inhaltsanalyse – Versuch einer Begriffsbestimmung und Systematisierung / Searching for the Core: Defining Qualitative Content Analysis*.
- Steinke, I. (2013). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung: Ein Handbuch*. Rowohlt-Taschenbuch-Verl.
- Sung, J. Y. & Oh, P. S. (2011). Sixth Grade Students' Content-Specific Competencies and Challenges in Learning the Seasons Through Modeling. *Research in Science Education*, 48(4), 839–864. <https://doi.org/10.1007/s11165-016-9589-8>
- Tempel, B. J. (2017). *Vermittlung von Modellkompetenz in den Unterrichtsfächern Biologie und Chemie: Modellierung, Validierung und Messung Professioneller Unterrichtswahrnehmung zukünftiger Lehrkräfte mithilfe eines Vignettentests* [Dissertation]. Pädagogische Hochschule Heidelberg, Heidelberg.
- Terzer, E., Hartig, J. & Upmeyer zu Belzen, A. (2013). Systematische Konstruktion eines Tests zu Modellkompetenz im Biologieunterricht unter Berücksichtigung von Gütekriterien. *Zeitschrift für Didaktik der Naturwissenschaften*, 19(1), 51–76.
- Tobin, R. G., Lacy, S. J., Crissman, S. & Haddad, N. (2018). Model-based reasoning about energy: A fourth-grade case study. *Journal of Research in Science Teaching*.
- Torres, J. & Vasconcelos, C. (2015). Nature of science and models: Comparing portuguese prospective teachers' views. *EURASIA Journal of Mathematics, Science and Technology Education*, 11(6), 1473–1494. <https://doi.org/10.12973/eurasia.2015.1407a>
- Torrijos-Muelas, M., González-Víllora, S., Osma, A. R. B. & Guijarro, E. (2023). Teachers and pre-service teachers' scientific competencies: a methodological development for a systematic review. *International Journal of Science Education*, 1–32. <https://doi.org/10.1080/09500693.2023.2190851>
- Treagust, D. F., Chittleborough, G. & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, 24(4), 357–368. <https://doi.org/10.1080/09500690110066485>
- Treagust, D. F., Chittleborough, G. D. & Mamiala, T. L. (2004). Students' Understanding of the Descriptive and Predictive Nature of Teaching Models in Organic Chemistry. *Research in Science Education*, 34(1), 1–20. <https://doi.org/10.1023/B:RISE.0000020885.41497.ed>
- Unesco. (2015). *UNESCO science report: Towards 2030*. UNESCO Publ.

## Literaturverzeichnis

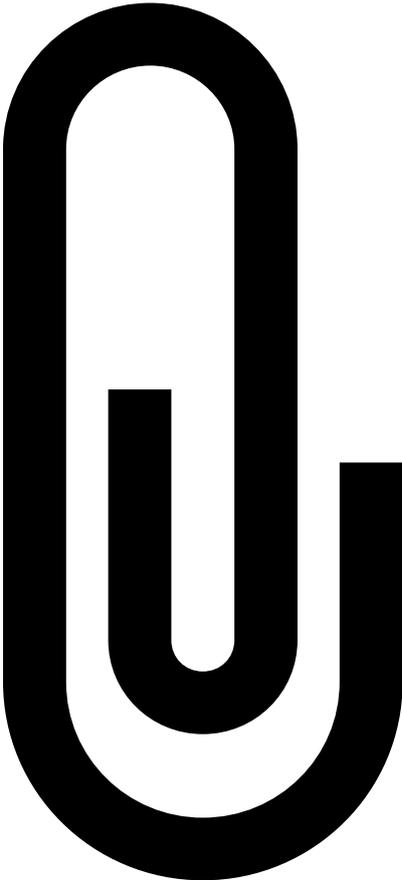
- Upmeier zu Belzen, A. (2014). Black Box: Modellierung von Prozessen naturwissenschaftlicher Erkenntnisgewinnung. In D. Ludwig, C. Weber & O. Zauzig (Hrsg.), *Das materielle Modell. Objektgeschichten aus der wissenschaftlichen Praxis* (S. 99–106). Wilhelm Fink.
- Upmeier zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- Upmeier zu Belzen, A., van Driel, J. H. & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger & J. H. van Driel (Hrsg.), *Models and Modeling in Science Education. Towards a Competence-Based View on Models and Modeling in Science Education* (1st ed. 2019, S. 3–19). Springer International Publishing. [https://doi.org/10.1007/978-3-030-30255-9\\_1](https://doi.org/10.1007/978-3-030-30255-9_1)
- van der Valk, T., van Driel, J. H. & Vos, W. de (2007). Common Characteristics of Models in Present-day Scientific Practice. *Research in Science Education*, 37(4), 469–488. <https://doi.org/10.1007/s11165-006-9036-3>
- van Driel, J. H. & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21, Artikel 11, 1141–1153. <https://doi.org/10.1080/095006999290110>
- van Driel, J. H. & Verloop, N. (2002). Experienced teachers' knowledge of teaching and learning of models and modelling in science education. *International Journal of Science Education*, 24(2), 1255–1272. <https://doi.org/10.1080/09500690210126711>
- van Fraassen, B. C. (2010). Scientific representation: Paradoxes of perspective. *Analysis*, 70(3), 511–514. <https://doi.org/10.1093/analys/anq042>
- van Meter, P. & Garner, J. (2005). The Promise and Practice of Learner-Generated Drawing: Literature Review and Synthesis. *Educational Psychology Review*, 17(4), 285–325. <https://doi.org/10.1007/s10648-005-8136-3>
- van Vorst, H., Dorschu, A., Fechner, S., Kauertz, A., Krabbe, H. & Sumfleth, E. (2015). Charakterisierung und Strukturierung von Kontexten im naturwissenschaftlichen Unterricht – Vorschlag einer theoretischen Modellierung. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 29–39. <https://doi.org/10.1007/s40573-014-0021-5>
- VCAA (2016). Victorian Curriculum: F-10. <https://victoriancurriculum.vcaa.vic.edu.au/science/curriculum/f-10>
- VERBI Software. (1989 – 2023). *MAXQDA [Computer software]*. Consult. Sozialforschung GmbH. Berlin.

- Vo, T., Forbes, C. T., Zangori, L. & Schwarz, C. V. (2015). Fostering third-grade students' use of scientific models with the water cycle: Elementary teachers' conceptions and practices. *International Journal of Science Education*, 37(15), 2411–2432.
- Vo, T., Forbes, C. T., Zangori, L. & Schwarz, C. V. (2019). Longitudinal investigation of primary in-service teachers' modelling the hydrological phenomena. *International Journal of Science Education*, 41(18), 2788–2807. <https://doi.org/10.1080/09500693.2019.1698786>
- Vollmeyer, R. & Funke, J. (1999). Personen-und Aufgabenmerkmale beim komplexen Problemlösen. *Psychologische Rundschau*, 50, 213–219.
- Wartofsky, M. W. (1966). *The model muddle: Proposals for an immodest realism*. *Boston Studies in the Philosophy of Science: Bd. 48*. Springer. <https://doi.org/10.1007/978-94-009-9357-0>
- Weinert, F. E. (2001a). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Hogrefe & Huber.
- Weinert, F. E. (Hrsg.). (2001b). *Leistungsmessungen in Schulen*. Beltz.
- Weiss, I., Pasley, J., Smith, P., Banilower, E. & Heck, D. (2003). *A study of K-12 mathematics and science education in the United States*. Horizon.
- Werner, S., Förtsch, C., Boone, W., Kotzebue, L. von & Neuhaus, B. J. (2019). Investigating how german biology teachers use three-dimensional physical models in classroom instruction: A video study. *Research in Science Education*, 49(2), 437–463. <https://doi.org/10.1007/s11165-017-9624-4>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wilmont, I., Barendsen, E. & Hoppenbrouwers, S. (2019). A case study of executive functions in real process modeling sessions. In H. A. Proper & J. Stirna (Hrsg.), *Lecture Notes in Business Information Processing. Advanced Information Systems Engineering Workshops* (Bd. 349, S. 17–28). Springer International Publishing. [https://doi.org/10.1007/978-3-030-20948-3\\_2](https://doi.org/10.1007/978-3-030-20948-3_2)
- Windschitl, M. & Thompson, J. (2006). Transcending simple forms of school science investigation: The impact of preservice instruction on teachers' understanding of model-based inquiry. *American Educational Research Journal*,
- Windschitl, M., Thompson, J. & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941–967. <https://doi.org/10.1002/sce.20259>
- Woolley, J. S., Deal, A. M., Green, J., HATHENBRUCK, F., Kurtz, S. A., Park, T. K. H., Pollock, S. V., Transtrum, M. B. & Jensen, J. L. (2018). Undergraduate students demonstrate common false scientific reasoning strategies. *Thinking skills and creativity*, 27, 101–113.

## Literaturverzeichnis

- Yenilmez Turkoglu, A. & Oztekin, C. (2016). Science teacher candidates' perceptions about roles and nature of scientific models. *Research in Science & Technological Education*, 34(2), 219–236. <https://doi.org/10.1080/02635143.2015.1137893>
- Zangori, L., Forbes, C. T. & Schwarz, C. V. (2015). Exploring the effect of embedded scaffolding within curricular tasks on third-grade students' model-based explanations about hydrologic cycling. *Science & Education*, 24(7-8), 957–981.
- Zangori, L., Peel, A., Kinslow, A., Friedrichsen, P. & Sadler, T. D. (2017). Student development of model-based reasoning about carbon cycling and climate change in a socio-scientific issues unit. *Journal of Research in Science Teaching*, 54(10), 1249–1273. <https://doi.org/10.1002/tea.21404>
- Zimmerman, C. (2005). The development of scientific reasoning skills. *Developmental review*, 20(1), 99–149.

9. Anhang



## 9.1 Zusammenfassung und Abstract

Modelle und das Modellieren sind in den Naturwissenschaften von zentraler Wichtigkeit und auch in der Lehre der Naturwissenschaften unverzichtbar. Dabei haben Modelle viele Anwendungszwecke; besonders epistemische Perspektiven auf Modelle und das Modellieren werden in Curricula weltweit betont. Um die curricularen Vorgaben erreichen zu können, benötigen Lehrkräfte der Naturwissenschaften selbst die entsprechende Modellierkompetenz, allerdings ist bisher kaum erforscht, wie Lehrkräfte der Naturwissenschaften selbst Modellieren. Die vorliegende Untersuchung nahm daher eine empirische Beschreibung individueller Modellierungsstrategien in den Fokus, indem fallbasiert und prozessorientiert eine Typologie von Modellierungsstrategien 36 Lehramtsstudierender der Biologie entwickelt wurde. Basierend auf dem Auftreten, der Häufigkeit und der Abfolge der Tätigkeiten wurden sechs typische Modellierungsstrategien differenziert und detailliert beschrieben. Die Ergebnisse der Untersuchung erweitern die bislang vorwiegend produktbasierten Kenntnisse naturwissenschaftsdidaktischer Forschung zu Modellen und dem Modellieren um eine Prozessfacette. Neben der erarbeiteten Typologie wurden zudem Erkenntnisse über die Dimensionalität und Extensionalität der Modellierkompetenz gewonnen, indem Zusammenhänge zwischen den identifizierten Modellierungsstrategien und weiteren Variablen untersucht wurden. Hierbei zeigte sich, dass verbreitete theoretische Annahmen über die Dimensionalität und Extensionalität der Modellierkompetenz selten empirisch gestützt werden konnten.

---

Models and modeling are of grave importance to science and the teaching of science. While various purposes can be pursued with models and modeling, curricula worldwide focus on epistemic perspectives on models and modeling. To achieve these requirements, science teachers need sufficient modeling competence themselves, however, studies describing the modeling processes of science teachers are widely lacking. Therefore, the present study aimed to describe individual modeling strategies of 36 pre-service biology teachers empirically, by doing a case based and process-oriented analysis. Based on the occurrence, frequency and sequence of individual modeling activities, six typical modeling strategies were differentiated and described in detail. The results extend the formerly product-oriented state of research regarding models and modeling. Additionally, the dimensionality and extensionality of the theoretical construct of modeling competence was further explored, as relations between the identified modeling strategies and various variables were investigated. Hereby, it was shown that common theoretical assumptions regarding the dimensionality and extensionality of modeling competence may lack empirical validity.

## 9.2 Liste der aus dieser Dissertation hervorgegangenen Vorveröffentlichungen

Im Rahmen dieser Dissertation sind die folgenden Publikationen entstanden:

1. 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>
2. Göhner, M., & Krell, M. (2018). Modellierungsprozesse von Lehramtsstudierenden der Biologie. *Erkenntnisweg Biologiedidaktik*, 17, 45–63.
3. Göhner, M., & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52, 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
4. Göhner, M., & Krell, M. (2020). Was ist schwierig am Modellieren? Identifikation und Beschreibung von Problemen in Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer. *Zeitschrift für Didaktik der Naturwissenschaften*, 27, 155–180. <https://doi.org/10.1007/s40573-021-00131-4>
5. Göhner, M., & Krell, M. (2022). Analyzing the relationships between pre-service biology teachers' modeling processes, scientific reasoning competencies and general cognitive abilities. In M. Ergazaki & K. Kampourakis (Hrsg.), *Contributions from Biology Education Research* (S. 231–244). Heidelberg: Springer Nature.
6. Göhner, M., Bielik, T., & Krell, M. (2022). Investigating the dimensions of the modeling competence among pre-service science teachers: Meta-modeling knowledge, modeling practice and modeling product. *Journal of Research in Science Teaching*, 59(8), 1354–1387. <https://doi.org/10.1002/tea.21759>

Weiterhin wirkte der Autor an diesen Publikationen mit, die teilweise Ideen oder Daten der vorliegenden Untersuchung aufgreifen:

7. Ammonet, R., Göhner, M. F., Bielik, T., & Krell, M. (Manuscript submitted for publication). Why most Definitions of Modeling Competence in Science Education fall short: Analyzing the Relevance of Volition for Modeling. *Science Education*.
8. Göhner, M., & Krell, M. (2023). Umweltfaktoren beeinflussen das Libellenverhalten: Modelle als Werkzeuge zur Erkenntnisgewinnung nutzen. *Unterricht Biologie*, 481, 20–25.
9. Krell, M., Göhner, M., Bielik, T., Wesener, F., & Tietjen, B. (2023). Digitale Modellierung von Ökosystemen im Biologieunterricht. *MNU-Journal*. (4), 307–313.
10. Meister, S., Krell, M., Göhner, M., & Upmeier zu Belzen, A. (2021). Pre-service biology teachers' responses to first-hand anomalous data during modelling Processes. *Research in Science Education*, 51, 1459–1479. <https://doi.org/10.1007/s11165-020-09929-7>



# Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review

Maximilian Göhner<sup>1</sup> · Moritz Krell<sup>1</sup>

Eingegangen: 10. Januar 2020 / Angenommen: 29. Februar 2020 / Online publiziert: 16. März 2020  
© Der/die Autor(en) 2020

## Zusammenfassung

Im Rahmen ihrer Entwicklung und Etablierung als empirisch arbeitende Disziplinen findet in den Naturwissenschaftsdidaktiken vermehrt eine Auseinandersetzung mit methodischen und methodologischen Fragen statt. Hierzu gehört unter anderem die Frage danach, was *gute* fachdidaktische Forschung überhaupt ausmacht. Obwohl die qualitative Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung als ein etabliertes Verfahren gilt, unterscheiden sich Vorschläge hinsichtlich anzulegender Gütekriterien und umzusetzender Maßnahmen der Qualitätssicherung. Im Sinne einer exemplarischen Sammlung erprobter Forschungspraxis wird in diesem Beitrag ein Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung etablierten Gütekriterien und Maßnahmen zur Qualitätssicherung bei der Umsetzung von Verfahren der qualitativen Inhaltsanalyse gegeben. Hierzu werden 50 in der *Zeitschrift für Didaktik der Naturwissenschaften* publizierte Artikel, die eine qualitative Inhaltsanalyse umsetzen, mit einem deduktiv-induktiv erstellten Kategoriensystem untersucht. Es werden 11 Gütekriterien identifiziert, wobei vorwiegend die klassischen Gütekriterien Validität, Reliabilität und Objektivität in den untersuchten Artikeln beschrieben werden. Von 16 identifizierten Maßnahmen der Qualitätssicherung werden am häufigsten die Modi der Kategorienbildung und die Interrater-Übereinstimmung beschrieben. Es ergibt sich ein signifikant positiver Zusammenhang zwischen dem Publikationsjahr und der Zahl der durchschnittlich beschriebenen Maßnahmen zur Qualitätssicherung. Allerdings werden die einzelnen Maßnahmen der Qualitätssicherung, über alle Artikel hinweg betrachtet, nicht konsistent jeweils einem Gütekriterium zugeordnet. Die Ergebnisse sollen die Diskussion in den Fachdidaktiken anregen und zur weiteren Etablierung und Systematisierung des Verfahrens der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung beitragen.

**Schlüsselwörter** Qualitative Forschung · Qualitative Inhaltsanalyse · Gütekriterien · Objektivität · Reliabilität · Validität

---

✉ Maximilian Göhner  
maximilian.goehner@fu-berlin.de

<sup>1</sup> Didaktik der Biologie, Freie Universität Berlin,  
Schwendenerstraße 1, 14195 Berlin, Deutschland

## Qualitative Content Analysis in Science Education Research Under the Consideration of Quality Criteria: a Review

### Abstract

In the course of developing and establishing the field of science education, methodical and methodological questions have become increasingly important. This includes questioning what poses as *good* qualitative research. Although, qualitative content analysis is considered an established method in science education research, there are different ideas regarding which quality criteria and measures of quality control have to be taken into account. To give an exemplary collection of reliable research practice, the present article reviews the quality criteria and measures of quality control used in science education publications utilizing a qualitative content analysis. 50 publications of the German journal *Zeitschrift für Didaktik der Naturwissenschaften* were investigated with a coding scheme, that was deductively-inductively developed. 11 quality criteria were identified, with validity, reliability and objectivity being the most common. 16 measures of quality control were identified, whereby descriptions of the development process of the categories and interrater-agreement were most commonly described. Also, there is a significant positive correlation between the year of publication and the number of measures of quality control described. However, these measures of quality control are inconsistently assigned to quality criteria, when all publications are taken into account. These results may spark further discussion in the science education community and support utilizing the method of qualitative content analysis more systematically.

**Keywords** Qualitative research · Qualitative content analysis · Quality criteria · Objectivity · Reliability · Validity

### Einleitung

Die Fachdidaktiken haben sich in Deutschland in den letzten Jahrzehnten zu eigenständigen, empirisch arbeitenden wissenschaftlichen Disziplinen entwickelt (Bayrhuber et al. 2011; Leuders 2015). Das Spezifische der Fachdidaktiken liegt hierbei in ihrem Fachbezug, wobei sie methodisch mit benachbarten Disziplinen verbunden sind (z. B. empirische Bildungsforschung, Psychologie). Im Rahmen ihrer Entwicklung und Etablierung als empirisch arbeitende Disziplinen findet in den Naturwissenschaftsdidaktiken vermehrt eine Auseinandersetzung mit methodischen und methodologischen Fragen statt (Leuders 2015); was auch an der Publikation von Lehrbüchern und Sammelbänden zu Methoden in der naturwissenschaftsdidaktischen Forschung erkennbar ist (z. B. Groß et al. 2019; Sandmann und Schmiemann 2016; Krüger et al. 2014; Reinisch et al. in Vorbereitung). Hierzu gehört auch die Frage danach, was *gute* fachdidaktische Forschung überhaupt ausmacht. Aktuelle Reviews zeigen, dass es nicht nur Hinweise auf eine mangelnde Replizierbarkeit von Studien in der sozialwissenschaftlichen und erziehungswissenschaftlichen Forschung gibt (Camerer et al. 2018; Makel und Plucker 2014), sondern auch, dass insbesondere mit Bezug auf qualitative Arbeiten sowohl in den Erziehungswissenschaften (Grunenberg 2004) als auch in den Sozialwissenschaften (Barusch et al. 2011) die praktische Umsetzung qualitätssichernder Maßnahmen oft sehr unterschiedlich ausfällt. Da die hohe Vielfalt qualitativer Methoden (Howe und Eisenhart 1990) und die im Vergleich zu quantitativer Forschung große Bedeutsamkeit von interpretativer Sinnkonstruktion (Patton 2001) die Etablierung von allgemeingültigen Gütekriterien für die qualita-

tive fachdidaktische und erziehungswissenschaftliche Forschung erschweren (vgl. Steinke 2013; Flick 2010), werden methodenspezifisch anzulegende Gütekriterien und Maßnahmen diskutiert (vgl. „methodological awareness“; Seale 2002). Dies zeigt sich auch in einer Forderung des Roundtables der Deutschen Forschungsgemeinschaft (DFG) zur „Qualität qualitativer Forschung in der Erziehungswissenschaft“ (Helsper et al. 2001). Dort wurde die exemplarische Sammlung erprobter Forschungspraxen gefordert, um WissenschaftlerInnen eine Orientierung bei der Umsetzung von qualitätssichernden Maßnahmen in qualitativen Forschungsprojekten zu bieten.

Zu den in der qualitativen sozialwissenschaftlichen und erziehungswissenschaftlichen Forschung etablierten Methoden gehören beispielsweise die verschiedenen Verfahren der qualitativen Inhaltsanalyse (Schreier 2014; Mayring 2019). Auch in der naturwissenschaftsdidaktischen Forschung kommt der qualitativen Inhaltsanalyse eine besondere Bedeutung zu, die sich etwa in spezifisch naturwissenschaftsdidaktischen, methodologischen Beiträgen zeigt (z. B. Gropengießer 2005; Krüger und Riemeier 2014). Die vorliegende Analyse erarbeitet einen Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung etablierten Gütekriterien und Maßnahmen zur Qualitätssicherung bei der Umsetzung von Verfahren der qualitativen Inhaltsanalyse. Durch diese Sammlung erprobter Forschungspraxen (Helsper et al. 2001) trägt die Studie dazu bei, Verfahren der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung weiter zu etablieren und mögliche qualitätssichernde Maßnahmen zu systematisieren.

## Qualitative Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung

Unter dem Begriff *qualitative Inhaltsanalyse* werden verschiedene Varianten eines methodischen Verfahrens zusammengefasst, welches auf die „Systematisierung und Interpretation von manifesten und latenten Kommunikationsinhalten unter Verwendung eines Kategoriensystems“ abzielt (Stamann et al. 2016, S. 13). Traditionell wurden qualitativ inhaltsanalytische Verfahren zu den qualitativen Forschungsmethoden gezählt (z. B. Lamnek 2010). Aktuell verortet Mayring (2019) diese allerdings zwischen dem qualitativen und quantitativen Forschungsparadigma, da einzelne Varianten jeweils spezifische Nähe zu quantitativer oder qualitativer Methodologie aufweisen. Eine generelle Einordnung der qualitativen Inhaltsanalyse als qualitative Methode sei demnach nicht möglich (Mayring 2019). Schreier (2014) unterscheidet zwischen acht verschiedenen Varianten der qualitativen Inhaltsanalyse, wobei zwei Varianten von ihr besonders hervorgehoben werden, die sie als Basisformen der qualitativen Inhaltsanalyse bezeichnet. Die erste Basisform, die inhaltlich-strukturierende Inhaltsanalyse, identifiziert die für eine Fragestellung relevanten Aspekte aus vorliegendem Material (z. B. Texten), indem systematisch und iterativ ein Kategoriensystem entwickelt, erprobt und modifiziert wird (vgl. Kuckartz 2016; Mayring 2015; Schreier 2012). Hierbei zeigen aktuelle Diskussionen jedoch weiteren Abgrenzungsbedarf (Mayring 2019). So unterscheidet sich zum Beispiel auch innerhalb der Verfahren, die von Schreier (2014) zur Basisform der inhaltlich-strukturierenden qualitativen Inhaltsanalyse zugeordnet werden, das Verhältnis von deduktiv und induktiv abgeleiteten Kategorien. Dies kann je nach Forschungsgegenstand variieren (vgl. Kuckartz 2016; Schreier 2012; Mayring 2015), wobei Mayring (2015) die Bedeutung der deduktiven Kategorienbildung für die inhaltlich-strukturierende qualitative Inhaltsanalyse besonders betont. Die zweite Basisform, die qualitative Inhaltsanalyse durch Extraktion (vgl. Gläser und Laudel 2013), zielt auf den Einbezug des gesamten untersuchten Materials unter möglichst hoher ökonomischer Effizienz ab. Deduktiv entwickelte Kategorien dienen als Rahmen für die induktive Verortung aller in den Texten enthaltenen und durch Extraktion zusammengefassten Informationen. Zur weiteren Systematisierung der verschiedenen Varianten schlagen Stamann et al. (2016) vor, neben der Basisform (inhaltlich-strukturierend, extrahierend), der Art der Kategorien (formal, skalierend, inhaltlich) oder des Kategoriensystems (statisch, dynamisch, linear-hierarchisch, netzwerkartig) und den oben genannten Modi der Kategorienbildung (induktiv, deduktiv) auch die weitere Vorgehensweise mit den qualitativ-inhaltsanalytisch gewonnenen Daten (z. B. die Analyse von Kodierhäufigkeiten oder die Entwicklung einer Typologie) als Bezugspunkte zu betrachten.

Diese Vielfalt an Varianten der qualitativen Inhaltsanalyse illustriert, dass unterschiedliche methodologische Positionen zur qualitativen Inhaltsanalyse bestehen, die sich auch in aktuellen Diskussionen und Abgrenzungsversuchen wiederfinden (z. B. Mayring 2019).

Für die qualitative Inhaltsanalyse als etabliertes Verfahren der erziehungswissenschaftlichen und fachdidaktischen Forschung sind klar definierte Qualitätskriterien erforderlich, die für die Planung und Umsetzung von Forschungsvorhaben, für die Rezeption und Einordnung von Forschungsergebnissen sowie für die externe Begutachtung (z. B. *peer-review*, Forschungsförderung) leitend sein können (Helsper et al. 2016). Vergleichbar mit der Vielfalt qualitativer Forschungsmethoden im Allgemeinen (Howe und Eisenhart 1990), erschwert auch die Vielfalt der qualitativ-inhaltsanalytischen Verfahren im Besonderen die Antwort auf die Frage danach, was eine *gute* qualitative Inhaltsanalyse ausmacht und welche Gütekriterien an qualitativ-inhaltsanalytische Verfahren angelegt werden sollten. So argumentieren einige AutorInnen, dass die qualitative Inhaltsanalyse aus einer quantitativen Tradition hervorgeht und demnach auch klassische Gütekriterien wie Reliabilität und Validität bei deren Umsetzung angelegt werden sollten (vgl. Krippendorff 1980; Schreier 2012; Mayring 2015), wobei Objektivität meist auf Grund des hohen interpretativen Anteils der qualitativen Inhaltsanalyse mit Intersubjektivität gleichgesetzt (Mayring 2015) oder als Gütekriterium für die qualitative Inhaltsanalyse abgelehnt wird (Schreier 2012). Weitere Gütekriterien orientieren sich am Prozess des Verfahrens und können spezifische Maßnahmen zur Qualitätssicherung umfassen. So nennt Mayring (2002) beispielsweise die Verfahrensdokumentation, die argumentative Interpretationsabsicherung, die Regelgeleitetheit, die Nähe zum Gegenstand, die kommunikative Validierung und die Triangulation als Gütekriterien der inhaltlich-strukturierenden qualitativen Inhaltsanalyse. Etwas allgemeiner und orientiert an den zentralen Prinzipien der qualitativen Sozialforschung beschreibt Lamnek (2010) darüber hinaus Offenheit, Kommunikativität, Naturalistizität und Interpretativität als Merkmale einer guten qualitativen Inhaltsanalyse. Insgesamt wird damit deutlich, dass vielfältige Vorschläge zu Gütekriterien und entsprechenden Maßnahmen der Qualitätssicherung vorliegen. Diese Vielfältigkeit wird dem Anspruch klar definierter Qualitätskriterien (Helsper et al. 2001) allerdings nicht gerecht. Beispielsweise weist Mayring (2015) darauf hin, dass die oftmals als Inter-Koderreliabilität bezeichnete Beurteilerübereinstimmung ein Maß für die Auswertungsobjektivität ist, während Schreier (2012) die Beurteilerübereinstimmung entsprechend ihrer Bezeichnung der Reliabilität zuordnet.

## Ziel und Fragestellungen

Vor dem Hintergrund der oben dargestellten Relevanz qualitativ-inhaltsanalytischer Verfahren in der naturwissenschaftsdidaktischen Forschung ist es das Ziel dieser Analyse, einen Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung gängige Praxis bei der Beschreibung von Gütekriterien und qualitätssichernden Maßnahmen im Hinblick auf qualitativ-inhaltsanalytische Verfahren zu geben. Die vorliegende Analyse identifiziert häufig genannte Gütekriterien und qualitätssichernde Maßnahmen und stellt damit eine exemplarische Sammlung erprobter Forschungspraxis als Orientierung für Forschungsprojekte dar (Helsper et al. 2001). Ziel ist es ferner, damit die qualitative Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung weiter zu etablieren.

Die folgenden Fragen leiteten die vorliegende Analyse:

1. Welche Gütekriterien werden für die qualitative Inhaltsanalyse in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschrieben?
2. Welche Maßnahmen der Qualitätssicherung von qualitativen Inhaltsanalysen werden in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschrieben?
3. Inwieweit werden für die qualitative Inhaltsanalyse in der deutschsprachigen naturwissenschaftsdidaktischen Forschung einzelne Maßnahmen der Qualitätssicherung konsistent bestimmten Gütekriterien zugeordnet?

## Methoden

Um im Sinne eines systematischen Reviews (vgl. Bennett et al. 2005) die gängige Praxis der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschreiben zu können, wurden als Stichprobe alle bis zum 16.10.2018 publizierten Artikel der *Zeitschrift für Didaktik der Naturwissenschaften* mit dem Stichwort „Inhaltsanalyse“ im Fließtext ausgewählt<sup>1</sup>. Als Datengrundlage wurde die *Zeitschrift für Didaktik der Naturwissenschaften* gewählt, da sie als Verbandszeitschrift der Gesellschaft für Didaktik der Chemie und Physik (GDGP) und der Fachsektion Didaktik der Biologie (FDdB) im Verband Biologie, Biowissenschaften und Biomedizin (VBIO) als repräsentativ für deutschsprachige naturwissenschaftsdidaktische Forschung aufgefasst werden kann. Weitere Zeitschriften wurden aus forschungsökonomischen Gründen nicht in die Analyse miteinbezogen. Zudem wurde nicht weitergehend zwischen verschiedenen Varianten der qualitativen Inhaltsanalyse unter-

schieden (Schreier 2014), da in den identifizierten Artikeln größtenteils keine Angabe hierzu gemacht wurde.

58 der 256 insgesamt in diesem Zeitraum publizierten Artikel entsprachen dem Auswahlkriterium. Acht dieser 58 Artikel wurden nach der ersten Sichtung ausgeschlossen, da in ihnen die qualitative Inhaltsanalyse nicht als primäre Methode beschrieben, sondern zum Beispiel nur im Ausblick diskutiert wurde. Die 50 verbliebenen Artikel wurden anschließend qualitativ-inhaltsanalytisch untersucht (siehe separates Literaturverzeichnis; Anhang A). Hierbei wurde entsprechend der inhaltlich-strukturierenden Variante der qualitativen Inhaltsanalyse nach Schreier (2012) verfahren, um beschriebene Gütekriterien (Frage 1) und Maßnahmen zur Qualitätssicherung (Frage 2) in den Artikeln (Analyseeinheit) zu identifizieren, zu konzeptualisieren und die Artikel im Hinblick darauf zu beschreiben (vgl. Schreier 2014).

Ein Kategoriensystem bestehend aus Gütekriterien und Maßnahmen zur Qualitätssicherung wurde deduktiv-induktiv entwickelt. Als theoretische Grundlage für die deduktive Kategorienentwicklung wurden alle Texte herangezogen, die in den Methodenbeschreibungen der 50 analysierten Artikel im Zusammenhang mit der qualitativen Inhaltsanalyse referenziert werden. Das deduktive Vorgehen wurde gewählt, um die Beschreibung der Gütekriterien und Maßnahmen der Qualitätssicherung in der qualitativen Inhaltsanalyse in einzelnen Studien mit methodologischen Empfehlungen kontrastieren zu können. Da oft verschiedene Ausgaben des gleichen Textes referenziert wurden, wurden hier die älteste und die jüngste angegebene Auflage untersucht ( $N=21$ , siehe separates Literaturverzeichnis; Anhang B). Es wurden nur diejenigen Gütekriterien und Maßnahmen in das Kategoriensystem aufgenommen, die von den AutorInnen des jeweiligen Artikels explizit der qualitativen Inhaltsanalyse zugeordnet wurden. Weitere Gütekriterien und Maßnahmen der Qualitätssicherung, etwa die Datenerhebung betreffend, sind nicht Gegenstand der vorliegenden Analyse.

Die induktive Erweiterung des Kategoriensystems erfolgte auf der Basis der Kodierung von 50% der analysierten Artikel. Dabei wurden Abschnitte kodiert, die mindestens einen ganzen Satz umfassen (Kodiereinheit). Zur weiteren Einordnung von Beschreibungen wurde gegebenenfalls die referenzierte methodologische Literatur herangezogen (Kontexteinheit). Die Artikel wurden unabhängig durch zwei Personen ausgewertet (Erstauswertung: Erstautor des Artikels; Zweitauswertung: geschulte studentische Hilfskraft) und es wurde die Interrater-Übereinstimmung geprüft, wobei Cohens Kappa als zufallskorrigiertes Maß der Beurteilerübereinstimmung gewählt wurde (Cohens Kappa=0,70; „gut“, Wirtz und Caspar 2002). Nicht-übereinstimmungen wurden im Rahmen einer diskursiven Konsensfindung gelöst, dabei wurden gegebenenfalls wei-

<sup>1</sup> Der Zugriff erfolgte über <http://archiv.ipn.uni-kiel.de/zfdn/> und <https://link.springer.com/journal/40573>.

tere Kategorien hinzugefügt, und die bisherigen Kategorien um Definitionen, Kodierhinweise, Stichworte und Ankerbeispiele ergänzt.

Das sich hieraus ergebende Kategoriensystem wurde innerhalb der eigenen Arbeitsgruppe im Sinne eines Peer-Debriefings diskutiert (vgl. Steinke 2013), dabei wurde es insbesondere hinsichtlich seiner Vollständigkeit, der Struktur und den Kodierhinweisen bewertet (vgl. Lisch und Kriz 1978). Anschließend wurde das beschriebene Verfahren auf die übrigen 50 % der Artikel angewendet (Cohens Kappa=0,60; „gut“, Wirtz und Caspar 2002). Für das finale Kategoriensystem (Anhang C) ergaben sich so 27 Kategorien, wobei 11 Kategorien auf die Gütekriterien entfallen und 16 Kategorien Maßnahmen der Qualitätssicherung beschreiben (Tab. 1).

Alle Kodierungen wurden computergestützt mit dem Programm MAXQDA vorgenommen. Zur Beantwortung der Forschungsfragen 1 und 2 wurde jeweils eine Häufigkeitstabelle erstellt, die anzeigt in wie vielen Artikeln die einzelnen Gütekriterien (Frage 1) und qualitätssichernden Maßnahmen (Frage 2) beschrieben werden. Für die Gegenüberstellung von Gütekriterien und qualitätssichernden Maßnahmen (Frage 3) wurde eine Kreuztabelle erstellt, die aufzeigt, wie oft welche Gütekriterien mit welchen qualitätssichernden Maßnahmen zusammenhängend genannt werden.

## Ergebnisse

### Gütekriterien (Frage 1)

Insgesamt beschreiben 50 % der untersuchten Artikel mindestens eines der betrachteten Gütekriterien (Tab. 2). Validität wird in 15 Artikeln (30 %) beschrieben und stellt damit das am häufigsten beschriebene Gütekriterium dar. Hierbei kann Validität als Gütekriterium an das gesamte Verfahren der Auswertung (z. B. Gropengießer 1997), an das Kategoriensystem (z. B. Kühn 2011), an einzelne Kategorien (z. B. Kulgemeyer und Tomczyszyn 2015), an den Entwicklungsprozess der Kategorien (z. B. Mittelsten Scheid und Höble 2008), an die Untersuchungsergebnisse (z. B. Weitzel und Gropengießer 2009) oder an die Interpretation der Untersuchungsergebnisse angelegt werden (z. B. Kulgemeyer und Tomczyszyn 2015). Zusätzlich werden verschiedene Validitätsaspekte unterschieden, beispielsweise die interne und externe Validität (z. B. Brovelli 2014) oder die Inhaltsvalidität, Kriteriumsvalidität, Handlungsvalidität und prognostische Validität (z. B. Kulgemeyer und Tomczyszyn 2015).

In elf Artikeln (22 %) und somit jeweils ähnlich häufig wie die Validität werden die beiden anderen klassischen Gütekriterien Reliabilität und Objektivität beschrieben. Reliabilität wird hierbei ausschließlich im Zusammen-

**Tab. 1** Übersicht über die im Kategoriensystem enthaltenen Gütekriterien und Maßnahmen der Qualitätssicherung (das vollständige Kategoriensystem ist Anhang C zu entnehmen)

<b>Gütekriterien</b>	
Validität/Gültigkeit	
Reliabilität	
Objektivität/Intersubjektivität	
Transparenz	
Offenheit	
Nähe zum Gegenstand/Authentizität/Naturalizität	
Kommunikativität	
Interpretativität	
Praktische Relevanz	
Praktikabilität/Ökonomie	
Ethische Überlegungen	
<b>Maßnahmen der Qualitätssicherung</b>	
Auswahl/Standardisierung des Settings	
Dokumentation	
Argumentative Interpretationsabsicherung	
Systematisches Vorgehen	
Triangulation	
Kommunikative Validierung	
Theoretisches Sampling	
Formale Aspekte des Kategoriensystems	
Induktive Bildung des Kategoriensystems	
Deduktive Bildung des Kategoriensystems	
Training der BeurteilerInnen	
Diskussion des Kategoriensystems	
Diskussion der Kodierungen	
Konsensbildung	
Prüfung der Intrarater-Übereinstimmung	
Prüfung der Interrater-Übereinstimmung	

**Tab. 2** Nennung der Gütekriterien in den 50 ausgewählten Artikeln der ZfDN, mehrere Kodierungen des gleichen Gütekriteriums werden pro Artikel nur einmal gezählt

Gütekriterien	Artikelanzahl
Validität/Gültigkeit	15
Reliabilität	11
Objektivität/Intersubjektivität	11
Nähe zum Gegenstand/Authentizität/Naturalizität	5
Praktikabilität/Ökonomie	5
Offenheit	3
Praktische Relevanz	3
Transparenz	3
Ethische Überlegungen	2
Kommunikativität	2
Interpretativität	1

hang mit bestimmten qualitätssichernden Maßnahmen beschrieben (vgl. Tab. 4). In den betrachteten Artikeln treten Objektivität und Intersubjektivität nicht gemeinsam auf, beide Begriffe werden jedoch als Gütekriterium für das Kategoriensystem (z. B. Kodierungen anhand von objektiven Textmerkmalen; Starauschek 2006), einzelne Kategorien (z. B. Baur 2018) oder einzelne Kodierungen und davon abgeleitete Interpretationen (z. B. Nehring et al. 2016) verwendet. Schmelzing et al. (2010) nennen darüber hinaus die Auswertungsobjektivität als einen Teilaspekt der Objektivität.

Die Nähe zum Gegenstand (Mayring 2015), auch Authentizität (vgl. Brovelli et al. 2013) oder Naturalistizität genannt (vgl. Lamnek 2010), wird in 10 % der untersuchten Artikel beschrieben. Als Gütekriterium kann die Nähe zum Gegenstand dabei zum einen an die Datenaufnahme angelegt werden, zum Beispiel indem eine besonders authentische Situation geschaffen wird (z. B. Brovelli et al. 2013). Zum anderen wird die Nähe zum Gegenstand im Zusammenhang mit Kategoriensystemen beschrieben, in denen eine „möglichst authentische Erfassung des Bedeutungsgehaltes“ stattfindet (z. B. Kasper und Mikelskis 2008).

Praktikabilität oder ökonomische Überlegungen werden vor allem in Hinsicht auf formale Aspekte oder vorbereitende Schritte der qualitativen Inhaltsanalyse beschrieben. Solche Überlegungen umfassen zum Beispiel die Begründung des Einsatzes von Computerprogrammen (z. B. Reitschert und Höhle 2007), die Reduzierung des Umfangs von eingesetzten Instrumenten zur Datenaufnahme (z. B. Kulgemey-

er und Tomczyszyn 2015) sowie den Stichprobenumfang und die Stichprobenbeschaffenheit (z. B. Uhdn 2016). Beschreibungen der praktischen Relevanz, die im Zusammenhang mit der qualitativen Inhaltsanalyse getroffen werden, umfassen Implikationen für die schulische (Baur 2018) und universitäre Lehre (Kulgemeyer und Tomczyszyn 2015) sowie für die Gestaltung von Prüfungsanforderungen (Florian et al. 2015).

Die weiteren von Lamnek (2010) definierten Gütekriterien Offenheit, Kommunikativität und Interpretativität, die sich an den zentralen Prinzipien der qualitativen Sozialforschung orientieren, werden seltener beschrieben. So begründet Offenheit in drei Artikeln die Auswahl der qualitativen Inhaltsanalyse als Auswertungsmethode (Baalmann et al. 2004; Riemeier et al. 2010; Baur 2018), während Kommunikativität (z. B. Mittelsten Scheid und Höhle 2008) und Interpretativität (Baalmann et al. 2004) als Eigenschaften der qualitativen Inhaltsanalyse hervorgehoben, aber nicht weiter ausgeführt werden.

Ethische Überlegungen zum Einsatz der qualitativen Inhaltsanalyse werden in zwei Artikeln expliziert und heben die Rolle der ProbandInnen als Forschungspartner hervor (z. B. Gropengießer 1997; Uhdn 2016).

## Maßnahmen der Qualitätssicherung (Frage 2)

In 45 Artikeln (90 %) werden eine oder mehrere Maßnahmen der Qualitätssicherung beschrieben, dabei wird eine Zunahme der Anzahl beschriebener Maßnahmen mit

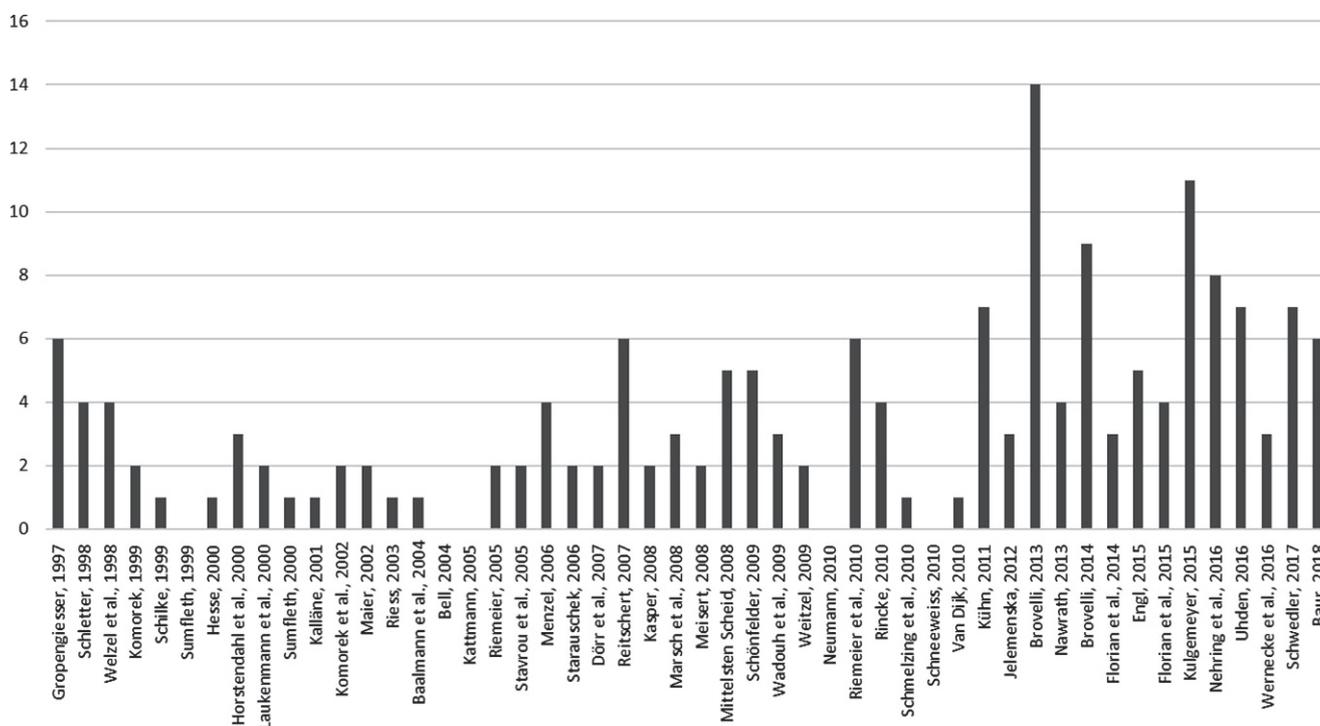


Abb. 1 Anzahl beschriebener Maßnahmen je untersuchtem Artikel

**Tab. 3** Beschreibung qualitätssichernder Maßnahmen in den 50 ausgewählten Artikeln der ZfDN, mehrere Kodierungen der gleichen Maßnahme werden pro Artikel nur einmal gezählt

Maßnahme der Qualitätssicherung	Artikelanzahl
Induktive Bildung des Kategoriensystems	29
Deduktive Bildung des Kategoriensystems	28
Prüfung der Interrater-Übereinstimmung	21
Dokumentation	12
Systematisches Vorgehen	12
Theoretisches Sampling	12
Kommunikative Validierung	10
Triangulation	10
Diskussion der Kodierungen	9
Diskussion des Kategoriensystems	9
Konsensbildung	7
Argumentative Interpretationsabsicherung	5
Formale Aspekte des Kategoriensystems	5
Training der BeurteilerInnen	4
Auswahl/Standardisierung des Settings	1
Prüfung der Intrarater-Übereinstimmung	0

der Zeit deutlich (Abb. 1), wobei vor allem Gropengießer (1997) und Brovelli et al. (2013) relativ umfangreich Maßnahmen entgegen diesem Trend beschreiben. Es ergibt sich ein signifikant positiver Zusammenhang (Pearson Korrelation) zwischen dem Publikationsjahr und der Zahl der durchschnittlich beschriebenen Maßnahmen zur Qualitätssicherung ( $N=50$ ;  $r=0,51$ ;  $p<0,001$ ; großer Effekt).

Die beschriebenen Maßnahmen der Qualitätssicherung (Tab. 3) beziehen sich am häufigsten auf den Modus der Kategorienbildung (vgl. Stamann et al. 2016). In acht Artikeln (16%) wird eine rein induktive Bildung des Kategoriensystems beschrieben. Sieben Artikel (14%) beschreiben eine rein deduktive Bildung des Kategoriensystems. Der Großteil der Artikel (21 Artikel, 42%) kombiniert eine induktive und deduktive Bildung des Kategoriensystems.

Die nächsthäufige Maßnahme der Qualitätssicherung der qualitativen Inhaltsanalyse in den untersuchten Artikeln ist die Prüfung der Interrater-Übereinstimmung (21 Artikel, 42%). Zu den hierbei berechneten Maßen gehören Cohens Kappa (z. B. Engl et al. 2015) sowie die prozentuale Beurteilerübereinstimmung (z. B. Staraschek 2006). Eine Prüfung der Intrarater-Übereinstimmung wird in keinem der untersuchten Artikel beschrieben.

Relativ häufig werden auch die von Mayring (2002) vorgeschlagenen qualitätssichernden Maßnahmen<sup>2</sup> Verfahrensdokumentation (12 Artikel, 24%), argumentative Interpretationsabsicherung (5 Artikel, 10%), systematisches Vorgehen (12 Artikel, 24%), kommunikative Validierung (10 Ar-

<sup>2</sup> Von Mayring (2002) als Gütekriterien der qualitativen Forschung bezeichnet.

tikel, 20%) und Triangulation (10 Artikel, 20%) beschrieben. Zu den expliziten Beschreibungen der Verfahrensdokumentation gehören die Entwicklungsprozesse von Kodierregeln (z. B. Reitschert und Höbke 2007) oder Kodiermanualen (z. B. Marsch et al. 2009) sowie die Begründung der Form der Datenaufnahme (z. B. Riemeier 2005) und deren Limitationen (z. B. Jelemenská 2012). Argumentative Interpretationsabsicherung wird von verschiedenen AutorInnen genannt, aber nicht näher ausgeführt (z. B. Schönfelder und Bögeholz 2009). Systematisches Vorgehen wird entweder als Grundvoraussetzung für die Datenauswertung (z. B. Mittelsten Scheid und Höbke 2008) beschrieben oder begründet Teile des Vorgehens wie die Durchführung einer Pilotstudie (z. B. Mittelsten Scheid und Höbke 2008) oder andere Maßnahmen der Qualitätssicherung wie eine kommunikative Validierung oder die Diskussion der Kodierungen (z. B. Riemeier et al. 2010). Kommunikative Validierungen und Triangulationen werden jeweils in zehn Artikeln (20%) beschrieben. Zur kommunikativen Validierung werden zumeist retrospektiv Interviews mit ProbandInnen genutzt (z. B. Kulgemeyer und Tomczyszyn 2015), zum Teil auch in Verbindung mit Struktur-Lege-Techniken (z. B. Jelemenská 2012). Zur Triangulation werden sowohl verschiedene Methoden der Datenerhebung und -auswertung (z. B. Gropengießer 1997) als auch Daten verschiedener Erhebungszeitpunkte genutzt (z. B. Schwedler 2017).

Hinsichtlich diskursiver Maßnahmen der Qualitätssicherung können die interne, zwischen den BeurteilerInnen stattfindende Diskussion der Kodierungen (Kulgemeyer und Tomczyszyn 2015) sowie die Diskussion des Kategoriensystems mit mehr oder weniger externen ExpertInnen (Brovelli et al. 2013) unterschieden werden. Beide Maßnahmen werden in jeweils neun Artikeln (18%) beschrieben. Weniger oft werden dagegen Prozesse der Konsensbildung (7 Artikel, 14%) oder das Training der BeurteilerInnen erläutert (4 Artikel, 8%). Auch formale Aspekte des Kategoriensystems (5 Artikel, 10%) wie beispielsweise eine vorgegebene Länge der Kodiereinheiten, sowie die Standardisierung des Settings (1 Artikel, 2%) werden selten beschrieben.

### Zuordnung von Gütekriterien und Maßnahmen der Qualitätssicherung (Frage 3)

Die 174 insgesamt beschriebenen Maßnahmen der Qualitätssicherung werden zu einem Großteil (74 Maßnahmen, 43%) explizit Gütekriterien zugeordnet, allerdings unterscheiden sich die zugeordneten Gütekriterien in den betrachteten Artikeln. Keine der betrachteten Maßnahmen wird konsistent über mehrere Artikel hinweg einem einzigen Gütekriterium konsistent zugeordnet (Tab. 4).

Die Interrater-Übereinstimmung wird der Validität (in 4 Artikeln), der Reliabilität (in 7 Artikeln) sowie der Ob-

**Tab. 4** Kreuztabelle der zusammenhängend beschriebenen Gütekriterien und qualitätssichernden Maßnahmen. Zeilen (Maßnahmen der Qualitätssicherung) und Spalten (Gütekriterien) sind absteigend nach ihrer Häufigkeit sortiert

	Validität/ Gültigkeit	Reliabilität	Objektivität/ Intersubjektivität	Nähe zum Gegenstand/ Authentizität/ Naturalistizität	Trans- parenz	Praktikabilität/ Ökonomie	Offenheit	Praktische Relevanz	Ethische Überlegun- gen	Kommunika- tivität	Interpreta- tivität	Summe
<i>Prüfung der Interrater- Übereinstimmung</i>	4	7	6	–	–	–	–	–	–	–	–	18
<i>Induktive Bildung des Kategoriensystems</i>	3	3	1	2	–	1	1	–	–	–	–	11
<i>Dokumentation</i>	3	1	2	2	2	–	–	–	–	–	–	10
<i>Systematisches Vorgehen</i>	5	–	1	2	–	1	–	–	–	–	–	9
<i>Triangulation</i>	7	1	–	1	–	–	–	–	–	–	–	9
<i>Diskussion der Kodierun- gen</i>	5	3	1	–	–	–	–	–	–	–	–	9
<i>Diskussion des Kategorien- systems</i>	5	–	1	2	–	–	–	–	–	–	–	8
<i>Kommunikative Validierung</i>	4	–	–	1	1	1	–	–	–	–	–	7
<i>Argumentative Interpretati- onsabsicherung</i>	4	–	–	1	–	–	–	–	–	–	–	5
<i>Deduktive Bildung des Kategoriensystems</i>	2	–	–	1	1	–	1	–	–	–	–	5
<i>Theoretisches Sampling</i>	2	–	–	–	–	1	–	–	–	–	–	3
<i>Training der BeurteilerIn- nen</i>	–	2	1	–	–	–	–	–	–	–	–	3
<i>Formale Aspekte des Kate- goriensystems</i>	–	–	1	1	–	–	–	–	–	–	–	2
<i>Konsensbildung</i>	1	–	1	–	–	–	–	–	–	–	–	2
<i>Auswahl/Standardisierung des Settings</i>	–	–	–	–	–	–	–	–	–	–	–	0
<i>Summe</i>	45	17	15	13	4	4	2	0	0	0	0	–

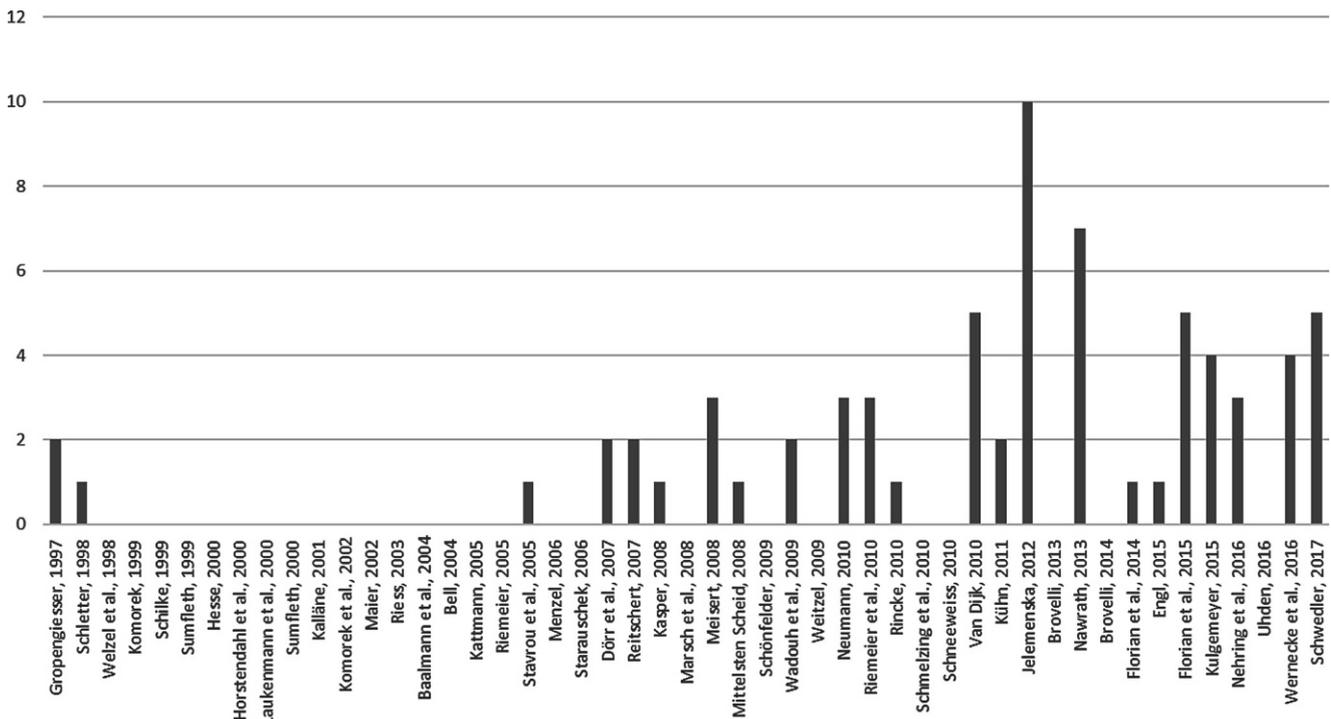


Abb. 2 Anzahl beschriebener Maßnahmen je untersuchtem Artikel, die einem Gütekriterium zugeordnet wurden

ektivität (in 6 Artikeln) zugeordnet. Dabei findet eine Zuordnung zu mehreren Gütekriterien teilweise auch innerhalb eines Artikels statt, zum Beispiel schreibt Baur (2018): „Fünfundzwanzig Prozent der Daten (Losprinzip) wurden bei der Kategoriensystementwicklung nochmals von einer zweiten Raterin ausgewertet, um die Objektivität und Reliabilität zu bestimmen.“ Auch die induktive Bildung des Kategoriensystems wird der Validität (in 3 Artikeln), der Reliabilität (in 7 Artikeln) und der Nähe zum Gegenstand (in 2 Artikeln), sowie in jeweils einem Artikel der Objektivität, Praktikabilität und Offenheit zugeordnet. Insgesamt werden die meisten beschriebenen Maßnahmen auch dem am häufigsten beschriebenen Gütekriterium der Validität zugeordnet.

Werden nur diejenigen Maßnahmen berücksichtigt, die explizit einem beliebigen Gütekriterium zugeordnet wurden (Abb. 2), ergibt sich ebenso ein signifikant positiver Zusammenhang zwischen dem Publikationsjahr und der Zahl der durchschnittlich beschriebenen und einem Gütekriterium zugeordneten Maßnahmen ( $N=50$ ;  $r=0,44$ ;  $p<0,01$ ; mittlerer Effekt).

## Diskussion

In der qualitativen Forschung werden Gütekriterien nach wie vor umfangreich diskutiert (z.B. Steinke 2013) und auch für die qualitative Inhaltsanalyse existieren vielfälti-

ge theoretische Vorschläge, Gütekriterien und Maßnahmen der Qualitätssicherung anzulegen (Mayring 2015; Schreier 2012; Lamnek 2010). Die vorliegende Untersuchung zeigt für die publizierten Artikel in der *Zeitschrift für Didaktik der Naturwissenschaften*, dass auch in der praktischen Umsetzung im Laufe der letzten Jahre vermehrt Gütekriterien und Maßnahmen der Qualitätssicherung für die qualitative Inhaltsanalyse beschrieben werden (Abb. 1). Dieser positive Trend könnte auf eine umfassendere Qualitätssicherung der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung hindeuten, die mit der zunehmenden Etablierung der Fachdidaktiken als eigenständige, empirisch arbeitende Disziplinen zusammenhängen kann (Bayrhuber et al. 2011; Leuders 2015). Als Gütekriterien werden dabei, entgegen der theoretischen Vielfalt, vorwiegend Validität, Reliabilität und Objektivität angelegt, die sonst klassischerweise der quantitativen Forschung zugeordnet werden (Tab. 2). Die dominierende Position der Gütekriterien Validität, Reliabilität und Objektivität könnte dabei Hinweis auf die zunehmende Etablierung methodologischer Positionen (Mayring 2015; Schreier 2012) oder eine vermehrte Nutzung integrativer Ansätze sein, die quantitative und qualitative Methoden zusammenführen (Kelle 2008). Die Dominanz der Validität als meistgenanntes Gütekriterium passt außerdem zu aktuellen Entwicklungen in der quantitativen Forschung. Hier wird ein „argument-based approach to validation“ vorgeschlagen (Kane 2013), nach dem durch Berücksichtigung verschiedener Evidenzquellen der argu-

mentative Nachweis erbracht werden soll, dass die intendierten Testwertinterpretationen legitim sind. Dabei werden vielfältige Argumente für die Validität der intendierten Interpretation eines Testwertes geprüft. Hierbei kann nicht von der Validität einer Methode oder eines Tests im Allgemeinen gesprochen werden, diese ist vielmehr in Abhängigkeit des gegebenen Untersuchungskontexts darzulegen (Hartig, Frey und Jude 2012).

Betrachtet man die verschiedenen identifizierten Maßnahmen der Qualitätssicherung ergibt sich auf den ersten Blick ein ebenfalls relativ einheitliches Bild: Die Modi der Kategorienbildung und die Interrater-Übereinstimmung stellen klar die am häufigsten genannten Maßnahmen dar (Tab. 3), was sich auch theoretisch wiederfindet, so weist beispielsweise Mayring (2015) der Interrater-Übereinstimmung eine besondere Bedeutung zu. Auf den zweiten Blick zeigen sich zwischen den Artikeln Unterschiede im Hinblick auf den Zusammenhang zwischen den beschriebenen Maßnahmen der Qualitätssicherung und ihrer Zuordnung zu den Gütekriterien. Hier existiert weder theoretisch (vgl. „Objektivität“ Mayring 2015; Schreier 2012) noch praktisch (Baur 2018; Kulgemeyer und Tomczyszyn 2015; Kühn 2011) ein methodologischer Konsens, der als erprobte Forschungspraxis aufgefasst werden kann (vgl. Helsper et al. 2001). Diese Unterschiede könnten in der qualitativen Inhaltsanalyse selbst begründet sein, die durch ihren Variantenreichtum immer wieder Begriffsbestimmungen und Systematisierungsversuchen unterliegt (Ramsenthaler 2013; Schreier 2014; Stamann et al. 2016). Allerdings findet sich hierfür in den untersuchten Artikeln kaum ein Hinweis, da bis auf den Modus der Kategorienbildung selten klar definiert oder benannt wird, welche Variante der qualitativen Inhaltsanalyse dem jeweiligen Forschungsprojekt zugrunde liegt. Weiterhin könnten diese Unterschiede in der Natur der qualitativen Forschungsprojekte liegen, wobei abgewogen werden muss, in welcher Detailliertheit Methoden, Gütekriterien und Maßnahmen der Qualitätssicherung überhaupt konkretisiert werden können, ohne dass die Offenheit und Flexibilität qualitativer Methoden zu Gunsten von Systematik und Transparenz verloren geht (Denzin 2016). Um eine Diskussion über die Zusammenhänge zwischen spezifischen Gütekriterien und Maßnahmen der Qualitätssicherung zu vermeiden, können letztere auch hier analog zur quantitativen Forschung im Sinne eines „argument-based approach to validation“ (Kane 2013) als Argumente für die Validität der abgeleiteten Interpretationen aufgefasst werden. So kann jede Maßnahme der Qualitätssicherung als eine Evidenzquelle für eine valide Interpretation der Untersuchungsergebnisse aufgefasst werden. Dies bedeutet aber auch, dass ein Zusammenspiel mehrerer Maßnahmen der Qualitätssicherung sowie deren umfassende qualitative Beschreibungen an Bedeutung gewinnen. Gerade in der Anzahl der be-

schriebenen Maßnahmen der Qualitätssicherung und deren qualitativer Beschreibung unterscheiden sich auch die hier untersuchten Artikel, wobei sich die Unterschiede mitunter aus den spezifischen Projekten heraus ergeben. So werden forschungsökonomisch begründet häufig nur Teile der Daten oder sogar nur das Kodiermanual (z.B. Schmelzing et al. 2010) auf Interrater-Übereinstimmung hin überprüft, obwohl kleine Werte der Beurteilerübereinstimmung auch als Indikator und Entscheidungshilfe für weitere qualitätssichernde Maßnahmen genutzt werden können (z.B. Kulgemeyer und Tomczyszyn 2015). Häufig schwer zu begründen sind darüber hinaus die Grenzwerte der Beurteilerübereinstimmung, die sich im Allgemeinen nach der Komplexität des vorliegenden Datenmaterials beziehungsweise des zu erfassenden Merkmals richten (Wirtz und Caspar 2002). Hier kann eine inhaltliche Begründung (z.B. in Nehring et al. 2016) für Nachvollziehbarkeit und Transparenz sorgen. Insgesamt werden Maßnahmen der Qualitätssicherung, die vor und nach der eigentlichen Kodierung stattfinden, selten und wenig ausführlich beschrieben, hierzu gehören besonders diskursive und daher komplex zu dokumentierende Prozesse wie das Training der BeurteilerInnen, das Vergleichen der Kodierungen und die Bildung eines finalen Konsenses. Hierbei spielt natürlich der zeitliche Aufwand eine große Rolle und sollte für jede qualitätssichernde Maßnahme abgewogen werden (Praktikabilität/Ökonomie). Beispielsweise erfordert die Überprüfung der Intrarater-Übereinstimmung, die in keinem der analysierten Artikel durchgeführt wurde (Tab. 3), einen zusätzlichen Kodierdurchgang. Von der naturwissenschaftsdidaktischen Community als besonders hochwertig eingeschätzte Forschungsprojekte können hier aufzeigen, wie detailliert methodenspezifische Gütekriterien beschrieben sein können, um eine flexible, systematische und transparente Evaluation zu ermöglichen.

Natürlich unterliegen die Ergebnisse des vorliegenden Artikels auch einigen Limitationen. So ist die qualitative Inhaltsanalyse nicht nur Gegenstand, sondern auch Methode der vorliegenden Analyse, was auf einer methodologischen Ebene paradox erscheint: Die qualitative Inhaltsanalyse als Forschungsmethode wird auf Gütekriterien und Maßnahmen der Qualitätssicherung hin untersucht, wobei methodisch selbst eine qualitative Inhaltsanalyse genutzt wird, die wiederum die Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigen soll. In der praktischen Umsetzung konnten so viele Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigt und reflektiert werden (vgl. Methode), wobei auch hier aus forschungsökonomischen Gründen auf die Überprüfung der Intrarater-Übereinstimmung verzichtet wurde. Darüber hinaus ergeben sich weitere methodische Einschränkungen. Die *Zeitschrift für Didaktik der Naturwissenschaften* wurde als Datengrundlage ausgewählt, unter der Annahme, dass diese als Ver-

bandszeitschrift der GDGP und FDdB im VBIO eine geeignete Zeitschrift für qualitativ hochwertige und repräsentative naturwissenschaftsdidaktische Forschung im deutschsprachigen Raum darstellt. Zur weiteren Generalisierung der hier beschriebenen Ergebnisse wäre eine Ausweitung auf weitere Zeitschriften nötig; zum Beispiel internationale Journals, Zeitschriften mit einem explizit qualitativen Fokus oder solche, die über die naturwissenschaftsdidaktische Forschung hinausgehen (z. B. *Journal of Research in Science Teaching*, *Zeitschrift für interpretative Schul- und Unterrichtsforschung*). Dabei könnten auch weitere Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigt werden und die vorliegenden Ergebnisse erweitern, zum Beispiel bezüglich der Datenerhebung. Zusätzlich zu bedenken sind natürlich auch formale und stilistische Einflüsse auf die Beschreibung der Gütekriterien und Maßnahmen der Qualitätssicherung. So ist es denkbar, dass in den jeweiligen Forschungsprojekten durchaus Gütekriterien angelegt und Maßnahmen der Qualitätssicherung durchgeführt worden sind, diese dann im publizierten Artikel aber nicht beschrieben wurden.

## Fazit

Die vorliegende Analyse gibt einen Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung etablierten Gütekriterien und Maßnahmen zur Qualitätssicherung bei der Umsetzung von Verfahren der qualitativen Inhaltsanalyse. Dabei können die hier beschriebenen Ergebnisse WissenschaftlerInnen bereits als Ansatzpunkt für Gütekriterien und Maßnahmen der Qualitätssicherung dienen, die an die eigene Forschung angelegt werden können. Zugleich sollen die vorliegenden Ergebnisse aber vor allem zu Diskussionen in den Fachdidaktiken anregen und so zur weiteren Etablierung und Systematisierung des Verfahrens der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung beitragen. Die Güte qualitativ inhaltsanalytischer Forschung kann im Sinne des „argument-based approach to validation“ (Kane 2013) im Einzelfall und unter Berücksichtigung der intendierten Interpretation von Forschungsergebnissen bewertet werden. Hierdurch zeigt sich die Umsetzung von Maßnahmen zur Qualitätssicherung nicht als Routineverfahren, für das allgemeingültige Regeln vorgeschrieben werden könnten (Hartig et al. 2012). Vielmehr sollte Gütesicherung als ein theoriegeleiteter und komplexer Prozess verstanden werden, bei dem in Abhängigkeit des Forschungskontexts begründet und entschieden werden muss, welche Maßnahmen zur Qualitätssicherung Evidenz für die intendierte Interpretation von Daten liefern können. Neben der Erarbeitung von Gütekriterien für die externe Begutachtung (vgl. Helsper et al. 2016) wird damit einmal mehr die Forderung nach ei-

ner umfassenden und gut zugänglichen Sammlung erprobter Forschungspraxen unterstützt (Helsper et al. 2001), die für WissenschaftlerInnen auch bei der Planung und Umsetzung von Forschungsvorhaben in den Fachdidaktiken leitend sein kann.

**Förderung** Die Autoren danken der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung des Projekts TypMoL (Projektnummer 327507949).

**Funding** Open Access funding provided by Projekt DEAL.

**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>.

## Anhang A

### Verzeichnis der betrachteten ZfDN-Artikel

- Baalman, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung – Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 7–28.
- Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 24.
- Bell, T. (2004). Komplexe Systeme und Strukturprinzipien der Selbstregulation – Konstruktion grafischer Darstellungen, Transfer und systemisches Denken. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 183–204.
- Brovelli, D. (2014). Integrierte naturwissenschaftliche Lehrerbildung – Entwicklung professioneller Kompetenz bei Lehramtsstudierenden. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 21–32.
- Brovelli, D., Bölsterli, K., Rehm, M., & Wilhelm, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettestest

- mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat. *Unterrichtswissenschaft*, 41(4), 306–329.
- Engl, L., Schumacher, S., Sitter, K., Größler, M., Niehaus, E., Rasch, R., . . . Risch, B. (2015). Entwicklung eines Messinstrumentes zur Erfassung der Protokollierfähigkeit – initiiert durch Video-Items. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 223–229.
  - Florian, C., Sandmann, A., & Schmiemann, P. (2014). Modellierung kognitiver Anforderungen schriftlicher Abituraufgaben im Fach Biologie. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 175–189.
  - Florian, C., Schmiemann, P., & Sandmann, A. (2015). Aufgaben im Zentralabitur Biologie – eine kategoriengestützte Analyse charakteristischer Aufgabenmerkmale schriftlicher Abituraufgaben. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 69–86.
  - Gropengießer, H. (1997). Schülervorstellungen zum Sehen. *Zeitschrift für Didaktik der Naturwissenschaften*, 3.
  - Heran-Dörr, E., Wiesner, H., & Kahlert, J. (2007). Schülerorientierung oder Orientierung an Schülervorstellungen? Wie Lehrkräfte vor und nach einer internetunterstützten Fortbildungsmaßnahme über physikbezogenen Sachunterricht denken. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 162–180.
  - Hesse, M. (2000). Erinnerungen an die Schulzeit – Ein Rückblick auf den erlebten Biologieunterricht junger Erwachsener. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 187–201.
  - Horstendahl, M., Fischer, H. E., & Rolf, R. (2000). Konzeptuelle und motivationale Aspekte der Handlungsregulation von Schülerinnen und Schülern im Experimentalunterricht der Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 7–25.
  - Jelemenská, P. (2012). Lehrervorstellungen zum Lehren und Lernen von Evolution – Eine Fallstudie zum fachdidaktischen Coaching. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 231–261.
  - Kalläne, B., & Schilke, K. (2001). Hier und anderswo – Eine Analyse der Lehrpläne für den Sachunterricht. *Zeitschrift für Didaktik der Naturwissenschaft*, 7(2001), 538.
  - Kasper, L., & Mikelskis, H. (2008). Lernen aus Dialogen und Geschichten im Physikunterricht – Ergebnisse einer Evaluationsstudie zum Thema Erdmagnetismus. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 7–25.
  - Kattmann, U. (2005). Lernen mit anthropomorphen Vorstellungen? – Ergebnisse von Untersuchungen zur Didaktischen Rekonstruktion in der Biologie. *Zeitschrift für Didaktik der Naturwissenschaften*, 11, 165–174.
  - Komorek, M. (1999). Eine Lernprozessstudie zum deterministischen Chaos. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(3).
  - Komorek, M., Wendorff, L., & Duit, R. (2002). Expertenbefragung zum Bildungswert der nichtlinearen Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 8(1), 33–51.
  - Kühn, S. M. (2011). Weiterentwicklung der Aufgabekultur im naturwissenschaftlichen Unterricht der gymnasialen Oberstufe und im Abitur. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 35–55.
  - Kulgemeyer, C., & Tomczyszyn, E. (2015). Physik erklären – Messung der Erklärens-fähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 111–126.
  - Laukenmann, M., Bleicher, M., Fuss, S., Gläser-Zikuda, M., Mayring, P., & Rhöneck, C. (2000). Eine Untersuchung zum Einfluss emotionaler Faktoren auf das Lernen im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 139–155.
  - Maier, U. (2002). Eine qualitative Interviewstudie zum Einfluss des Lehrerverhaltens auf Lernemotionen von Schülern im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 8, 85–102.
  - Marsch, S., Hartwig, C., & Krüger, D. (2009). Lehren und Lernen im Biologieunterricht: Ein Kategoriensystem zur Beurteilung konstruktivistisch orientierter Lernumgebungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 109–130.
  - Meisert, A. (2008). Vom Modellwissen zum Modellverständnis – Elemente einer umfassenden Modellkompetenz und deren Fundierung durch lernerseitige Kriterien zur Klassifikation von Modellen. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 243–261.
  - Menzel, S., & Bögeholz, S. (2006). Vorstellungen und Argumentationsstrukturen von SchülerInnen der elften Jahrgangsstufe zur Biodiversität, deren Gefährdung und Erhaltung. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 199–217.
  - Mittelsten Scheid, N., & Höble, C. (2008). Wie Schüler unter Verwendung syllogistischer Elemente argumentieren: Eine empirische Studie zu Niveaus von Argumentation im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 145–165.
  - Nawrath, D., & Komorek, M. (2013). Kontextorientierung aus Sicht von Physiklehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 233–257.
  - Nehring, A., Stiller, J., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Videostudie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 77–96.
  - Neumann, K., Viering, T., & Fischer, H. (2010). Die Entwicklung physikalischer Kompetenz am Beispiel des

- Energiekonzepts. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 285–298.
- Reitschert, K., & Höble, C. (2007). Wie Schüler ethisch bewerten: Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sek. I. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 125–143.
  - Riemeier, T. (2005). Schülervorstellungen von Zellen, Teilung und Wachstum. *Zeitschrift für Didaktik der Naturwissenschaften*, 11(1), 52–72.
  - Riemeier, T., Jankowski, M., Kersten, B., Pach, S., Raabe, I., Sundermeier, S., & Gropengießer, H. (2010). Wo das Blut fließt. Schülervorstellungen zu Blut, Herz und Kreislauf beim Menschen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 77–93.
  - Riess, W. (2003). Die Kluft zwischen Umweltwissen und Umwelthandeln als pädagogische Herausforderung – Entwicklung und Erprobung eines Prozessmodells zum Umwelthandeln in alltäglichen Anforderungssituationen. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 147–159.
  - Rincke, K. (2010). Alltagssprache, Fachsprache und ihre besonderen Bedeutungen für das Lernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16(2010), 235–260.
  - Schilke, K. (1999). Lernvoraussetzungen von Kindern zum Thema Dinosaurier. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(2), 3–14.
  - Schletter, J. C., & Bayrhuber, H. (1998). Lernen und Gedächtnis – Kompartimentalisierung von Schülervorstellungen und wissenschaftlichen Konzepten. *Zeitschrift für Didaktik der Naturwissenschaften*, 4, 3.
  - Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 189–207.
  - Schneeweiss, H., & Gropengießer, H. (2010). Schülerkonzepte zu Mikroben. *Zeitschrift für Didaktik der Naturwissenschaften*, 16.
  - Schönfelder, S., & Bögeholz, S. (2009). Bewertungskompetenz in der reflexiven Leitbildarbeit eines Umweltbildungszentrums – Ein Beitrag zur Professionalisierung des pädagogischen Personals. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 247–263.
  - Schwedler, S. (2017). Was überfordert Chemiestudierende zu Studienbeginn? *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 165–179.
  - Starauschek, E. (2006). Der Einfluss von Textkohäsion und gegenständlichen externen piktoralen Repräsentationen auf die Verständlichkeit von Texten zum Physiklernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 127–157.
  - Stavrou, D., Komorek, M., & Duit, R. (2005). Didaktische Rekonstruktion des Zusammenspiels von Zufall und Gesetzmäßigkeit in der nichtlinearen Dynamik. *Zeitschrift für Didaktik der Naturwissenschaften*, 11, 147–164.
  - Sumfleth, E., & Kleine, E. (1999). Analogien im Chemieunterricht – eine Fallstudie am Beispiel des „Balls der einsamen Herzen“. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(3), 39–56.
  - Sumfleth, E., & Telgenbüscher, L. (2000). Zum Einfluss von Bildmerkmalen und Fragen zum Bild beim Chemielernen mit Hilfe von Bildern – Beispiel Massenspektrometrie. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 59–78.
  - Uhdén, O. (2016). Verständnisprobleme von Schülerinnen und Schülern beim Verbinden von Physik und Mathematik. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 13–24.
  - Van Dijk, E. M., & Kattmann, U. (2010). Evolution im Unterricht: Eine Studie über fachdidaktisches Wissen von Lehrerinnen und Lehrern. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 7–21.
  - Wadouh, J., Sandmann, A., & Neuhaus, B. (2009). Vernetzung im Biologieunterricht – deskriptive Befunde einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 69–87.
  - Weitzel, H., & Gropengießer, H. (2009). Vorstellungsentwicklung zur stammesgeschichtlichen Anpassung: Wie man Lernhindernisse verstehen und förderliche Lernangebote machen kann. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 287–305.
  - Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., ... von Aufschnaiter, S. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden – Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, 4(1), 29–44.
  - Wernecke, U., Schwanewedel, J., Schütte, K., & Harms, U. (2016). Wie wird Energie im Biologieschulbuch dargestellt? – Entwicklung eines Kategoriensystems und exemplarische Anwendung auf eine Schulbuchreihe. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 215–229.

## Anhang B

### Verzeichnis der referenzierten methodologischen Literatur

- Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.

- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. (2. Aufl.) Berlin: Springer.
- Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research* 31, 659–671.
- Erickson, F. (1998). *Qualitative research methods for science education*. Dordrecht: Kluwer. 1155–1173.
- Fatke, R. (1997). Fallstudien in der Erziehungswissenschaft. In B. Friebertshäuser & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim: Juventa. 56–68.
- Flick, U., von Kardorff, E., & Steinke, I. (Hrsg.). (2012). *Qualitative Forschung: Ein Handbuch*. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.
- Früh, W. (2007). *Inhaltsanalyse: Theorie und Praxis* (6. Aufl.). Konstanz: UVK.
- Gläser, J., & Laudel, G. (2004). *Experteninterviews und qualitative Inhaltsanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gropengießer, H. (2001). *Didaktische Rekonstruktion des „Sehens“: Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung*. Oldenburg: Carl-von-Ossietzky-Universität.
- Gropengießer, H. (2003). *Wie man Vorstellungen der Lerner verstehen kann: Lebenswelten, Denkwelten, Sprechwelten. Beiträge zur Didaktischen Rekonstruktion*. (2. Aufl.). Oldenburg: Didaktisches Zentrum.
- Gropengießer, H. (2005). Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lernforschung. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (2. Aufl.). Weinheim: Beltz. 172–189.
- Haft, H. (1984). Inhaltsanalyse. In H. Haft & H. Kordes (Hrsg.), *Methoden der Erziehungs- und Bildungsforschung*. Stuttgart: Klett-Cotta. 411–418.
- Huber, G. L. (1989). Qualität versus Quantität in der Inhaltsanalyse. In W. Bos & C. Tarnai (Hrsg.), *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie*. Münster, New York: Waxmann. 32–47.
- Lamnek, S. (1989). *Qualitative Sozialforschung*. München: Psychologie Verlags Union.
- Lamnek, S. (2010). *Qualitative Sozialforschung*. (5. Aufl.). Weinheim: Beltz.
- Lisch, R., & Kriz, J. (1978). *Grundlagen und Modelle der Inhaltsanalyse*. Reinbek: Rowohlt.
- Mayring, P. (1990). *Einführung in die qualitative Sozialforschung*. München: Psychologie Verlags Union.
- Mayring, P. (1990). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Deutscher Studien Verlag.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Eds.), *Handbuch qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag. 601–613.
- Mayring, P., Gläser-Zikuda, M., & Ziegelbauer, S. (2005). Auswertung von Videoaufnahmen mit Hilfe der Qualitativen Inhaltsanalyse – ein Beispiel aus der Unterrichtsforschung. *Medien Pädagogik*.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. (2. Aufl.). Thousand Oaks, CA: SAGE.
- Reinhoffer, B. (2005). Lehrkräfte geben Auskunft über ihren Unterricht. Ein systematisierender Vorschlag zur deduktiven und induktiven Kategorienbildung in der Unterrichtsforschung. In P. Mayring & M. Gläser-Zikuda (Eds.), *Die Praxis der qualitativen Inhaltsanalyse* (2. Aufl.). Weinheim: Beltz Verlag. 123–141.
- Wang, H. A. (1998). Science textbook studies reanalysis: Teachers „friendly“ content analysis methods? Paper presented at the National Association for Research In Science Teaching, San Diego.

Tab. C.1 Vollständiges Kategoriensystem

Gütekriterien	Kodierhinweise	Definition	Stichworte	Ankerbeispiel
Validität/ Gültigkeit	Validität/Gültigkeit wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Gültigkeit beinhaltet die Übereinstimmung von Ergebnissen mit dem durch die Untersuchung (Messung) vorgegebenen, theoretisch-begrifflich zu erfassenden.“ (Lisch und Kriz 1978, S. 84) <sup>a</sup>	Validität Gültigkeit Glaubwürdigkeit Validitätsformen (z. B. Augenschein-, Inhalts-, Konstruktvalidität, ...)	„Um eine Aussage zur Güte der Analyse machen zu können, muss abgesichert werden, dass die identifizierten Probleme adäquat analysiert werden, d. h. die Angemessenheit der Bedeutungsrekonstruktion des Materials muss durch eine Überprüfung verschiedener Expertenurteile validiert werden.“ (Uhden 2016, S. 17)
Reliabilität	Reliabilität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Zuverlässigkeit beinhaltet die Reproduzierbarkeit von Ergebnissen unter den gleichen intersubjektiven Bedingungen – insbesondere also die Forderung, dass andere Forscher bei Anwendung desselben Erhebungsinstruments in Interaktion mit demselben Untersuchungsgegenstand zu demselben Ergebnis gelangen.“ (Lisch und Kriz 1978, S. 84)	Reliabilität Zuverlässigkeit Verlässlichkeit Trennschärfe	„Fünfundzwanzig Prozent der Daten (Losprinzip) wurden bei der Kategoriensystementwicklung nochmals von einer zweiten Raterin ausgewertet, um die Objektivität und Reliabilität zu bestimmen.“ (Baur 2018, S. 6)
Objektivität/ Intersubjektivität	Objektivität/Intersubjektivität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Die Methode soll vom analysierenden Subjekt abgelöst werden, d. h. die Ergebnisse müssen intersubjektiv nachvollziehbar und damit auch reproduzierbar, kommunizierbar und kritisierbar sein.“ (Früh 2007, S. 42)	Objektivität Intersubjektivität Reproduzierbarkeit	„Schließlich wurde die Zuordnung der Antworten zu den von uns formulierten Kategorien durch einen Intersubjektivitätstest innerhalb der Arbeitsgruppe überprüft.“ (Welzel et al. 1998, S. 33)
Transparenz	Transparenz wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Die systematische, offengelegte und damit kritisierbare Vorgehensweise der empirischen Wissenschaft verlangt [die transparente Offenlegung] der Vorstellungen des Forschers, als auch de[s] ansivierete[n] Realitätsausschnitt[s]“ (Früh 2007, S. 20)	Transparenz Offenlegung Dokumentation Nachvollziehbarkeit	„Umstände und Verfahren der Erhebung, der Aufbereitung des Materials und der Auswertung werden nachvollziehbar und detailliert dokumentiert.“ (Gropengießer 1997, S. 76)
Offenheit	Offenheit wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Ggf. wird der Umgang mit unerwarteten Ergebnissen explizit beschrieben	„Qualitative Inhaltsanalyse versucht insoweit diese Offenheit zu praktizieren, als kein vorab entwickeltes inhaltsanalytisches Schema mit Analyseeinheiten, -dimensionen und -kategorien auf die zu untersuchenden Kommunikationsinhalte angelegt wird. Vielmehr versucht sie, den Inhalt selbst sprechen zu lassen und daraus die Analyse zu entfallen.“ (Lamnek 2010, S. 462)	Offenheit Interaktivität	„Dieses Ziel setzt eine subjektbezogene Datenerhebung mit einem offenen und interaktiven Verfahren voraus, wie es mit qualitativen Untersuchungsmethoden vorliegt.“ (Riemeier et al. 2010, S. 79)
Nähe zum Gegenstand/ Authentizität/ Naturalistizität	Die Nähe zum Gegenstand wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Das Untersuchungsfeld der qualitativen Sozialforschung „ist die natürliche Welt, die mit naturalistischen Methoden erfasst und beschrieben werden soll.“ (Lamnek 1989, S. 196)	Nähe zum Gegenstand Authentizität Naturalistizität Berücksichtigung natürlicher Gegebenheiten Vollständigkeit	„In einem zweiten Schritt wurden die Vignetten Lehrkräften und Fachdidaktikern vorgelegt und argumentativ im Blick auf die authentische Abbildung von Unterrichtssituationen validiert.“ (Brovelli 2014, S. 27)
Praktikabilität/ Ökonomie	Praktikabilität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Es werden ggf. praktische/ökonomische Vorgaben oder Einschränkungen und ihre Auswirkungen auf die qualitative Inhaltsanalyse beschrieben	„Der pragmatische Sinn jeder Inhaltsanalyse besteht letztlich darin, unter einer bestimmten forschungsleitenden Perspektive Komplexität zu reduzieren. Textmengen werden hinsichtlich theoretisch interessierender Merkmale klassifizierend beschrieben.“ (Früh 2007, S. 44)	Praktikabilität Ökonomie Zeitintensivität Finanzielle Überlegungen	„Die Festlegung der Stichprobengröße ergab sich aus dem zeitintensiven Untersuchungsdesign.“ (Baur 2018, S. 7)
Praktische Relevanz	Die praktische Relevanz wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Die Ergebnisse der qualitativen Inhaltsanalyse werden ggf. als praxisrelevant beschrieben, sodass TeilnehmerInnen oder Öffentlichkeit davon profitieren können	„Even if a study’s findings are „valid“ and transferable, we still need to know what the study does for its participants -both researchers and researched- and for its consumers.“ (Miles und Huberman 1994, S. 280)	Praktische Relevanz Berufspraxis Externe Wirkung	„Die dargestellten Fehler, Fehlkonzepte und schülerspezifischen Vorgehensweisen können für die Schulpraxis einen wichtigen Ausgangspunkt darstellen.“ (Baur 2018, S. 11)
Ethische Überlegungen	Ethische Überlegungen werden im Zusammenhang mit der qualitativen Inhaltsanalyse beschrieben. Hierzu gehört bspw. die Wertschätzung der TeilnehmerInnen	„Researchers are obliged ethically to anticipate what will be done in data collection, analysis, and reporting, and to explain to those studied why it will be done that way rather than some other.“ (Erickson 1998, S. 1456)	Ethische Überlegungen Vertrauenswürdigkeit Glaubwürdigkeit Fairness Zumutbarkeit	„Die Technik der interaktiven Tafeln erlaubt es, Tafelvideos synchron zu der Tonaufnahme aufzuzeichnen, so dass die Schüler nicht einer Beobachtungssituation mittels Videokamera ausgesetzt werden mussten. Als Motivation und Gegenleistung für die freiwillige Teilnahme wurde den Schülern ein Gutschein über zwei Nachhilfestunden [...] ausgehändigt, wodurch eine angemessene Gegenleistung geliefert wurde und zudem auch schwächere Schüler zur Teilnahme bewegt werden konnten.“ (Uhden 2016, S. 16)

Tab. C.1 (Fortsetzung)

Kommunikativität	Kommunikativität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Der Akt der Kommunikation wird als zentral dargestellt und der Güte der qualitativen Inhaltsanalyse zugeordnet	„[...] kommunikative] Verständigung wird bei der Interpretation von Kommunikationsinhalten durch qualitative Inhaltsanalyse besonders wichtig. Um die Realität der jeweiligen Situationen zu erfassen, muss auch die empirische Sozialforschung einen kommunikativen Akt initiieren, der der Alltagskommunikation ähnelt“ (Lamnek 1989, S. 195)	Kommunikativität	„Dieses Ziel setzt eine subjektbezogene Datenerhebung mit einem offenen und interaktiven Verfahren voraus, wie es mit qualitativen Untersuchungsmethoden vorliegt.“ (Riemeier 2010, S. 79)
Interpretativität	Interpretativität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Es werden Überlegungen oder Methoden dazu geäußert, wie Äußerungen von ProbandInnen interpretiert werden können	„eine wissenschaftlich modifizierte Form des alltagsweltlichen Fremdverstehens [...], um aus der naturalistischen, quasi-alltagsweltlichen Untersuchungssituation Handlungsmuster herauszufiltern. Dieser Vorgang besteht aus zwei Phasen: (1) Nachvollzug der alltagsweltlichen Deutungen und Bedeutungszuweisungen und (2) typisierende Konstruktion eines Musters.“ (Lamnek 2010, S. 464)	Interpretativität	„Um individuelle Bedeutungsstrukturen zu erheben, ist es notwendig, in einem offenen, interaktiven und problemzentrierten Verfahren interpretativ die Komplexität dieser Bedeutungsstrukturen zu erfassen.“ (Baalmann et al. 2004, S. 10)
<b>Maßnahmen der Qualitätssicherung</b>	<b>Kodierhinweise</b>	<b>Definition</b>	<b>Stichworte</b>	<b>Ankerbeispiel</b>
Auswahl/ Standardisierung des Settings	Eine Form der Standardisierung des Settings wird beschrieben und explizit erläutert, welche Relevanz dies für die qualitative Inhaltsanalyse hat. Eine Beschreibung des Settings oder standardisierender Schritte der Datenaufnahme reicht nicht aus	„[man versucht] im subjektiven, inneren Erleben der Befragten vergleichbare Situationen zu erzeugen, indem sich Interviewer, Beobachter usw. individuell auf die untersuchten Personen einstellen. [...] Dahinter steht die Überlegung, dass man unterschiedlichen Probanden Fragestellungen auch unterschiedlich präsentieren muss, um ihnen zu einem vergleichbaren Verständnis der Fragestellung zu verhelfen.“ (Bortz und Döring 1995, S. 302)	Setting Standardisierung gleiche Bedingungen Untersuchungskontext	„da die Adressaten wegen der Notwendigkeit einer Standardisierung der Erhebungsumgebung zum Zwecke der Testentwicklung vorher geschult werden mussten.“ (Kulgemeyer 2015, S. 124)
Dokumentation	Die Dokumentation der qualitativen Inhaltsanalyse wird auf Metaebene beschrieben und begründet. (z. B. Warum wurde etwas besonders ausführlich/nicht dokumentiert, wie wurden Entscheidungsprozesse dokumentiert?)	„Als eine Möglichkeit dazu erscheint die Forderung, bei Benutzung quantitativer Verfahren die Bedeutungsdimensionen der Daten qualitativ zu erschließen und offenzulegen, bei Benutzung qualitativer Verfahren die Interpretationsprozesse zu systematisieren und zu dokumentieren“ (Huber 1989, S. 41)	Dokumentation Kodiermanual Datensicherung Schutz der Identität von ProbandInnen	„Die Datenaufnahme erfolgte mithilfe digitaler Videotechnik, um eine eindeutige Sprecheridentifikation zu gewährleisten und über die verbale Sprache hinaus nonverbale Kommunikationselemente wie Gestik und Mimik dokumentieren zu können.“ (Riemeier 2005, S. 46)
Argumentative Interpretationsabsicherung	Es wird beschrieben, dass die Interpretation der Daten einem logisch argumentativen Vorgehen folgt	„Die Interpretationen werden argumentativ abgeleitet. Dies geschieht einerseits in der Explikation, aber auch in den Strukturierungen.“ (Gropengießer 2001, S. 132)	Argumentation Interpretation Regel Kodierhinweis	„In einem zweiten Schritt wurden die Vignetten Lehrkräften und Fachdidaktikern vorgelegt und argumentativ im Blick auf die authentische Abbildung von Unterrichtssituationen validiert.“ (Brovelli 2014, S. 27)
Systematisches Vorgehen	Das systematische Vorgehen der eigenen Studie wird reflektiert (Metaebene) und begründet (Vor-/Nachteile, Gütekriterien)	„Regelgeleitetheit: Das Material wird in Analyseeinheiten zerlegt und schrittweise bearbeitet, einem Ablaufmodell folgend.“ (Gropengießer 2005, S. 10)	Systematisches Vorgehen Pilotstudie mehrfaches Kodieren regelgeleitetes Vorgehen	„Dazu wurde zunächst ein Teil des Materials gesichtet, zusammengefasst und ein Kategoriensystem induktiv abgeleitet. [...] Es] wurden bereits in der Theorie bekannte Kategorien [...] deduktiv formuliert, dann aber induktiv in größere Kategorien integriert [...]. Anschließend wurde der Rest des Materials regelgeleitet kodiert.“ (Schwedler 2017, S. 170)
Triangulation	Es wird beschrieben, dass die vorliegenden Untersuchungsergebnisse mit anderen Ergebnissen verglichen wurden. Hierbei können andere Methoden auf dieselben Daten oder dieselben Methoden auf andere Daten angewendet werden. Außerdem können Vergleiche mit Untersuchungsergebnissen stattfinden, denen dieselben Annahmen oder Theorien zugrunde liegen	„In der Sozialforschung wird mit dem Begriff <i>Triangulation</i> die Betrachtung eines Forschungsgegenstandes von (mindestens) zwei Punkten bezeichnet.“ (Flick et al. 2012, S. 309)	Triangulation Triangulationsformen (z. B. Datentriangulation, Investigatortriangulation, Theorientriangulation, ...)	„Zum „Erfassen der Lernervorstellungen“ wurden drei Untersuchungen durchgeführt: [...] Erhebung des Stands der Forschung zu empirischen Untersuchungen im Themenbereich Blut und Blutkreislauf und eine anschließende Reinterpretation dieser Ergebnisse, zweitens leitfadensstrukturierte Interviews und drittens Interviewphasen von Vermittlungsexperimenten.“ (Riemeier et al. 2010, S. 79)
Kommunikative Validierung	Es wird beschrieben, dass StudienteilnehmerInnen in die Interpretation miteinbezogen werden (z. B. durch Follow-Up Interviews oder andere Formen der reaktiven Interaktion). Eine Kommunikation zwischen den Auswertenden ist nicht zu kodieren (vgl. Kodierkonferenz)	„Another response to the criticism that traditional qualitative research invites abuses of authority by researchers (including those of self-deception in data identification and analysis) has been for researchers to try to redress the imbalance of power by sharing it more fully with those who are studied. Both participatory action research and practitioner research are attempts to address the power/knowledge issues involved in social research [...].“ (Erickson 1998, S. 1468)	Kommunikative Validierung Interaktives Erfragen Bestätigung von Interpretationen Validierungsinterview	„Die entstandenen Strukturlegebilder wurden am nächsten Tag den Kindern zum Zwecke einer kommunikativen Validierung vorgelegt und erläutert. Nahmen die Kinder Korrekturen oder Ergänzungen vor, so wurden diese in der Qualitativen Inhaltsanalyse berücksichtigt.“ (Riess 2003, S. 154)

Tab. C.1 (Fortsetzung)

Theoretisches Sampling	Es wird eine gezielte Auswahl von StudienteilnehmerInnen oder Datenquellen beschrieben und begründet, inwieweit diese Auswahl die Qualität der qualitativen Inhaltsanalyse erhöht (z. B. durch die Erhöhung der Wahrscheinlichkeit einer theoretischen Sättigung)	„Die Auswahl der zu untersuchenden Fälle wird hier nicht nach dem Zufallsprinzip, sondern theoriegeleitet gezielt vom Forscher selbst getroffen (theoretisch-systematische Auswahl, bewußte Auswahl, theoretische Stichprobe).“ (Bortz und Döring 1995, S. 310)	Sampling Theoretisches Sampling Datenauswahl	„Um vorunterrichtliche Lernervorstellungen untersuchen zu können, wurde die Stichprobe in allen hier dargestellten Untersuchungen so gewählt, dass die Probanden noch keinen Unterricht zum Thema Blut und Blutkreislauf hatten.“ (Riemeier et al. 2010, S. 79)
Formale Aspekte des Kategoriensystems	Es werden formale Aspekte oder Voraussetzungen des Kategoriensystems beschrieben, hierzu gehören bspw. Trennschärfe, Anzahl oder Beschaffenheit der Kategorien	„[Die] Kodiereinheit legt fest, welches der kleinste Materialbestandteil ist, der ausgewertet werden darf, was der minimale Textteil ist, der unter eine Kategorie fallen kann.“ (Lamnek 1989, S. 213)	Formale Aspekte des Kategoriensystems Analyseeinheit Kodiereinheit Kontexteinheit Kategorienanzahl Kategorienlänge Trennschärfe	„Die Bestimmung der Länge von Events für die Kodierungen wurde über die Struktur von Unterrichtsaktivitäten vorgenommen.“ (Nehring et al. 2016, S. 87)
Induktive Bildung des Kategoriensystems	Es wird beschrieben, dass Kategorien aus den aufgenommenen Daten heraus gebildet werden. Auch eine später im Prozess auftretende datenbasierte Erweiterung oder Ausschärfung des Kategoriensystems wird kodiert	„Induktive Kategorienbildung hat als Grundgedanken, dass die Verfahrenswesen zusammenfassender Inhaltsanalyse genutzt werden, um schrittweise Kategorien aus einem Material zu entwickeln.“ (Flick et al. 2012, S. 472)	Induktiv Datengestützt Ausgehend von Daten	„Aus diesen Antworten haben wir fünf Haupt- und 33 Unterkategorien von Zielen für das Experimentieren gebildet.“ (Welzel et al. 1998, S. 32)
Deduktive Bildung des Kategoriensystems	Es wird beschrieben, dass Kategorien aus Vorannahmen oder Theorien gebildet werden. Auch eine später im Prozess stattfindende theoriebasierte Umstrukturierung oder Zusammenfassung des Kategoriensystems wird kodiert	„Die aus den theoretischen Vorüberlegungen abgeleiteten Variablen bilden unmittelbar die Grundlage für die Auswertungskategorien. Dadurch wird der Zusammenhang zum existierenden Wissen über den Untersuchungsgegenstand und zu den die Untersuchung strukturierenden theoretischen Vorüberlegungen hergestellt.“ (Gläser und Laudel 2004, S. 198)	Deduktiv Theoriegeleitet Literaturbasiert	„das Datenmaterial wurde nach einem im Vorfeld gebildeten Kategoriensystem strukturiert [...] Die Auswertekategorien sind aufgrund biologiedidaktischer und fachwissenschaftlicher Überlegungen aufgestellt worden.“ (Hesse 2000, S. 190)
Training der BeurteilerInnen	Es wird beschrieben, dass die BeurteilerInnen geschult werden	„Die Frage ist nun, wann „steht XY sinngemäß da“ und wie weit darf der Codierer die Textinformationen interpretieren? Diese Kompetenz wird bei der Codierschulung geübt und später im Reliabilitätstest überprüft.“ (Früh 2007, S. 52)	Training Schulung	„Zur Sicherung der Qualität des Kategoriensystems wurden verschiedene Maßnahmen durchgeführt, dazu gehörte insbesondere ein intensives, mehrstufiges Training der jeweils zwei fachspezifischen Beurteiler“ (Kühn 2011, S. 42)
Diskussion des Kategoriensystems durch Experten	Es wird beschrieben, dass mehrere Personen das Kategoriensystem (oder mehrere Kategoriensysteme) diskutieren	„Dieses systematische Vorgehen wird ergänzt und abgesichert mit der Kontrolle der Auswertung durch eine andere Person. Zusätzlich begleiteten die Mitglieder der Arbeitsgruppe den Prozess kritisch [...]“ (Gropengießer 2001, S. 132)	Diskussion des Kategoriensystems (Experten-)Diskussion	„Mit diesem Ziel wurde die Existenz und die zugehörige analysierende Bezeichnung und Beschreibung der Verständnisprobleme zu mehreren Zeitpunkten der Analyse mit ein oder zwei Physikdidaktikern diskutiert.“ (Uhlen 2016, S. 17)
Diskussion der Kodierungen	Es wird beschrieben, dass mehrere BeurteilerInnen ihre Kodierungen im Einzelfall diskutieren	„Das „klassische“ methodologische Kriterium der Objektivität wird in der Inhaltsanalyse durch Diskursfähigkeit der Ergebnisse gesichert.“ (Haft 1984, S. 418)	Diskussion der Kodierungen (Kodier-)Konferenz Ergebnisdiskussion	„Um eine objektive Vorgehensweise gewährleisten zu können [...], wurde die Auswertung von drei Personen zunächst unabhängig durchgeführt und anschließend gemeinsam diskutiert.“ (Schönfelder 2009, S. 254)
Konsensbildung	Es wird beschrieben, dass die BeurteilerInnen eine gemeinsame Kodierung erarbeiten	„Konsens [...] kann im Verlaufe fachlicher Diskussionen erzielt werden. Eine Konsensbildung in einem heterogenen Forscherteam ist ein stärkeres Indiz für Validität.“ (Bortz und Döring 1995, S. 310)	Konsensbildung Abgleich	„Die Kodierungen und Interpretationen wurden von mindestens zwei Auswertern vorgenommen und Differenzen bis zum Konsens diskutiert.“ (Marsch et al. 2009, S. 116)
Intrarater-Übereinstimmung	Es wird beschrieben, dass die Intrarater-Übereinstimmung in Form eines gängigen Maßes (bspw. Cohens Kappa) berechnet wurde. Ein Vergleich der Interpretationen allein reicht nicht aus (vgl. Diskussion der Kodierungen)	„Die Intra-Koderreliabilität wird überprüft, indem nach Abschluss der Analyse zumindest Teile des Materials erneut durchgearbeitet werden, ohne auf die zuerst erfolgten Kodierungen zu sehen.“ (Mayring 2010, S. 603)	Intrarater-Übereinstimmung Intrarater-Reliabilität Zweitkodierung mit zeitlichem Abstand	<i>Kein Beispiel vorhanden</i>
Interrater-Übereinstimmung	Es wird beschrieben, dass die Interrater-Übereinstimmung in Form eines gängigen Maßes (bspw. Cohens Kappa) berechnet wurde. Ein Vergleich der Interpretationen allein reicht nicht aus (vgl. Diskussion der Kodierungen)	„Die Inter-Koderreliabilität [...] wird überprüft, indem zumindest ein Ausschnitt des Materials einem zweiten Kodierer [...] vorgelegt wird.“ (Mayring 2010, S. 604)	Interrater-Übereinstimmung Interrater-Reliabilität Übereinstimmung mehrerer KodiererInnen	„die von dem zweiten Rater hinsichtlich der Existenz eines Energiebezugs bewertet wurden. Die Berechnung des Kappa-Koeffizienten als Maß für die Interraterreliabilität ergab $\kappa = 0,726$ .“ (Wernecke et al. 2016)

<sup>3</sup>Die hier aufgeführte Definition von Validität ist abzugrenzen vom im Artikel verwendeten *argument-based approach to validation* nach Kane (2013)

## Literatur

- Baalmann, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung – Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 7–28.
- Barusch, A., Gringeri, C., & George, M. (2011). Rigor in qualitative social work research: a review of strategies used in published articles. *Social Work Research*, 35(1), 11–19.
- Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 24, 115–129.
- Bayrhuber, H., Harms, U., Muszynski, B., Ralle, B., Rothgangel, M., Schön, L., et al. (Hrsg.). (2011). *Empirische Fundierung in den Fachdidaktiken*. Münster: Waxmann.
- Bennett, J., Lubben, F., Hogarth, S., & Campbell, B. (2005). Systematic reviews of research in science education: rigour or rigidity? *International Journal of Science Education*, 27(4), 387–406.
- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. (2. Aufl.) Berlin: Springer.
- Brovelli, D. (2014). Integrierte naturwissenschaftliche Lehrerbildung – Entwicklung professioneller Kompetenz bei Lehramtsstudierenden. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 21–32.
- 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.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2, 637–644.
- Denzin, N. (2016). *The qualitative manifesto: a call to arms*. Taylor and Francis. <https://doi.org/10.4324/9780429449987>.
- Engl, L., Schumacher, S., Sitter, K., Größler, M., Niehaus, E., Rasch, R., et al. (2015). Entwicklung eines Messinstrumentes zur Erfassung der Protokollierfähigkeit – initiiert durch Video-Items. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 223–229.
- Erickson, F. (1998). Qualitative research methods for science education. Dordrecht: Kluwer. 1155–1173.
- Flick, U. (2010). Gütekriterien qualitativer Forschung. In G. Mey & K. Muck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 395–407). Wiesbaden: VS.
- Flick, U., von Kardorff, E., & Steinke, I. (Hrsg.). (2012). *Qualitative Forschung: Ein Handbuch*. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.
- Florian, C., Schmiemann, P., & Sandmann, A. (2015). Aufgaben im Zentralabitur Biologie – eine kategoriengestützte Analyse charakteristischer Aufgabenmerkmale schriftlicher Abituraufgaben. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 69–86.
- Früh, W. (2007). *Inhaltsanalyse: Theorie und Praxis* (6. Aufl.). Konstanz: UVK.
- Gläser, J., & Laudel, G. (2013). Life with and without coding: two methods for early-stage data analysis in qualitative research aiming at causal explanations. *Forum Qualitative Sozialforschung*, 14. <https://doi.org/10.17169/fqs-14.2.1886>
- Gläser, J., & Laudel, G. (2004). Experteninterviews und qualitative Inhaltsanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gropengießer, H. (1997). Schülervorstellungen zum Sehen. *Zeitschrift für Didaktik der Naturwissenschaften*, 3, 71–87.
- Gropengießer, H. (2001). Didaktische Rekonstruktion des „Sehens“: Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung. Oldenburg: Carl-von-Ossietzky-Universität.
- Gropengießer, H. (2005). Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lernforschung. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (2. Aufl.). Weinheim: Beltz. 172–189.
- Groß, J., Hammann, M., Schmiemann, P., & Zabel, J. (Hrsg.). (2019). *Biologiedidaktische Forschung: Erträge für die Praxis*. Berlin: Springer.
- Grunenberg, H. (2004). Empirische Befunde zur Qualität qualitativer Sozialforschung. Resultate einer Analyse von Zeitschriftenartikeln. In U. Kuckartz, H. Grunenberg & A. Lauterbach (Hrsg.), *Qualitative Datenanalyse: computergestützt: Methodische Hintergründe und Beispiele aus der Forschungspraxis* (S. 65–80). Wiesbaden: VS.
- Haft, H. (1984). Inhaltsanalyse. In H. Haft & H. Kordes (Hrsg.), *Methoden der Erziehungs- und Bildungsforschung*. Stuttgart: Klett-Cotta. 411–418.
- Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Berlin Heidelberg: Berlin.
- Helsper, W., Herwartz-Emden, L., & Terhart, E. (2001). Qualität qualitativer Forschung in der Erziehungswissenschaft. *Zeitschrift für Pädagogik*, 14(2), 251–269.
- Helsper, W., Kelle, H., & Koller, H.-C. (2016). Qualitätskriterien der Begutachtung qualitativer Forschungsvorhaben in der Erziehungswissenschaft. Ergebnisse eines DFG-Roundtable. *Zeitschrift für Pädagogik*, 62(5), 738.
- Hesse, M. (2000). Erinnerungen an die Schulzeit – Ein Rückblick auf den erlebten Biologieunterricht junger Erwachsener. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 187–201.
- Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: a prolegomenon. *Educational Researcher*, 19(4), 2–9.
- Huber, G. L. (1989). Qualität versus Quantität in der Inhaltsanalyse. In W. Bos & C. Tarnai (Hrsg.), *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie*. Münster, New York: Waxmann. 32–47.
- Jelemenská, P. (2012). Lehrervorstellungen zum Lehren und Lernen von Evolution – Eine Fallstudie zum fachdidaktischen Coaching. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 231–261.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kasper, L., & Mikelskis, H. (2008). Lernen aus Dialogen und Geschichten im Physikunterricht – Ergebnisse einer Evaluationsstudie zum Thema Erdmagnetismus. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 7–25.
- Kelle, U. (2008). Qualitative vs. quantitative Forschung — die Debatte. In U. Kelle (Hrsg.), *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: Theoretische Grundlagen und methodologische Konzepte* (2. Aufl., S. 25–55). Wiesbaden: VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-91174-8\\_2](https://doi.org/10.1007/978-3-531-91174-8_2)
- Krippendorff, K. (1980). *Validity in content analysis*
- Krüger, D., & Riemer, T. (2014). Die qualitative Inhaltsanalyse – eine Methode zur Auswertung von Interviews. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 133–145). Berlin, Heidelberg: Springer.
- Krüger, D., Parchmann, I., & Schecker, H. (Hrsg.). (2014). *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin, Heidelberg: Springer.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz.
- Kühn, S. M. (2011). Weiterentwicklung der Aufgabenkultur im naturwissenschaftlichen Unterricht der gymnasialen Oberstufe und im Abitur. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 35–55.
- Kulgemeyer, C., & Tomczyszyn, E. (2015). Physik erklären – Messung der Erklärfähigkeit angehender Physiklehrkräfte in einer

- simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 111–126.
- Lamnek, S. (2010). *Qualitative Sozialforschung*. Weinheim: Beltz.
- Lamnek, S. (1989). *Qualitative Sozialforschung*. München: Psychologie Verlags Union.
- Leuders, T. (2015). Empirische Forschung in der Fachdidaktik. Eine Herausforderung für die Professionalisierung und die Nachwuchsqualifizierung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33(2), 215–234.
- Lisch, R., & Kriz, J. (1978). *Grundlagen und Modelle der Inhaltsanalyse*. Reinbek: Rowohlt.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: replication in the education sciences. *Educational Researcher*, 43(6), 304–316.
- Marsch, S., Hartwig, C., & Krüger, D. (2009). Lehren und Lernen im Biologieunterricht: Ein Kategoriensystem zur Beurteilung konstruktivistisch orientierter Lernumgebungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 109–130.
- Mayring, P. (2002). *Einführung in die Qualitative Sozialforschung*. Weinheim: Beltz.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Beltz.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Hrsg.), *Handbuch qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag. 601–613.
- Mayring, P. (2019). Qualitative Inhaltsanalyse – Abgrenzungen, Spielarten, Weiterentwicklungen. *Forum Qualitative Sozialforschung*, 20(3), 1–15.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. (2. Aufl.). Thousand Oaks, CA: SAGE.
- Mittelsten Scheid, N., & Hößle, C. (2008). Wie Schüler unter Verwendung syllogistischer Elemente argumentieren: Eine empirische Studie zu Niveaus von Argumentation im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 14(2008), 145–165.
- Nehring, A., Stiller, J., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Videostudie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 77–96.
- Patton, M. (2001). *Qualitative research & evaluation methods*. Saint Paul: SAGE.
- Ramsenthaler, C. (2013). Was ist „Qualitative Inhaltsanalyse?“. In M. Schnell, C. Schulz, H. Kolbe, & C. Dunger (Hrsg.), *Der Patient am Lebensende: Eine Qualitative Inhaltsanalyse* (S. 23–42). Wiesbaden: Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-531-19660-2\\_2](https://doi.org/10.1007/978-3-531-19660-2_2)
- Reinisch, B., Helbig, K., & Krüger, D. (Hrsg.). (in Vorbereitung). *Biologiedidaktische Vorstellungsforschung: Zukunftsweisende Praxis*. Berlin: Springer.
- Reitschert, K., & Hößle, C. (2007). Wie Schüler ethisch bewerten: Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sek. I. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 125–143.
- Riemeier, T. (2005). Schülervorstellungen von Zellen, Teilung und Wachstum. *Zeitschrift für Didaktik der Naturwissenschaften*, 11(1), 52–72.
- Riemeier, T., Jankowski, M., Kersten, B., Pach, S., Rabe, I., Sundermeier, S., et al. (2010). Wo das Blut fließt. Schülervorstellungen zu Blut, Herz und Kreislauf beim Menschen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 77–93.
- Riemeier, T., Jankowski, M., Kersten, B., Pach, S., Rabe, I., Sundermeier, S., & Gropengießer, H. (2010). Wo das Blut fließt. Schülervorstellungen zu Blut, Herz und Kreislauf beim Menschen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 77–93.
- Riess, W. (2003). Die Kluft zwischen Umweltwissen und Umwelthandeln als pädagogische Herausforderung – Entwicklung und Erprobung eines Prozessmodells zum Umwelthandeln in alltäglichen Anforderungssituationen. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 147–159.
- Sandmann, A., & Schmiemann, P. (Hrsg.). (2016). *Biologiedidaktische Forschung: Schwerpunkte und Forschungsgegenstände*. Berlin: Logos.
- Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 189–207.
- Schönfelder, S., & Bögeholz, S. (2009). Bewertungskompetenz in der reflexiven Leitbildarbeit eines Umweltbildungszentrums – Ein Beitrag zur Professionalisierung des pädagogischen Personals. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 247–263.
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: SAGE.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum qualitative Sozialforschung*, 15(1), 1–27.
- Schwedler, S. (2017). Was überfordert Chemiestudierende zu Studienbeginn? *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 165–179.
- Seale, C. (2002). Quality issues in qualitative inquiry. *Qualitative Social Work*, 1(1), 97–110.
- Stamann, C., Janssen, M., & Schreier, M. (2016). Qualitative Inhaltsanalyse – Versuch einer Begriffsbestimmung und Systematisierung. *Forum qualitative Sozialforschung*, 17, 1–16.
- Staraschek, E. (2006). Der Einfluss von Textkohäsion und gegenständlichen externen piktoralen Repräsentationen auf die Verständlichkeit von Texten zum Physiklernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 127–157.
- Steinke, I. (2013). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung: Ein Handbuch*. Reinbek bei Hamburg: Rowohlt.
- Uhden, O. (2016). Verständnisprobleme von Schülerinnen und Schülern beim Verbinden von Physik und Mathematik. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 13–24.
- Weitzel, H., & Gropengießer, H. (2009). Vorstellungsentwicklung zur stammesgeschichtlichen Anpassung: Wie man Lernhindernisse verstehen und förderliche Lernangebote machen kann. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 287–305.
- Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., Paulsen, Albert; Robinault, Karine; & von Aufschneider, S. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden – Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, 4(1), 29–44.
- Wernecke, U., Schwanewedel, J., Schütte, K., & Harms, U. (2016). Wie wird Energie im Biologieschulbuch dargestellt? – Entwicklung eines Kategoriensystems und exemplarische Anwendung auf eine Schulbuchreihe. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 215–229.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.

## Modellierungsprozesse von Lehramtsstudierenden der Biologie

Maximilian Göhner & Moritz Krell

maximilian.goehner@fu-berlin.de – moritz.krell@fu-berlin.de

Freie Universität Berlin: Didaktik der Biologie,

Schwendenerstraße 1, 14195 Berlin

---

### **Zusammenfassung**

*Um im naturwissenschaftlichen Unterricht Modellierungsprozesse von SchülerInnen unterstützen zu können, benötigen Lehrkräfte neben dem Modellverstehen auch Fähigkeiten des Modellierens als Teil ihrer professionellen Kompetenz. Zur Beschreibung von Modellierungsprozessen wurden zwei angehende Biologielehrkräfte in dieser Studie dazu aufgefordert, zeichnerisch ein Modell des Inneren einer Blackbox zu entwickeln. Ihr Vorgehen dabei wurde videografiert, zusätzlich wurde das Modellverstehen der ProbandInnen mit einem Fragebogen erfasst. Die Modellierungsprozesse wurden eventbasiert codiert und ausgewertet, dabei konnten in einem Fall nur explorierende und herstellende Tätigkeiten des Modellierens beobachtet werden (expressives Modellieren). Im zweiten Fall zeigten sich auch Tätigkeiten, bei denen wiederholt Vorhersagen vom Modell über das Original abgeleitet wurden (zyklisches Modellieren). Es gibt keine Hinweise auf einen direkten Zusammenhang zwischen dem umgesetzten Modellierungsprozess und dem Modellverstehen.*

### **Abstract**

*To guide modeling practices of students in science lessons, science teachers need meta-modeling knowledge as well as modeling skills as part of their professional competencies. Aiming to describe modeling processes, two pre-service biology teachers were presented with a black box and challenged to draw a model of its inner system. Their modeling processes were videotaped and their meta-modeling knowledge was assessed through a questionnaire. Modeling processes were coded and analyzed based on events. One pre-service teacher showed activities limited to the exploration of the black box and the development of an accurate representation of it (expressive modeling). The other pre-service teacher additionally used his developed model repeatedly for predictions regarding the*

*black box's behavior (cyclic modeling). No indications were found for a consistent relation between meta-modeling knowledge and expressed modeling processes.*

## 1 Einleitung

Modelle sind in den Naturwissenschaften von zentraler Bedeutung (LAUBICHLER & MÜLLER, 2007; ODENBAUGH, 2005) und werden auch in Unterrichtskontexten als wichtig bewertet, zum Beispiel zur Förderung eines allgemeinen Wissenschaftsverständnisses (WINDSCHITL, THOMPSON & BRAATEN, 2008). Modelle können als zweckgerichtet entwickelte Entitäten (*epistemic artifacts*; KNUUTTILA, 2005) aufgefasst werden, in denen ausgewählte Eigenschaften eines Originals berücksichtigt, andere hingegen vernachlässigt sind. Modelle werden in den Naturwissenschaften für vielfältige Zwecke entwickelt, zum Beispiel für die Untersuchung von Phänomenen, die Entwicklung von Erklärungen oder für die Darstellung und Kommunikation (LAUBICHLER & MÜLLER, 2007; ODENBAUGH, 2005).

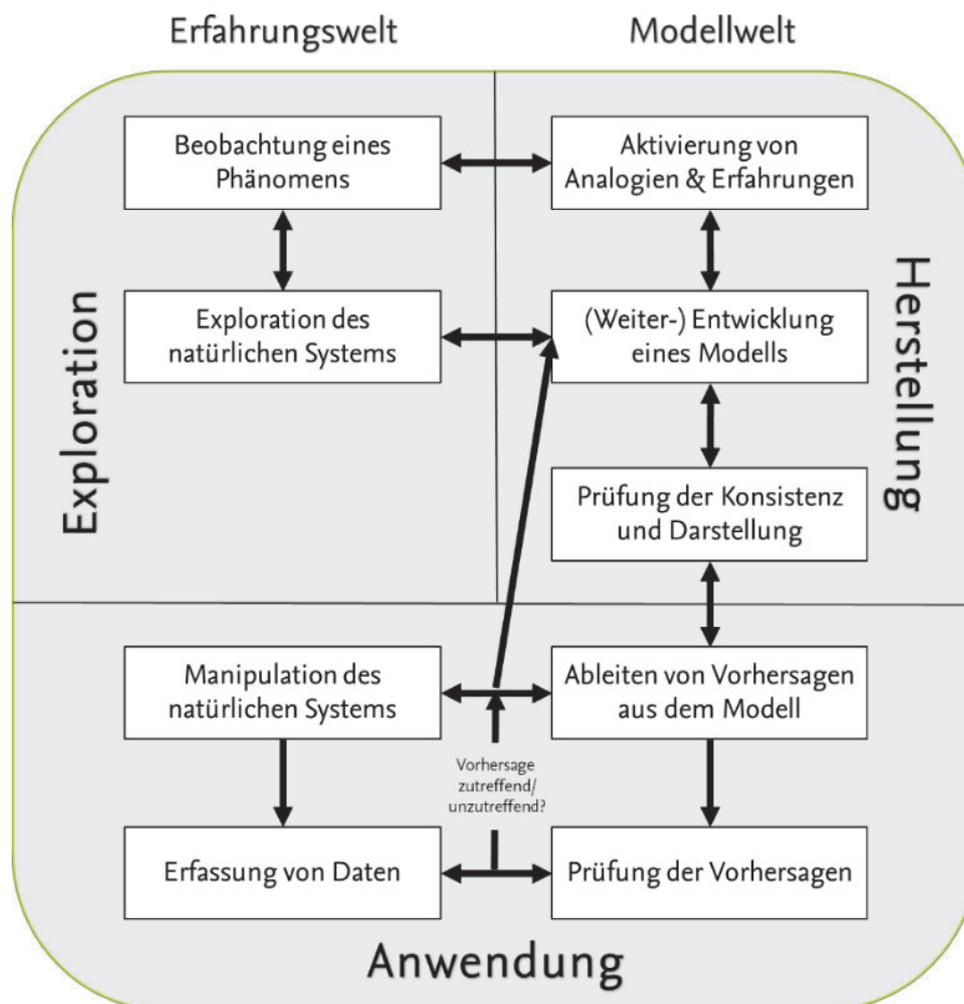
Modellbasiertes Lernen im naturwissenschaftlichen Unterricht kann kognitive, meta-kognitive, soziale und erkenntnisgewinnende Fähigkeiten fördern und so insgesamt den Lernerfolg erhöhen (LOUCA & ZACHARIA, 2012). Dazu werden Modelle im Unterricht als Medien zur Veranschaulichung oder Erklärung genutzt (Herstellungsperspektive; UPMEIER ZU BELZEN & KRÜGER, 2010) sowie prädiktiv als Werkzeuge der Erkenntnisgewinnung eingesetzt (Anwendungsperspektive; UPMEIER ZU BELZEN & KRÜGER, 2010). Schulische Steuerungsdokumente ordnen Modelle in den Kompetenzbereich Erkenntnisgewinnung ein und fordern explizit, dass SchülerInnen selbst modellieren sowie Hypothesen aus Modellen ableiten (KMK, 2005). Lehrkräfte naturwissenschaftlicher Fächer benötigen daher Modellkompetenz (d. h. Fähigkeiten des Modellierens und ein ausgeprägtes Modellverstehen; NICOLAOU & CONSTANTINOU, 2014) als Teil ihrer professionellen Kompetenz (GÜNTHER, FLEIGE, UPMEIER ZU BELZEN & KRÜGER, 2017), um entsprechende Lerngelegenheiten für SchülerInnen planen und durchführen zu können.

Bislang vorliegende Studien über die Modellkompetenz von Lehrkräften naturwissenschaftlicher Fächer konzentrieren sich vorwiegend auf die Erfassung, Beschreibung und Förderung des Modellverstehens (z. B. KRELL & KRÜGER, 2016). Studien über die Fähigkeiten im Bereich des praktischen Modellierens liegen bislang kaum vor (NICOLAOU & CONSTANTINOU, 2014). Die vorliegende Arbeit trägt dazu bei, diese Lücke in der naturwissenschaftsdidaktischen Forschung zu schließen.

## 2 Theorie

### 2.1 Modellieren und die Beschreibung von Modellierungsprozessen

Der Prozess des Modellierens wird idealtypisch als dynamischer und komplexer Zyklus beschrieben (GIERE, BICKLE & MAULDIN, 2006; GILBERT & JUSTI, 2016; KHAN, 2011; PASSMORE, GOUVEA & GIREE, 2014). Zur genaueren Beschreibung des Modellierungsprozesses wurden in der naturwissenschaftsdidaktischen Forschung unterschiedliche Ansätze vorgeschlagen (vgl. LÖHNER, VAN JOOLINGEN, SAVELSBERGH & VAN HOUTWOLTERS, 2005), oft wird der Modellierungsprozess dabei als drei- oder vierschrittig aufgefasst (KHAN, 2007; LOUCA & ZACHARIA, 2012). In dieser Studie wird das Prozessschema naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren zur Beschreibung des Modellierungsprozesses genutzt (KRELL, UPMEIER ZU BELZEN & KRÜGER, 2016; Abbildung 1).



**Abbildung 1:** Prozessschema naturwissenschaftlicher Erkenntnisgewinnung durch Modellieren (verändert nach KRELL ET AL., 2016).

Dieses unterscheidet drei übergeordnete Phasen des Modellierens: Exploration, Herstellung und Anwendung (LEISNER-BODENTHIN, 2006; LOUCA & ZACHARIA, 2012; UPMEIER ZU BELZEN & KRÜGER, 2010).

Ausgehend vom dargestellten Prozessschema beginnt der Modellierungsprozess mit der Beobachtung eines Phänomens und der Exploration des zugrundeliegenden Systems, bei der Daten erfasst und dokumentiert werden (KRELL, WALZER, HERGERT & KRÜGER, 2017; LOUCA & ZACHARIA, 2012). Die Aktivierung von Analogien und Erfahrungen ermöglicht es dann ein Modell zu entwickeln, welches die für das Phänomen als relevant erachteten Variablen enthält (BROWN & CLEMENT, 1989). Das Modell kann zuerst auf interne Konsistenz sowie eine angemessene Darstellung des beobachteten Phänomens geprüft werden (Herstellungsperspektive bzw. Modelle als Medien; KRELL et al., 2017; UPMEIER ZU BELZEN & KRÜGER, 2010).

Erst in der untersuchungsleitenden Anwendung des Modells durch das Ableiten von Vorhersagen wird der zyklische Charakter des Modellierens zur Erkenntnisgewinnung klar; das Modell wird durch die empirische Überprüfung der abgeleiteten Vorhersagen evaluiert. Werden die abgeleiteten Vorhersagen falsifiziert, muss das Modell optimiert oder gänzlich verworfen werden. In beiden Fällen schließt sich eine erneute Testung im beschriebenen Sinne an (Anwendungsperspektive bzw. Modelle als Werkzeuge; GIÈRE et al., 2006; KRELL et al., 2017; UPMEIER ZU BELZEN & KRÜGER, 2010).

## **2.2 Modellkompetenz als Teil der professionellen Kompetenz von Biologielehrkräften**

Anknüpfend an SHULMAN (1986) wird innerhalb der professionellen Kompetenz von Lehrkräften zwischen den Dimensionen des Professionswissen *content knowledge*, *pedagogical content knowledge* und *pedagogical knowledge* unterschieden (BAUMERT & KUNTER, 2013). Entsprechend wird in den ländergemeinsamen inhaltlichen Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung gefordert, dass angehende Biologielehrkräfte im Laufe ihres Studiums „Kenntnisse und Fertigkeiten ... im hypothesengeleiteten Experimentieren und Modellieren“ erwerben (KMK, 2017, S. 22), das heißt Modelle als Werkzeuge der Erkenntnisgewinnung verstehen und nutzen können (UPMEIER ZU BELZEN & KRÜGER, 2010). Die hier betrachteten Facetten der Modellkompetenz, Modellverstehen und Fähigkeiten des Modellierens, werden dem *content knowledge* zugeordnet (GÜNTHER et al., 2017; JUSTI & VAN DRIEL, 2005). Modellverstehen umfasst dabei Wissen über Modelle und das Modellieren, welches in die fünf Dimensionen *Eigenschaften von Modellen*, *Alternative*

*Modelle, Zweck von Modellen, Testen von Modellen und Ändern von Modellen* strukturiert werden kann (Kompetenzmodell der Modellkompetenz; UPMEIER ZU BELZEN & KRÜGER, 2010). Neben den kognitiven Fähigkeiten des Modellierens sind zur praktischen Umsetzung von Modellierungsprozessen (vgl. Abbildung 1) auch die volitionale, motivationale und soziale Bereitschaft zur Problemlösung relevant (UPMEIER ZU BELZEN & KRÜGER, 2010).

Bisherige Studien weisen darauf hin, dass angehende und praktizierende Lehrkräfte häufig über eher limitiertes Modellverstehen verfügen und Modelle oft primär als Medien betrachten (CRAWFORD & CULLIN, 2005; JUSTI & GILBERT, 2003; KRELL & KRÜGER, 2016; VAN DRIEL & VERLOOP, 1999). Entsprechend werden Modelle im naturwissenschaftlichen Unterricht nur selten im Sinne des zyklischen Modellierungsprozesses entwickelt, getestet und gegebenenfalls optimiert (CAMPBELL, OH, MAUGHN, KIRIAZIS & ZUWALLACK, 2015; KHAN, 2011; KRELL & KRÜGER, 2013, 2016). Campbell und Kollegen (z. B. CAMPBELL, OH & NEILSON, 2013) unterscheiden fünf Unterrichtsmuster (*modeling pedagogies*), nach denen Modelle im naturwissenschaftlichen Unterricht eingesetzt werden können: exploratives Modellieren (Untersuchung eines gegebenen Modellobjekts), expressives Modellieren (Herstellung eines Modells zur Erklärung), experimentelles Modellieren (einmaliges Ableiten und Testen von Hypothesen), evaluatives Modellieren (Vergleich alternativer Modelle) sowie zyklisches Modellieren (zyklischer Prozess des Entwickelns und Anwendens eines Modells; vgl. Abbildung 1). Im naturwissenschaftlichen Unterricht werden vorrangig die Unterrichtsmuster exploratives und expressives Modellieren umgesetzt (CAMPBELL et al., 2015).

Studien über Fähigkeiten und Vorgehensweisen beim praktischen Modellieren von Lehrkräften liegen kaum vor. KRELL und HERGERT (im Druck) beschreiben bei Biologie-Lehramtsstudierenden vornehmlich explorierende und herstellende Tätigkeiten und trennen dabei entsprechend des Prozessschemas (Abbildung 1) Explorations- und Herstellungsphasen zeitlich voneinander. SINS, SAVELSBERGH und VAN JOOLINGEN (2005) beobachteten bei SchülerInnen besonders häufig die Tätigkeit Zusammenfassung, dabei wird ein erreichter Fortschritt im Modellierungsprozess dokumentiert und paraphrasiert.

Es wird angenommen, dass ein ausgeprägtes Modellverstehen (CHENG & LIN, 2015; LOUCA & ZACHARIA, 2012; SCHWARZ et al., 2009), allgemeine kognitive Fähigkeiten (NEHRING, NOWAK, UPMEIER ZU BELZEN & TIEMANN, 2015), fachliches Vorwissen (RUPPERT, DUNCAN & CHINN, 2017) sowie Fähigkeiten naturwissenschaftlicher Erkenntnisgewinnung (KRELL, UPMEIER ZU BELZEN & KRÜGER, 2014) die Umsetzung von Modellierungsprozessen unterstützen.

### 3 Fragestellungen

Die vorliegende Studie untersucht drei Fragestellungen:

1. Welche Tätigkeiten setzen Biologie-Lehramtsstudierende beim Modellieren um?

Es wird erwartet, dass im Vorgehen der ProbandInnen vornehmlich explorierende und herstellende Tätigkeiten (vgl. Abbildung 1) umgesetzt werden (CAMPBELL et al., 2015; KHAN, 2011; KRELL & KRÜGER, 2013, 2016). Insbesondere Zusammenfassungen werden häufig und unabhängig von der Modellierungsphase auftretend erwartet (SINS et al., 2005).

2. Inwiefern lassen sich in den Modellierungsprozessen von Biologie-Lehramtsstudierenden die drei Phasen Exploration, Herstellung und Anwendung zeitlich voneinander abgrenzen (vgl. Abbildung 1)?

Es wird erwartet, dass Tätigkeiten, die theoriegeleitet derselben Phase zugeordnet werden, zusammen auftreten und somit Phasen der Exploration, Herstellung und Anwendung zeitlich voneinander abgegrenzt werden können (KRELL et al., 2017).

3. Inwiefern besteht ein Zusammenhang zwischen den gezeigten Tätigkeiten beim Modellieren und dem Modellverstehen von Biologie-Lehramtsstudierenden?

Es wird erwartet, dass ausgeprägtes Modellverstehen das vermehrte Auftreten von Tätigkeiten der Anwendungsphase unterstützt (CHENG & LIN, 2015; LOUCA & ZACHARIA, 2012; SCHWARZ et al., 2009).

## 4 Methodik

### 4.1 Setting

Um Modellierungsprozesse anregen und beobachten zu können wurden Biologie-Lehramtsstudierende mit einer „Wasser-Blackbox“ (KRELL et al., 2017) konfrontiert. Die Blackbox erwies sich bereits in einer früheren Studie mit SchülerInnen als wirksame Methode, um Modellierungsprozesse anzuregen (KOCH, KRELL & KRÜGER, 2015). Um Modellierungsprozesse und eine kreative Analogiebildung zu unterstützen (ORSENNE, 2015), wird den ProbandInnen die Aufgabe gestellt, an einer bereitgestellten Wandtafel ein Modell des Inneren der Blackbox zu zeichnen. Gleichzeitig werden die ProbandInnen zum lauten Denken aufgefordert (LEIGHTON & GIERL, 2007), was vor Beginn der Modellierungsaufgabe erklärt und geübt wird (SANDMANN, 2014).

Nach Betreten des vorbereiteten Raums (Abbildung 2) erhalten die ProbandInnen folgende Instruktion: „Hier ist die Blackbox, diese darfst du nicht öffnen. Du kannst in die Blackbox Wasser schütten und beobachten, was passiert. Ich empfehle mit dem mehrmaligen Eingießen von 400ml zu beginnen, um einen ersten Eindruck von der Blackbox zu bekommen.“ Darüber hinaus erfolgt keine weiteren Interaktion zwischen den ProbandInnen und der Versuchsleitung (Ausnahme ist ggf. eine Erinnerung an das laute Denken).



**Abbildung 2:** Setting der Blackbox-Untersuchung (links: mit der Aufgabe versehene Wandtafel für Modell und Notizen, beispielhaft beschriftet durch den Autor; mittig: Kamera; rechts: Blackbox und dazugehörige Untersuchungsutensilien).

Durch eine wiederholte Eingabe von 400 ml Wasser ergeben sich sechs verschiedene Eingabe/Ausgabe-Paare (Tabelle 1). ProbandInnen fällt es erfahrungsgemäß schwer die dritte Ausgabe zu erklären (KRELL et al., 2017), was zur weiteren Untersuchung der Blackbox motiviert (UPMEIER ZU BELZEN, 2014). Die Dauer der Untersuchung ist nicht vorgegeben und kann von den ProbandInnen frei bestimmt werden.

**Tabelle 1:** Eingabe/Ausgabe der Blackbox, ein Beispiel.

Eingabe	Ausgabe
400	0
400	400
400	600
400	400
400	0
400	1000

#### 4.2 Datenerhebung und Auswahl der ProbandInnen

Nach dem Prinzip der heterogenen Fallauswahl (KELLE & KLUGE, 2010; SCHREIER, 2010) werden ProbandInnen auf Basis ihrer Fähigkeiten im *wissenschaftlichen Denken* (Ko-WADiS-Fragebogen; HARTMANN et al., 2015) sowie ihrer allgemeinen kognitiven Fähigkeiten im *schlussfolgernden Denken*

(I-S-T 2000 R; LIEPMANN, BEAUDUCEL, BROCKE & AMTHAUER, 2007) ausgewählt. Zur Studienteilnahme werden Studierende eingeladen, die im Ko-WADiS-Fragebogen eine Standardabweichung und gleichzeitig im I-S-T 2000 R eine halbe Standardabweichung besser („hoch“) oder schlechter („niedrig“) als die jeweilige Normstichprobe abschneiden. Die Verteilung der bisher 13 ProbandInnen ist in Tabelle2 dargestellt.

**Tabelle2:** Verteilung der ProbandInnen auf die Hintergrundvariablen.

	wissenschaftliches Denken (Ko-WADiS)	
	hoch	niedrig
schlussfolgerndes Denken (I-S-T 2000 R)	hoch	7
	niedrig	3
		2
		1

Im Folgenden werden zur Diskussion der Fragestellungen die ersten beiden vollständig ausgewerteten Fälle (Pseudonyme Angelina und Raphael) exemplarisch vorgestellt. Diese weisen in den beiden Hintergrundvariablen überdurchschnittliche Performanz auf (d. h. hoch/hoch), was das Auftreten von Tätigkeiten der Anwendung begünstigen könnte (vgl. Fragestellungen 1 und 2). Zur Untersuchung von Fragestellung 3 wird das Modellverstehen der ProbandInnen vor Beginn der Modellierungsaufgabe mit einem offenen Fragebogen erfasst (fünf Fragen zu den Teilkompetenzen der Modellkompetenz; KRELL & KRÜGER, 2016).

### 4.3 Datenauswertung

Die Videografien werden wortwörtlich, aber sprachlich geglättet transkribiert. Dabei werden Schlüsselhandlungen, die den Codierprozess unterstützen, wie beispielsweise das Abwischen der Tafel oder das Vornehmen einer Eingabe von Wasser, in das Transkript aufgenommen (gekennzeichnet durch z. B. [Tafel], [Input]). Die Transkription und anschließende Codierung erfolgt mit Hilfe des Programms MAXQDA (VERBI SOFTWARE, 2018). Die Transkripte werden entsprechend der strukturierenden qualitativen Inhaltsanalyse (MAYRING, 2010) eventbasiert (SEIDEL, PRENZEL, DUIT & LEHRKE, 2003) auf der Basis eines Codierleitfadens mit 19 Tätigkeiten des Modellierens (sog. Codiereinheiten) codiert (KRELL et al., 2017). Die 19 Tätigkeiten können den neun in Abbildung 1 dargestellten Modellierungsschritten zugeordnet werden (Tabelle 3). Die Auswertung der Antworten des Fragebogens zum Modellverstehen erfolgt ebenfalls mit einem bereits erprobten Codierleitfaden (KRELL & KRÜGER, 2016).

**Tabelle 3:** Übersicht über die Tätigkeiten des Modellierens respektive Codiereinheiten (KRELL et al., 2017).

Code	Tätigkeit (Subcode)
<b>Beobachtung eines Phänomens</b>	1. Probanden nehmen Verhalten von Blackbox als spontan nicht erklärbar wahr oder erkennen fehlende Passung zwischen angenommenem Muster der Blackbox und Verhalten von Blackbox
<b>Exploration des Systems</b>	2. Probanden nehmen Eingabe von Input vor und/ oder beobachten den Output (explorativ; nicht hypothesengeleitet) 3. Probanden fassen das beobachtete Verhalten der Blackbox zusammen oder beschreiben es 4. Probanden nehmen Eingabe von Input vor und/ oder beobachten den Output (Mustererkennung) 5. Probanden bestätigen/ erkennen Muster
<b>Aktivierung von Analogien und Erfahrungen</b>	6. Probanden formulieren Ideen über Mechanismus der Blackbox und/ oder reflektieren Vor- und Nachteile ihrer Ideen und verwerfen ggf. Ideen (ohne zeichnerische Umsetzung)
<b>(Weiter-) Entwicklung eines Modells</b>	7. Probanden entwickeln auf Grundlage von Beobachtung, Analogien und/ oder Erfahrungen zeichnerisch ein Modell der Blackbox 8. Probanden entwickeln Modell weiter zur Optimierung von Funktionsfähigkeit, Ästhetik o. ä. (Modellobjekt) 9. Probanden entwickeln Modell weiter aufgrund fehlerhafter Übereinstimmungen mit Beobachtungen an Blackbox (retrospektiv) 10. Probanden verwerfen Modell aufgrund fehlerhafter Übereinstimmungen mit Beobachtungen an Blackbox (retrospektiv; vgl. Kategorie 12) oder mangelnder Konsistenz des Modellobjekts (vgl. Kategorie 11).
<b>Prüfung von Konsistenz und Darstellung</b>	11. Probanden reflektieren/ überprüfen/ bewerten Konsistenz des Modells (Modellobjekt) 12. Probanden vergleichen Eigenschaften des Modells mit Beobachtungen an Blackbox (retrospektiv)
<b>Feststellung Konsistenz &amp; Darstellung des Modells</b>	13. Probanden stellen Konsistenz und Angemessenheit des Modells fest
<b>Ableiten von Vorhersage (Hypothese) aus Modell</b>	14. Probanden nutzen Modell um Vorhersage über Output bei bestimmtem Input zu treffen
<b>Prüfung der Vorhersage (datenbasiert)</b>	15. Probanden nehmen Eingabe von Input vor und beobachten den Output (hypothesengeleitet aus Modell) 16. Probanden bestätigen Hypothese aus Modell durch Output von Blackbox 17. Probanden falsifizieren Hypothese aus Modell durch Output von Blackbox
<b>Ändern / Verwerfen des Modells (datenbasiert)</b>	18. Probanden entwickeln Modell aufgrund von falsifizierter Hypothese weiter (datenbasiert) 19. Probanden verwerfen Modell aufgrund von falsifizierter Hypothese (datenbasiert)

Alle Codierungen werden zweimal mit mindestens einwöchigem Abstand durch die Versuchsleitung vorgenommen. Anschließend codiert eine geschulte Projektmitarbeiterin das Material erneut, sodass Intrarater- und Interrater-Reliabilität mit Cohens Kappa geprüft werden können. Mit Ausnahme der Interrater-Reliabilität des Transkripts von Raphael (moderat) fallen sämtliche Kappa-Werte sehr hoch aus (vgl. Tabelle 4; Grenzwerte nach LANDIS & KOCH, 1977). Nach Bildung eines finalen Konsenses durch die beiden CodiererInnen werden die entstandenen Codierungen der Videos als „Codelines“ visualisiert.

**Tabelle 4:** Intrarater- und Interrater-Reliabilität (Cohens Kappa).

	Offener Fragebogen zur Erfassung des Modellverstehens (KRELL & KRÜGER, 2016)		Codierung der Videografie nach Tätigkeiten (KRELL et al., 2017)	
	Intrarater- Reliabilität	Interrater- Reliabilität	Intrarater- Reliabilität	Interrater- Reliabilität
Angelina	1,00	0,73	0,67	0,82
Raphael	1,00	1,00	0,75	0,45

## 5 Ergebnisse

### 5.1 Tätigkeiten des Modellierens (Fragestellung 1)

Angelina führt in 63 Minuten 62 Tätigkeiten aus. Dabei verbringt sie die Hälfte der Zeit mit der Zusammenfassung von Beobachtungen, ein weiteres Drittel machen die übrigen explorierenden Tätigkeiten aus (Tätigkeiten 1, 2, 4, 5; vgl. Tabelle 3), sodass sie insgesamt 84 % der Zeit mit Exploration verbringt. Die restliche Zeit verwendet sie auf Tätigkeiten der Herstellung. Angelina führt damit keine Tätigkeiten der Anwendung aus (Tabelle 5).

**Tabelle 5:** Übersicht über den zeitlichen Anteil und die Anzahl des Vorkommens der von Angelina und Raphael ausgeführten Tätigkeiten während des Modellierens.

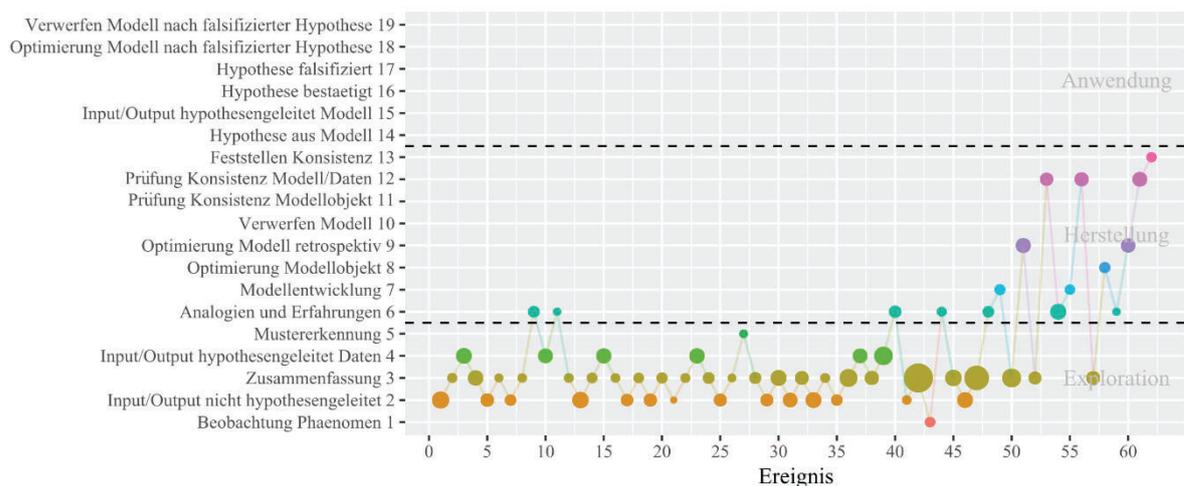
Tätigkeit	Angelina				Raphael				
	Zeit		Anzahl		Zeit		Anzahl		
	absolut [s]	%	absolut	%	absolut [s]	%	absolut	%	
Exploration	1	19,3	1	1	2	122,3	3	1	1
	2	731,1	20	14	23	682,5	18	16	21
	3	1807,2	49	24	39	946	25	17	22
	4	510,2	14	6	10	441	12	8	10
	5	8,4	0	1	2	31	0	2	3
Herstellung	6	211,1	6	7	11	20,3	1	3	4
	7	39,9	1	2	3	441,9	12	5	6
	8	23,8	0	1	2	0	0	0	0
	9	134,9	4	2	3	148,3	4	2	3
	10		0		0	105,8	3	2	3
	11		0		0	8,3	0	1	1
	12	173,9	5	3	5	338,3	9	5	6
	13	17,2	0	1	2	81,8	2	2	3
Anwendung	14		0		0	133,3	4	3	4
	15		0		0	162,7	4	4	5
	17		0		0	56,8	1	3	4
	18		0		0	58,1	2	2	3
	19		0		0	18,6	0	1	1
<b>Summe</b>		3677	100	62	100	3797	100	77	100

Raphael führt in 66 Minuten 77 Tätigkeiten aus und verbringt 25 % der Zeit mit dem Zusammenfassen seiner Beobachtungen. Insgesamt verwendet er 58 % der Zeit auf Tätigkeiten der Exploration und 31 % auf Tätigkeiten der Herstellung.

Darüber hinaus sind 15 Tätigkeiten (11 % der Zeit) der Anwendung zugeordnet (Tabelle 5).

## 5.2 Phasen des Modellierens (Fragestellung 2)

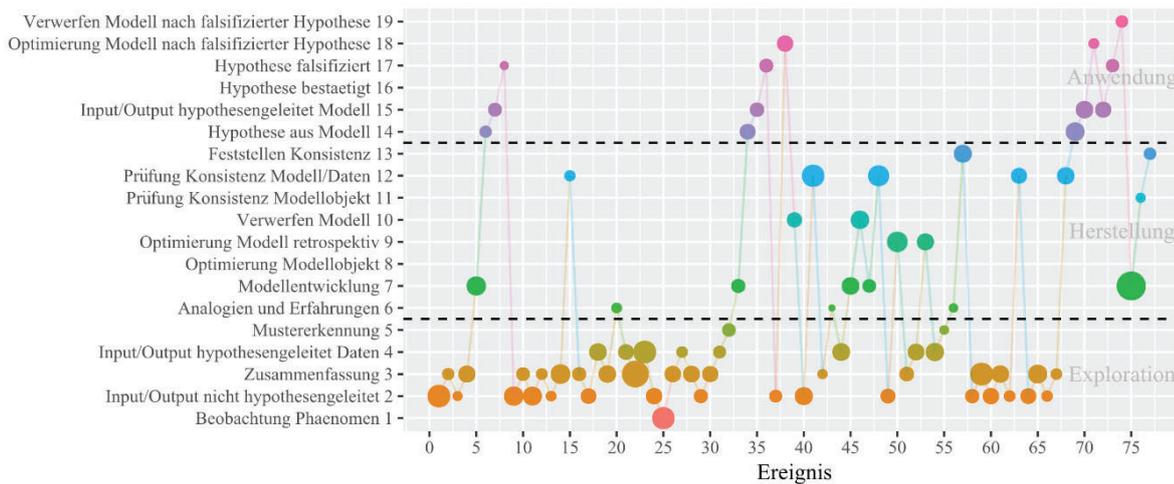
Angelina führt bis Ereignis 48 explorierende Tätigkeiten aus (Abbildung 3). Ein- und Ausgaben wechseln sich mit der Überprüfung des von ihr vermuteten Musters und Zusammenfassungen ab. Außerdem bildet sie mehrere Analogien, zum Beispiel vermutet sie „Säulen, die verschieden hoch befüllt sind“ in der Blackbox. Ab Ereignis 49 zeichnet sie, unterbrochen von weiteren Zusammenfassungen und Analogiebildungen, zwei Modelle, die sie prüft, indem sie diese retrospektiv mit den gewonnenen Daten vergleicht und anschließend optimiert. Sie beendet ihre Modellierung mit unsicherer Konsistenzfeststellung: „Ich gehe vorläufig jetzt erstmal von verschiedenen Behältern aus, die kippen und dann verschiedene Fassungsvermögen haben [Tafel], sodass sich erklärt, warum es so unregelmäßig wieder austritt das Wasser.“



**Abbildung 3:** Tätigkeitssequenz (Codeline) von Angelina. Zeitlich länger andauernde Tätigkeiten sind als größere Punkte dargestellt.

Raphael (Abbildung 4) zeichnet nach einer kurzen Phase explorierender Tätigkeiten, in der zwei Eingaben vorgenommen und zwei Ausgaben beobachtet werden, ein initiales Modell mit der Vermutung „in der Blackbox [müsste] irgendein Auffangbehälter sein.“ Aus diesem Modell leitet er die Vorhersage ab, dass bei erneuter Eingabe von Wasser dieselbe Menge ausfließen müsste, welche er anschließend falsifiziert. Darauf folgen erneut überwiegend explorierende Tätigkeiten, bis er bei Ereignis 32 ein Muster im Verhalten der Blackbox feststellt, woraufhin er ein weiteres Modell zeichnet. Von diesem leitet er erneut eine Vorhersage ab, die er testet und falsifiziert. Darauf folgen abwechselnd herstellende und explorierende Tätigkeiten, bis er bei Ereignis 57 sein Modell als konsistent erachtet: „Das würde auf jeden Fall Sinn ergeben. Jetzt würde ich ganz gern nur noch rausfinden, wie viele

Kammern das sind und wie viel die Kammern fassen.“ Dieser Fragestellung folgen weitere Explorations-, Herstellungs- und Anwendungstätigkeiten. Er beendet den Modellierungsprozess nach 65 Minuten mit der Herstellung seines finalen Modells. Dabei beschreibt er das Innere der Blackbox mit großer Sicherheit: „Auf jeden Fall gibt es sich mehr oder weniger komplett entleerende Ventile. Dann gibt es mindestens zwei unterschiedlich große Behälter, in die das Wasser gleichmäßig verteilt wird. Der eine ist natürlich deutlich größer als der andere.“



**Abbildung 4:** Tätigkeitssequenz (Codeline) von Raphael. Zeitlich länger andauernde Tätigkeiten sind als größere Punkte dargestellt.

### 5.3 Modellieren und Modellverstehen (Fragestellung 3)

Angelinas und Raphaels Modellverstehen erscheint eher limitiert. Ihre Antworten zum offenen Fragebogen beziehen sich durchweg auf Modelle als Medien (Niveau II der Modellkompetenz nach UPMEIER ZU BELZEN & KRÜGER, 2010). Zum Beispiel formuliert Angelina: „Der grundlegende Zweck [von Modellen] liegt in der Veranschaulichung komplexer Sachverhalte“, analog dazu beschreibt Raphael den Zweck von Modellen als „Veranschaulichung von Objekten, die am Original nicht so gut zu erkennen sind oder im Unterricht nicht im Original vorgeführt werden können“. Ein Einsatz von Modellen zur Erkenntnisgewinnung (Niveau III der Modellkompetenz nach UPMEIER ZU BELZEN & KRÜGER, 2010) wird weder von Angelina noch von Raphael erwähnt, beide argumentieren dagegen, dass Modelle ihre Gültigkeit verlieren, sobald sie neuen Erkenntnisse nicht entsprechen. Obwohl Raphael also Tätigkeiten aus dem Bereich der Anwendung durchführt und aus seinen Modellen Vorhersagen ableitet (Abbildung 4), zeigt er kein entsprechendes Modellverstehen.

## 6 Diskussion

### 6.1 Tätigkeiten des Modellierens (Fragestellung 1)

Entsprechend der Befunde von KRELL und HERGERT (im Druck) dominieren auch bei Angelina explorierende und herstellende Tätigkeiten. Parallel zu den Ergebnissen von SINS et al. (2005) treten insbesondere Zusammenfassungen sehr häufig auf, unabhängig von der vorhergehenden Tätigkeit. Dies erklären die Autoren mit der Wahrnehmung von Modellierungsaufgaben im Sinne des Ingenieursmodus statt des Wissenschaftsmodus (vgl. SCHAUBLE, KLOPPER & RAGHAVAN, 1991). Das heißt Angelina versucht das Modell durch kleinschrittige Anpassungen (*model fitting*) zu verbessern (Ingenieursmodus), anstatt es zur eigentlichen Problemlösung zu nutzen (Wissenschaftsmodus), also zur Erschließung des Inneren der Blackbox.

Raphael zeigt wiederholt Tätigkeiten der Anwendung, was ihm Herstellung, Evaluation und Verbesserung eines Modells ermöglicht, mit dem er selbst zufrieden ist und das er für erklärungsstark hält. Weiterführend könnte die Entwicklung der Modelle und gegebenenfalls ihre tatsächliche Erklärungsstärke beurteilt werden, wobei zu erwarten ist, dass Studierende, die mehr Tätigkeiten der Anwendung umsetzen, auch erklärungsstärkere Modelle entwickeln (SINS et al., 2005). Insgesamt überwiegen allerdings auch bei Raphael Tätigkeiten der Exploration und Herstellung in Häufigkeit und Länge. Mangelndes Vorwissen hinsichtlich der Blackbox (vgl. RUPPERT et al., 2017) sowie die eher offene Aufgabenstellung könnten hier mögliche Gründe für den hohen explorierenden Tätigkeitsanteil bei beiden ProbandInnen darstellen. Außerdem werden in dieser Studie nur gezeichnete Modelle als solche erfasst, darüber hinaus stattfindende gedankliche Modellbildung wird methodisch bedingt nicht berücksichtigt.

### 6.2 Tätigkeitsmuster des Modellierens (Fragestellung 2)

Der Tätigkeitsverlauf von Angelina besteht aus einer Explorations- und einer Herstellungsphase, wobei ihr Modellierungsprozess ohne Anwendungsphase endet. Ihr gelingt es nicht, das Modell losgelöst von ihren Beobachtungen zu betrachten und überprüfbare Vorhersagen zu entwickeln (LOUCA & ZACHARIA, 2015). Da diese fehlen, kann sie ihr Modell nicht testen und gegebenenfalls falsifizieren, obwohl sie die Konsistenz ihres Modells in Frage stellt. Ihr Modellierungsprozess kann damit analog zu den Beschreibungen von KRELL und HERGERT (im Druck) dem expressiven Modellieren (CAMPBELL et al., 2013; OH & OH, 2011) zugeordnet werden.

Raphael dagegen wechselt 17 Mal zwischen Tätigkeiten der Phasen Exploration, Herstellung und Anwendung entsprechend des zyklischen

Charakters des Modellierens (GIERE et al., 2006; GILBERT & JUSTI, 2016; KHAN, 2011; PASSMORE et al., 2014). Sein Tätigkeitsverlauf zeigt die von GILBERT und JUSTI (2016) beschriebene Komplexität und Dynamik sowie Non-Linearität des Modellierens, wobei er zwischen zeitlich abgrenzbaren Phasen der Exploration, Herstellung und Anwendung flexibel wechselt. Ein linear-zyklisches Vorgehen, zum Beispiel nach dem oftmals beschriebenen *Generation-Evaluation-Modification* Kreislauf (z. B. KHAN, 2011), ist nicht erkennbar (vgl. LOUCA & ZACHARIA, 2015).

### 6.3 Modellieren und Modellverstehen (Fragestellung 3)

Der Modellierungsprozess Angelinas zeigt sich konsistent zu ihrem Modellverstehen. Sie beschreibt Modelle als Medien zur Veranschaulichung komplexer Sachverhalte (vgl. JUSTI & GILBERT, 2003; KRELL & KRÜGER, 2016) und nutzt diese auch in ihrem Modellierungsprozess, um ihre Ideen zu visualisieren. Sie leitet keine Vermutungen vom Modell über das Original ab. Raphael zeigt zwar einen deutlich komplexeren Modellierungsprozess, der auch die Anwendung des entwickelten Modells beinhaltet, beschreibt Modelle im Fragebogen aber ebenso als Medien. Für die Annahme, ein ausgeprägtes Modellverstehen unterstütze den Prozess des Modellierens (CHENG & LIN, 2015; LOUCA & ZACHARIA, 2012; SCHWARZ et al., 2009) finden sich somit keine Indizien. Modellverstehen scheint insbesondere keine Limitation für komplexe, zyklische Modellierungsprozesse zu sein. Obwohl einige Autoren (CHENG & LIN, 2015; LOUCA & ZACHARIA, 2012; SCHWARZ et al., 2009) argumentieren, dass ausgeprägtes Modellverstehen Voraussetzung für effektives wissenschaftliches Modellieren ist, zeigt die Arbeit von BAILER-JONES (2002), dass auch WissenschaftlerInnen, die mit Modellen arbeiten, Modelle theoretisch als Medien beschreiben. Hinsichtlich allgemeiner kognitiver Fähigkeiten (*schlussfolgerndes Denken*), sowie Fähigkeiten im Bereich der naturwissenschaftlichen Erkenntnisgewinnung (*wissenschaftliches Denken*) zeichnet sich ein ähnliches Bild. Beide ProbandInnen zeigten in den Vortests überdurchschnittliche Performanz (Tabelle 2), aber unterschiedliche Modellierungsprozesse.

## 7 Fazit & Ausblick

Mithilfe des von KRELL et al. (2017) entwickelten Codierleitfadens konnten Tätigkeiten des Modellierens detailreich und trennscharf erfasst werden. Phasen der Exploration, Herstellung und Anwendung sind in den Modellierungsprozessen gut zeitlich abgrenzbar, sodass auf Grundlage einer größeren Stichprobe allgemeine Strategien des Modellierens aus den Einzelfallbeschreibungen abgeleitet werden können (KELLE & KLUGE, 2010).

Den Einfluss der Hintergrundvariablen (*wissenschaftliches Denken, schlussfolgerndes Denken, Modellverstehen*) auf den Prozess des Modellierens gilt es durch die Auswertung weiterer Fälle zu untersuchen.

## Danksagung

Die Autoren danken den an der Untersuchung beteiligten Studierenden und der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung des Projekts *TypMoL* (Projektnummer 327507949).

## Zitierte Literatur

- BAILER-JONES, D. M. (2002): Scientists' thoughts on scientific models. *Perspectives on Science*, 10(3), 275-301.
- BAUMERT, J. & KUNTER, M. (2013): The COACTIV model of teachers' professional competence. In M. KUNTER, J. BAUMERT, W. BLUM, U. KLUSMANN, S. KRAUSS, & M. NEUBRAND (Hrsg.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*. Boston, MA: Springer US. 8-25.
- BROWN, D. & CLEMENT, J. (1989): Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional science*, 18(4), 237-261.
- CAMPBELL, T., OH, P. S., MAUGHN, M., KIRIAZIS, N. & ZUWALLACK, R. (2015): A review of modeling pedagogies: Pedagogical functions, discursive acts, and technology in modeling instruction. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(1), 159-176.
- CAMPBELL, T., OH, P. S. & NEILSON, D. (2013): Reification of five types of modeling pedagogies with model-based inquiry (MBI) modules for high school science classrooms. In INFORMATION RESOURCES MANAGEMENT ASSOCIATION (Hrsg.), *K-12 Education: Concepts, methodologies, tools, and applications: Development and design methodologies*. Hershey, PA: IGI Global. 401-421.
- CHENG, M.-F. & LIN, J.-L. (2015): Investigating the relationship between students' views of scientific models and their development of models. *International Journal of Science Education*, 37(15), 2453-2475.
- CRAWFORD, B. & CULLIN, M. (2005): Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. BOERSMA, M. GOEDHART, O. DE JONG, & H. EIJKELHOF (Hrsg.), *Research and the quality of Science Education*. Dordrecht: Springer. 309-323.
- GIERE, R., BICKLE, J. & MAULDIN, R. (2006): *Understanding scientific reasoning*. Belmont, CA: Thomson Wadsworth.
- GILBERT, J. & JUSTI, R. (2016): *Modelling-based teaching in science education*. Cham: Springer.
- GÜNTHER, S. L., FLEIGE, J., UPMEIER ZU BELZEN, A. & KRÜGER, D. (2017): Interventionsstudie mit angehenden Lehrkräften zur Förderung von Modellkompetenz im Unterrichtsfach Biologie. In C. GRÄSEL & K. TREMPER (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals: Interdisziplinäre Betrachtungen, Befunde und Perspektiven*. Wiesbaden: Springer VS. 215-236.
- HARTMANN, S., MATHESIUS, S., STILLER, J., STRAUBE, P., KRÜGER, D. & UPMEIER ZU BELZEN, A. (2015): Kompetenzen der naturwissenschaftlichen Erkenntnisgewinnung als Teil des Professionswissens zukünftiger Lehrkräfte: Das Projekt Ko-WADiS. In B. KOCH-PRIEWE, A. KÖKER, J. SEIFRIED, & E. WUTTKE (Hrsg.), *Kompetenzerwerb an Hochschulen: Modellierung und Messung*. Bad Heilbrunn: Julius Klinkhardt KG. 39-58.
- JUSTI, R. & GILBERT, J. (2003): Teachers' views on the nature of models. *International Journal of Science Education*, 25(11), 1369-1386.
- JUSTI, R. & VAN DRIEL, J. (2005): The development of science teachers' knowledge on models and modelling: promoting, characterizing, and understanding the process. *International Journal of Science Education*, 27(5), 549-573.
- KELLE, U. & KLUGE, S. (2010): *Vom Einzelfall zum Typus: Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung. 2., überarb.* Wiesbaden: VS Verlag.

- KHAN, S. (2007): Model-based inquiries in chemistry. *Science Education*, 91, 877-905.
- KHAN, S. (2011): What's missing in model-based teaching. *Journal of Science Teacher Education*, 22(6), 535-560.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland]. (2017): *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. München: Wolters Kluwer.
- KNUUTTILA, T. (2005): Models, representation, and mediation. *Philosophy of Science*, 72(5), 1260-1271.
- KOCH, S., KRELL, M. & KRÜGER, D. (2015): Förderung von Modellkompetenz durch den Einsatz einer Blackbox. *Erkenntnisweg Biologiedidaktik*, 14, 93-108.
- KRELL, M. & HERGERT, S. (im Druck): The blackbox approach: Analyzing modeling strategies. In A. UPMEIER ZU BELZEN, D. KRÜGER, & J. VAN DRIEL (Hrsg.), *Towards a competence-based view on models and modeling in science education*: Springer.
- KRELL, M. & KRÜGER, D. (2013): Wie werden Modelle im Biologieunterricht eingesetzt? Ergebnisse einer Fragebogenstudie. *Erkenntnisweg Biologiedidaktik*, 12, 9-26.
- KRELL, M. & KRÜGER, D. (2016): Testing models: A key aspect to promote teaching activities related to models and modelling in biology lessons? *Journal of Biological Education*, 50(2), 160-173.
- KRELL, M., UPMEIER ZU BELZEN, A. & KRÜGER, D. (2014): Students' levels of understanding models and modelling in biology: Global or aspect-dependent? *Research in Science Education*, 44(1), 109-132.
- KRELL, M., UPMEIER ZU BELZEN, A. & KRÜGER, D. (2016): Modellkompetenz im Biologieunterricht. In A. SANDMANN & P. SCHMIEMANN (Hrsg.), *Biologiedidaktische Forschung: Schwerpunkte und Forschungsgegenstände*. Berlin: Logos. 83-102.
- KRELL, M., WALZER, C., HERGERT, S. & KRÜGER, D. (2017): Development and application of a category system to describe pre-service science teachers' activities in the process of scientific modelling. *Research in Science Education*.
- LANDIS, J. R. & KOCH, G. G. (1977): The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- LAUBICHLER, M. D. & MÜLLER, G. B. (2007): *Modeling biology: Structures, behavior, evolution*. Wien: MIT Press.
- LEIGHTON, J. & GIERL, M. (2007): *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge, NY: Cambridge University Press.
- LEISNER-BODENTHIN, A. (2006): Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 91-109.
- LIEPMANN, D., BEAUDUCCEL, A., BROCKE, B. & AMTHAUER, R. (2007): *Intelligenz-Struktur-Test 2000 R: IST 2000 R*. Göttingen: Hogrefe.
- LÖHNER, S., VAN JOOLINGEN, W. R., SAVELSBERGH, E. R. & VAN HOUT-WOLTERS, B. (2005): Students' reasoning during modeling in an inquiry learning environment. *Computers in Human Behavior*, 21(3), 441-461.
- LOUCA, L. T. & ZACHARIA, Z. C. (2012): Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471-492.
- LOUCA, L. T. & ZACHARIA, Z. C. (2015): Examining learning through modeling in k-6 science education. *Journal of Science Education and Technology*, 24(2), 192-215.
- MAYRING, P. (2010): Qualitative inhaltsanalyse. In G. MEY & K. MRUCK (Hrsg.), *Handbuch qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag. 601-613.
- NEHRING, A., NOWAK, K. H., UPMEIER ZU BELZEN, A. & TIEMANN, R. (2015): Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic variables. *International Journal of Science Education*, 37(9), 1343-1363.
- NICOLAOU, C. T. & CONSTANTINOU, C. P. (2014): Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13(1), 52-73.
- ODENBAUGH, J. (2005): Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology and philosophy*, 20(2), 231-255.
- OH, P. S. & OH, S. J. (2011): What teachers of science need to know about models: An overview. *International Journal of Science Education*, 33(8), 1109-1130.

- ORSENNE, J. (2015): *Aktivierung von Schülervorstellungen zu Modellen durch praktische Tätigkeiten der Modellbildung*. Humboldt-Universität, Berlin. Abgerufen unter: <https://edoc.hu-berlin.de/bitstream/handle/18452/18164/orsenne.pdf?sequence=1>.
- PASSMORE, C., GOUVEA, J. S. & GIÈRE, R. (2014): Models in science and in learning science: Focusing scientific practice on sense-making. In M. R. MATTHEWS (Hrsg.), *International handbook of research in history, philosophy and science teaching*. Dordrecht: Springer. 1171-1202.
- RUPPERT, J., DUNCAN, R. G. & CHINN, C. A. (2017): Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education*.
- SANDMANN, A. (2014): Lautes Denken die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. KRÜGER, I. PARCHEMANN, & H. SCHECKER (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin; Heidelberg: Springer. 179-188.
- SCHAUBLE, L., KLOPFER, L. E. & RAGHAVAN, K. (1991): Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28(9), 859-882.
- SCHREIER, M. (2010): Fallauswahl. In G. MEY & K. MRUCK (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag für Sozialwissenschaften, 238-251.
- SCHWARZ, C., REISER, B., DAVIS, E., KENYON, L., ACHÉR, A., FORTUS, D. ET AL. (2009): Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632-654.
- SEIDEL, T., PRENZEL, M., DUIT, R. & LEHRKE, M. (2003): *Technischer Bericht zur Videostudie "Lehr-Lern-Prozesse im Physikunterricht"*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN).
- SHULMAN, L. S. (1986): Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- SINS, P. H., SAVELSBERGH, E. R. & VAN JOOLINGEN, W. R. (2005): The difficult process of scientific modelling: An analysis of novices' reasoning during computer based modelling. *International Journal of Science Education*, 27(14), 1695-1721.
- UPMEIER ZU BELZEN, A. (2014): Black Box: Modellierung von Prozessen naturwissenschaftlicher Erkenntnisgewinnung. In D. LUDWIG, C. WEBER, & O. ZAUZIG (Hrsg.), *Das materielle Modell. Objektgeschichten aus der wissenschaftlichen Praxis*. Paderborn: Wilhelm Fink. 99-106.
- UPMEIER ZU BELZEN, A. & KRÜGER, D. (2010): Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41-57.
- VAN DRIEL, J. & VERLOOP, N. (1999): Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21(11), 1141-1153.
- VERBI Software. (2018): MAXQDA - Software für qualitative Datenanalyse. Berlin: Consult. Sozialforschung GmbH.
- WINDSCHITL, M., THOMPSON, J. & BRAATEN, M. (2008): Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941-967.



## Preservice Science Teachers' Strategies in Scientific Reasoning: the Case of Modeling

Maximilian Göhner<sup>1</sup>  · Moritz Krell<sup>1</sup> 

Published online: 22 July 2020  
© The Author(s) 2020

### Abstract

The development of scientific reasoning competencies is a key goal of science education. To better understand the complex construct of scientific reasoning, which includes modeling as one style of reasoning, thorough investigations of the underlying processes are needed. Therefore, in this study, a typology of preservice science teachers' modeling strategies was developed. Thirty-two preservice science teachers were videotaped while engaging in the modeling task of investigating a black box. Following a qualitative content analysis, sequences of modeling activities were identified. By transforming these sequences of modeling activities into state transition graphs, six types of modeling strategies were derived, differing in the homogeneity and complexity of their modeling processes. The preservice science teachers engaged in activities of (1) exploration only; (2a) exploration and development with a focus on development; (2b) exploration and development with a focus on exploration; (2c) exploration and development, balanced; (3a) exploration, development, and drawing predictions from a model once; or (3b) exploration, development, and repeatedly drawing predictions from a model. Finally, this typology is discussed regarding the process of its development and its potential to inform and guide further research as well as the development of interventions aiming to foster competencies in scientific modeling.

**Keywords** Scientific reasoning · Modeling · Preservice science teachers · Typology · Modeling strategies · Qualitative content analysis

### Introduction

The development of scientific reasoning competencies is considered a key goal of twenty-first century education (Osborne 2013), as those competencies are required for active participation in science- and technology-centered societies. Consequently, scientific reasoning

---

✉ Maximilian Göhner  
maximilian.goehner@fu-berlin.de

<sup>1</sup> Biology Education, Freie Universität Berlin, Schwendenerstraße 1, Berlin 14195, Germany

competencies are highlighted in curricular documents worldwide (e.g., BCMOE 2019; KMK 2005; NGSS Lead States 2013; VCAA 2016). To support the development of students' competencies, preservice science teachers need to obtain scientific reasoning competencies as part of their professional competencies (Khan and Krell 2009; Mathesius et al. 2016). These scientific reasoning competencies can be defined as “a complex construct, which encompass[es] the abilities needed for scientific problem solving as well as to reflect on this process at a meta-level” (Krell et al. 2018, p. 2; cf. Morris et al. 2012), including generic cognitive abilities, such as the use of analogies and decision making (Nersessian 2002), and the ability to apply content, procedural, and epistemic knowledge for problem solving (Kind and Osborne 2017). In line with this definition, research on scientific reasoning competencies is a similarly complex matter, leading to a high diversity of studies concentrating on different contexts and foci and, consequently, using different theoretical frameworks and methodical approaches (Kosłowski 2013; Rönnebeck et al. 2016). Additionally, it is argued that there is no single form of scientific reasoning; Kind and Osborne (2017), for example, have recently proposed six different styles of scientific reasoning, which include mathematical deduction, experimental evaluation, hypothetical modeling, categorization and classification, probabilistic reasoning, and historical-based evolutionary reasoning. For some styles of scientific reasoning, especially experimental evaluation (Hammann et al. 2008; Schauble et al. 1991a), individuals' processes, strategies, and understandings have been broadly investigated and described in science education research, but others, like hypothetical modeling, lack detailed research (Nicolaou and Constantinou 2014). Hence, a thorough investigation of the processes involved in different styles of scientific reasoning, and their comparison could lead to deeper insights into the structure of the complex construct of scientific reasoning and would help develop a “big picture of what [scientific reasoning] might be” (Kind and Osborne 2017, p. 27). The present study focuses on the style of *hypothetical modeling* and adds to the research on scientific reasoning, by presenting a process-oriented analysis of preservice science teachers' strategies in the process of scientific modeling.

### Analyzing Scientific Reasoning Processes

Research on the dynamic and process-oriented aspects of behavior often relies on the systematic observation of predefined behavior codes (Bakeman and Gottman 1997), which are used as the fundamental elements of analysis. In research regarding scientific reasoning, the term “activities” is most commonly used to describe these elements of analysis (e.g., Khan and Krell 20019; Krell et al. 2019). Thus, different frameworks with specific sets of activities can be flexibly utilized to account for varying settings, samples or goals (cf. Rönnebeck et al. 2016), levels of detail (e.g., Chinn and Malhotra 2002; Fischer et al. 2014; Lawson 2004), or theoretical emphases on specific styles of reasoning (e.g., on modeling; Giere et al. 2006). By observing and identifying the different epistemic activities of individuals in their respective order, the sequences of activities can be extracted, which simplify the complex process of scientific reasoning to a level where patterns can be made visible. To date, empirical studies have shown that the activities of scientific reasoning do not follow a strictly predetermined order (e.g., Harwood 2004; Klahr and Dunbar 1988; Reiff et al. 2002), so scientific reasoning can be considered a creative problem-solving process (Zimmerman 2005). However, systematic differences in the reasoning processes of novices and experts have been found (e.g., Čančula et al. 2015; Klahr 2002).

## How Typologies of Scientific Reasoning Can Be Beneficial for Science Education

Many studies aim to generalize patterns in scientific reasoning processes by proposing types or classes of reasoning strategies, which can then be used for diagnostic purposes or in targeted interventions (Klahr et al. 1993; Schauble et al. 1991b). One example of an analysis of scientific reasoning processes that can be seen as a central framework in science education research is the work of Klahr and Dunbar (1988), which has been used and adapted in numerous further studies (e.g., Aizpurua et al. 2018; Hammann et al. 2008; Krell 2018; Neumann et al. 2019). Based on case studies, Klahr and Dunbar (1988) developed the *Scientific Discovery as Dual Search* (SDDS) model. They argue that scientific reasoning is a problem-solving task taking place in two problem spaces: the hypothesis space and the experiment space. They distinguish two strategies based on the occurrence of an experimental activity at a specific moment in the scientific reasoning process, classifying subjects into “experimenters” and “theorists.” They further describe the differences regarding the occurrence and frequency of activities assigned to one space or the other between these two classes. When given a scientific problem, “theorists” search for new hypotheses in the hypothesis space, while “experimenters” rely on experimentation to induce regularities without explicitly stating other hypotheses. Studies have indicated that experts rather behave like “theorists” and tend to search the hypothesis space, which could result from differences in prior knowledge (e.g., Rasmussen 1981), allowing them to solve scientific problems more efficiently (Klahr 2002). Based on the SDDS model, Klahr et al. (1993) described commonly found differences in reasoning processes of adults and children. They identified different problems children encounter in the development of scientific reasoning skills. For example, children often have problems constraining their search in the hypothesis or experiment space, leading to the testing of inappropriate hypothesis or design of inappropriate experiments. These findings, which were based on the application of the SDDS model and their typology of “experimenters” and “theorists,” allowed them to draw implications for educational settings (e.g., should educational settings by design be more constrained in regard to the amount of hypotheses and possible experiments if they target children), which were later incorporated into further studies (Klahr and Nigam 2004).

Another study combining the identification of experimentation activities with the generalization of these patterns was implemented by Schauble et al. (1991a). They investigated strategy differences in good and poor learners in an experimental setting, focusing on five “classes of student behavior.” These classes include the general activity level and the four activities of planning, evidence generation, data management, and evidence interpretation. The authors distinguish between an “engineering” and a more advanced “scientific” model of investigating scientific phenomena. Subjects utilizing the engineering model carry out an investigation until a specific effect is reached and then stop, while subjects utilizing the scientific model examine all variables and their influence on a given system. Additionally, good learners use a control of variables strategy more frequently, generate more hypotheses, keep more systematic records, and develop more goal-oriented plans than poor learners. The authors further used the distinction between an engineering and a scientific model of experimentation to develop a learning progression in which students were able to stepwise develop their individual experimentation processes to increasingly resemble the latter (Schauble et al. 1991b). This effect was observed to be stronger if the problems to be solved were introduced in an order of increasing analytical complexity.

Both examples illustrate the fruitfulness of typologies of scientific reasoning for science education; typologies might guide further research and contribute to theory development (e.g., Klahr and Dunbar 1988) as well as be used to develop specific educational settings and interventions (Schauble et al. 1991a, b; Klahr et al. 1993; Klahr and Nigam 2004). However, while different strategies (and related typologies) have been proposed for scientific experimentation, such findings are widely missing for scientific modeling.

### Scientific Modeling as One Style of Scientific Reasoning

The relationship between scientific modeling and scientific reasoning is extensively discussed in science education research. As introduced above, six styles of scientific reasoning have been proposed (Kind and Osborne 2017), with each style having a specific set of ontological, procedural, and epistemic resources needed for reasoning. Kind and Osborne (2017) further argue that the six styles of scientific reasoning offer a comprehensive schema for the construct of scientific reasoning. The focus of this study is on one style that is quite established in science education research: scientific modeling (e.g., Clement 2000; Gilbert 2004; Windschitl et al. 2008). Although described as one of the six styles by Kind and Osborne (2017), various other authors specifically emphasize the importance of modeling for reasoning in science; Giere (1999), for example, claims that “scientific reasoning is to a large extent model-based reasoning” (p. 56). Similarly, Lehrer and Schauble (2015) describe scientific reasoning (or scientific thinking, respectively) as a modeling process that encompasses various other practices. Thus, modeling can also be seen as an overarching ability, needed for all processes of scientific reasoning. For this study, however, the specific relationship between modeling and other styles of reasoning in science (i.e., whether modeling is one of various styles or an overarching style encompassing the others) is not of central importance. Therefore, in the present article, scientific modeling will be considered as one style of scientific reasoning and thus needs to be specifically investigated in science education.

From an educational perspective, modeling has the potential to contribute to three of the main objectives of science education (cf. Hodson 2014): *learning science* by learning major models as products of science; *learning to do science* by developing practices to create and evaluate models; and *learning about science* by appreciating the role of models as hypothetical entities and epistemic tools in science (Justi and Gilbert 2003). Consequently, modeling is also highlighted in recent curricular documents as a core scientific practice (e.g., NGSS Lead States 2013), and it has been widely acknowledged that research-based learning essentially consists of the construction, testing, revision, and application of models (e.g., Schwarz et al. 2017; Windschitl et al. 2008). Therefore, scientific modeling has been identified as a potentially useful and relevant example of scientific reasoning and will be the focus of the present article.

### Towards a Typology of Modeling Strategies in Science Education

To identify modeling strategies, the present study can build on an extensive body of research on models and modeling in science education, whereby a lot of studies were published already during the 1980s (e.g., Brown and Clement 1989; Gilbert and Osborne 1980; Nersessian 1992) or more recently (e.g., Clement and Rea-Ramirez 2008; Krell et al. 2019; Passmore et al. 2014). In these studies, different facets of knowledge and abilities related to models and modeling of various sample groups have been investigated, such as those of students (e.g., Grosslight et al. 1991; Krell et al. 2014a, b), teachers (van Driel and Verloop 1999; Windschitl

et al. 2008), and experts (e.g., Bailer-Jones 2002; Clement 2008). Other studies have focused on model-based learning and teaching, that is, how to use modeling approaches to enable students to better understand scientific concepts (e.g., Acher et al. 2007; Passmore et al. 2009).

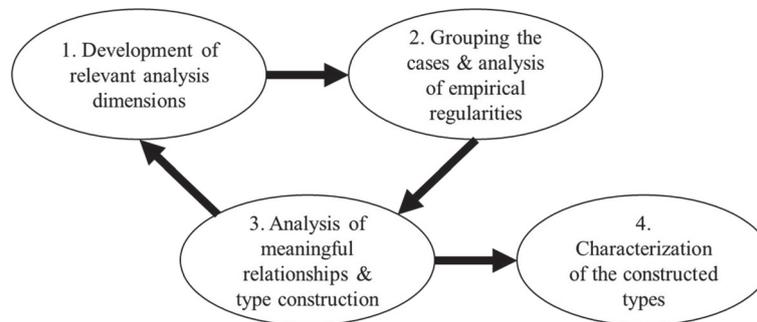
In all these studies, one common finding is that most sample groups seem to have difficulties in understanding the predictive power of models and their role as epistemic tools for scientific reasoning (Krell and Krüger 2016; Passmore et al. 2014). Additionally, even when models are used and understood as epistemic tools, multiple models representing alternative hypotheses are rarely considered (Grosslight et al. 1991). However, it is precisely this characteristic of models as epistemic tools for prediction that is emphasized in science education curricula and standard documents worldwide (e.g., BCMOE 2019; KMK 2005; NGSS Lead States 2013; VCAA 2016). Moreover, reviews propose that research regarding models and modeling in science education mostly focuses on cognitive and metacognitive aspects, so that “meta-modeling knowledge is overevaluated” (Nicolaou and Constantinou 2014, p. 72); detailed and process-oriented analyses of modeling strategies, which could be used for diagnostic purposes or interventions (cf. Klahr et al. 1993; Schauble et al. 1991b), are mostly missing in science education research (Louca and Zacharia 2012; Nicolaou and Constantinou 2014). One exception is the study by Sins et al. (2005), who investigated the types of reasoning, foci of conversation, and types of argumentation used by novices during computer-based modeling. The authors found that more successful students tended to justify their reasoning with prior knowledge and viewed the model as a whole, while the less successful students showed model-fitting behavior (cf. engineering model of experimentation; Schauble et al. 1991a). While the study of Sins et al. (2005) shows how the analysis of sequential activities can be further quantified by using, for example, *z*-scores, they did not distinguish between different strategies or types of modeling processes, which could potentially provide further insights into the nature of different modeling strategies applied by, for example, preservice science teachers.

The present article aims to fill this gap by developing a typology of modeling strategies of preservice science teachers, which can then be further used in science education to distinguish individual modeling processes and enable researchers and teachers to identify points of leverage for interventions on an individual level.

## Methodological Frame of the Study and Research Questions

To ensure the quality of the analysis and enable comparisons between the present study and the processes and strategies of reasoning involved in other styles of reasoning, the methods of analysis and the development of the typology and its interpretation must be made as transparent as possible. Therefore, different evaluation criteria for the classification and development of typologies are considered. The analysis is evaluated by the appropriateness of the considered attributes, the minimization of within-group variance, and the maximization of between-group variance (Bailey 1994). The model of empirically grounded type construction (Kluge 2000) is applied during the analysis to maintain methodical flexibility, while achieving a systematic and transparent development process for the typologies. The model describes four steps in the type construction process (Fig. 1). The first three steps can be repeated multiple times to evaluate the appropriateness of possible (combinations of) attributes.

Hence, in the present study, relevant analysis dimensions will be identified (step 1), grouped, and investigated in terms of the empirical regularities (step 2) and meaningful



**Fig. 1** Model of empirically grounded type construction (Kluge 2000)

relationships (step 3) within modeling processes to identify different types that can be characterized as modeling strategies (step 4). More specifically, the following four research questions will be discussed, whereby each question relates to one step of the model of empirically grounded type construction (Fig. 1):

- 1) Which attributes are appropriate for developing an empirically grounded typology of preservice science teachers' modeling strategies (cf. Fig. 1, step 1)?
- 2) To what extent can modeling processes of preservice science teachers be appropriately grouped based on the selected attributes and consideration of the criteria of minimizing within-group heterogeneity and maximizing between-group heterogeneity (cf. Fig. 1, step 2)?
- 3) To what extent can meaningful relationships be found between cases within the identified groups as well as between the identified groups (cf. Fig. 1, step 3)?
- 4) Which types of preservice science teachers' modeling strategies can be derived from the type construction process (cf. Fig. 1, step 4)?

## Methods

### Context

The context of this study is the first phase of science teacher education in Germany, which takes place at the university. In this first phase, science teachers are requested to develop basic professional knowledge and competencies that they will need as science teachers (Neumann et al. 2017), including knowledge and competencies related to inquiry and reasoning in science (KMK 2019). In Germany, preservice science teachers usually study two subjects (i.e., two future teaching subjects) within a bachelor's (six semesters) and a subsequent master's program (four semesters) before leaving university and entering the second phase of teacher education (internship). Existing empirical studies propose a significant positive development of German preservice science teachers' scientific reasoning competencies over their course of studies (e.g., Hartmann et al. 2015; Krüger et al. 2020). Students in later semesters are, therefore, expected to have advanced competencies related to inquiry and reasoning in science.

## Setting

To induce scientific reasoning processes, a black box setting was applied (Lederman and Abd-El-Khalick 2002). The black box approach is established in science education research to study processes of scientific thinking and modeling (Khan and Krell 2019; Lederman and Abd-El-Khalick 2002; Passmore and Svoboda 2012). In such approaches, the black box represents a natural phenomenon under study, and the exploration of the black box represents the process of scientific discovery. It was shown that black box approaches are suitable to elicit modeling processes, with models used as epistemic tools to discover the black box (Krell et al. 2019; Passmore and Svoboda 2012). In this study, the black box approach was preferably chosen over an authentic and content-rich scientific problem, in order to reduce the influence of prior knowledge on the preservice science teachers' modeling processes. However, it is recognized that this might be one limitation of the study as some authors emphasize the important role of content knowledge for scientific modeling (e.g., Ruppert et al. 2019). In the present article, a water black box was used. It can be explored by filling the black box with water (input), which then results in measurable outputs of water (see Krell et al. 2019 for a detailed description of the black box).

Participation in the black box activity was voluntary and not connected to any university courses or obligatory parts of the curriculum. Researchers and participants had no formal relationships to one another. Each participant was briefly introduced to the study and signed an informed consent. To gain additional insights into the scientific reasoning processes, all participants were asked to concurrently think aloud (Leighton and Gierl 2007), which was practiced with three short questions. After these preparations, the participants were brought into a room equipped with three video cameras, the black box, some pre-filled glasses of water, and a chalkboard. The researcher briefly explained the basic functionality of the black box using a prepared script and provided the following task: "Draw a model of the inside of the black box." Participants were informed that there are no time constraints (mean length of the activity for all participants: 1 h and 11 min). The first author stayed in the room during the process of modeling to prevent any technical errors and, if necessary, to remind the participants to concurrently think aloud; otherwise, he did not intervene.

## Sampling

Data from 32 participants were collected and analyzed for the present study. The participants' academic progress ranged from first (i.e., bachelor) to tenth semester (i.e., master). The participants were aged between 17 and 39 (mean age was 25). To further maximize the heterogeneity of participants and make the observation of different strategies more likely, theoretical sampling was used. Preservice science teachers were selected based on two criteria. The first criterion was the preservice science teachers' achievement in an established multiple-choice test instrument, which has been previously developed to assess scientific reasoning competencies and encompasses tasks about scientific investigations and scientific modeling (Krüger et al. 2020). The second criterion was the preservice science teachers' achievement in the so called I-S-T 2000R instrument (Liepmann et al. 2007). This paper-pencil instrument assesses general cognitive abilities. It also includes tasks of spatial transformation, which is being argued to be one ability needed for modeling (Nersessian 2002). Both criteria were assessed before participants were invited to the black box activity, and only those preservice science

teachers differing one standard deviation from the tested population's mean scores on both tests were asked to participate.

### Data Analysis

The participants' videos were transcribed verbatim, including all verbalizations as well as behavioral aspects deemed important for the analysis (these included making an input or observing an output, going to the board, or cleaning the board). The videos and corresponding transcripts were analyzed by the first author and one additional rater (trained student assistant) following a qualitative content analytic approach (Schreier 2012). To improve the comparability of the results and enhance the workflow, the transcripts were coded watching the videos simultaneously.

Each participant's modeling process was split into individual activities of scientific modeling. For this purpose, a framework consisting of 19 different activities of modeling was utilized (Table 1, i.e., category system), which was developed in a previous study (Khan and Krell 2019; Krell et al. 2019) and integrates frameworks used to describe the processes of analogical reasoning (Brown and Clement 1989) and scientific reasoning through models (Giere et al. 2006). This framework also allows the assignment of every activity to one of the three modeling phases of exploration, model development, or making predictions. While this previously developed category system includes all activities of modeling that have been identified within the whole sample of the previous study (Khan and Krell 2019; Krell et al. 2019), the present study further applies this category system in order to find patterns and identify types for individual preservice teachers' modeling strategies based on the sequence of these modeling activities. Additionally, it is worth mentioning that analogies (Code 6) can be activated throughout the whole process wherein participants are able to evaluate their ideas and mental models verbally; however, the methodical approach in the present study requires an externalized model in form of a drawing.

**Table 1** Category system with 19 modeling activities (Krell et al. 2019)

Number	Code	Phase
1	Perception of a phenomenon	Exploration
2	Exploration of the system: input/output (exploratory)	
3	Exploration of the system: summarizing/describing observations	
4	Exploration of the system: input/output (pattern detection)	
5	Exploration of the system: recognizing patterns	
6	Activation of analogies and experiences	
7	Development of model: graphically develop model	Development
8	Development of model: change model to optimize consistency	
9	Development of model: change model to optimize representation	
10	Development of model: reject model due to poor consistency/representation	
11	Evaluation of consistency and representation: evaluate consistency	
12	Evaluation of consistency and representation: evaluate representation	
13	Finding of consistency and representation	Prediction
14	Deduction of predictions	
15	Evaluation of predictions: input/output (to test predictions)	
16	Evaluation of predictions: confirmation of prediction	
17	Evaluation of predictions: falsification of prediction	
18	Modification/rejection of model: change model due to falsified predictions	
19	Modification/rejection of model: reject model due to falsified predictions	

Table 2 exemplifies how the transcripts were divided into single activities for a short transcript excerpt. This excerpt covers minutes 24 to 30 of one respondent's 66-min-long modeling process. Hence, the excerpt begins in the middle of the modeling process, after an initial exploration phase and the development of an initial model.

To ensure the reliability of the analysis (intrarater agreement), every video and corresponding transcript was coded twice by the first author. To ensure the intersubjectivity of the analysis (interrater agreement), the videos and transcripts were then coded again by the trained student assistant, followed by a discussion and resolution of any disagreements, which might have occurred. Cohen's Kappa suggests acceptable to high agreements (Table 3). Disagreements were discussed and resolved between both raters. The identified sequences of activities were then used for further analysis related to the steps of the model of empirically grounded type construction (Fig. 1). The respective approaches to data analysis are, therefore, explained in the sections below.

## Developing a Typology of Modeling Strategies

In the following, the findings are presented according to the research questions, which relate to the steps of the model of empirically grounded type construction. Two cycles will be presented

**Table 2** An excerpt of the transcript of one participant, exemplarily showing how the transcripts were divided into single activities

Time [min:s]	Paragraph	Text	Code
24:46	29	Let us see how much water I get back. [input and observing the output] Ok, the missing 100 came out, after I put in 2900 ml in total	2 exploration of the system: input/output (exploratory)
25:14	30	Now it should be empty again	3 exploration of the system: summarizing/describing observations
25:58	31	[input] I put in 400 again to fill up the first glass [observes output]	4 exploration of the system: input/output (pattern detection)
26:19	32	Ok, this could be what happened	5 exploration of the system: Recognizing patterns
26:43	33	[going to the board] Ok, now I will try to draw this step by step. At first, there is a chamber containing 400 ml. [leaves the board]	7 development of model: graphically develop model
27:11	34	I will put in 100 ml, to see if I can get it to overflow.	14 deduction of predictions
27:53	35	[input of 100 ml] Yes, something is overflowing, let us see if these are 100 ml. [observing the output]	15 evaluation of predictions: input/output (to test predictions)
28:24	36	Ok, either the chamber fits 50 more or the 50 ml went somewhere else	17 evaluation of predictions: Falsification of prediction
28:52	37	I will try again with the 100, let us see how much comes out. [input, observing the output] It is 50 again	2 exploration of the system: input/output (exploratory)
29:13	38	[going to the board] I get nearly half of it back, which would suggest that the chamber fits 400 and when it overflows one half is going out and another half in a different chamber. [leaving the board]	18 modification/rejection of model: change model due to falsified predictions

Transcribed behavioral aspects (taken from the video) are added in brackets

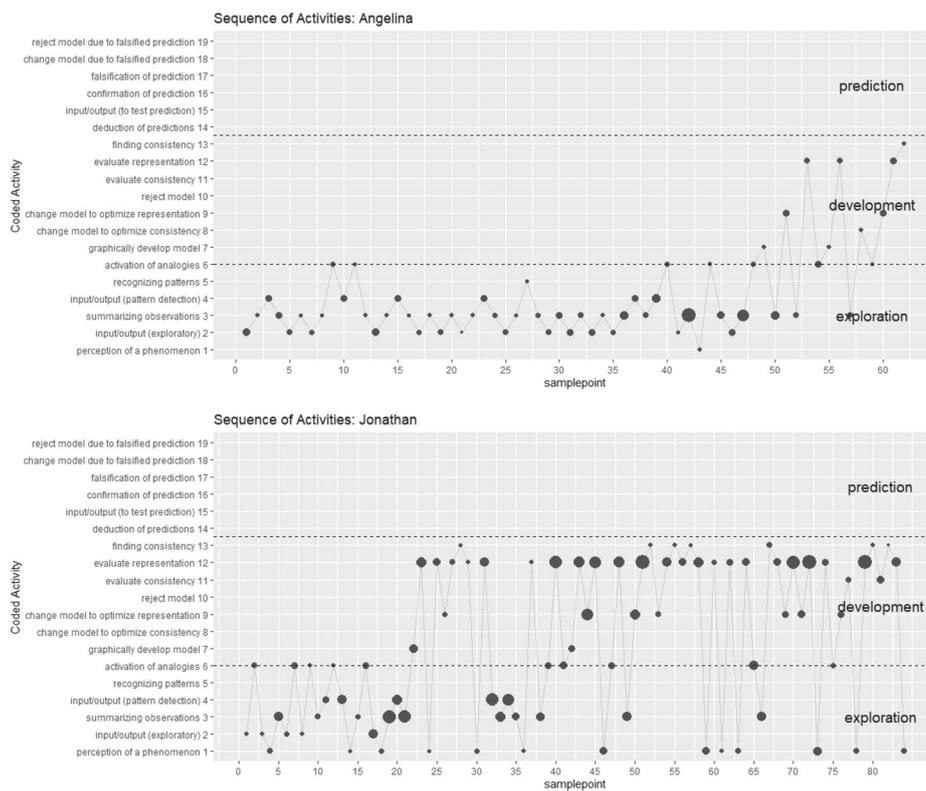
**Table 3** Overview of Cohen's Kappa calculations

	Min	Max	Mean
Intrarater-agreement	0.67	0.95	0.84
Interrater-agreement	0.45	0.93	0.77

focusing on the occurrence of activities (cycle 1) and the transitions between activities (cycle 2), respectively.

### Cycle 1, Step 1: Occurrence of Activities as a Relevant Analysis Dimension

Based on the studies described above, the occurrences of activities (Klahr and Dunbar 1988; Schauble et al. 1991a) were identified as the first relevant dimension of analysis. In the first cycle of the development of the typology, the sequence of activities conducted by the



**Fig. 2** Codelines of Angelina (top) and Jonathan (bottom). The conducted activities are shown in chronological order, whereby the size of each data point corresponds to the time spent on each activity. The dotted lines indicate the assignment of the activities to the three phases of exploration, model development, and prediction. While both participants engage in activities of exploration and model development, Angelina mostly focuses on activities of exploration, with activities of model development at the end. In contrast, Jonathan frequently switches between activities of exploration and model development. However, both Angelina and Jonathan conduct no activities of prediction

participants was visualized as codelines (see, e.g., Fig. 2), which is common practice in science education research (e.g., Louca and Zacharia 2015).

### **Cycle 1, Step 2: Grouping the Cases and Analysis of Empirical Regularities Based on the Occurrence of Activities**

Based on the visual assessment of these codelines, three groups of participants can be easily distinguished:

- (1) One single participant conducted only activities of exploration and did not draw a model at all.
- (2) Seventeen participants conducted activities of exploration and model development but did not derive any predictions from their models (cf. Fig. 2).
- (3) Fourteen participants conducted activities of exploration, model development, and prediction.

As the affiliation of the participants to their group can be directly identified from the occurrence of activities, no further interpretation was needed, suggesting a sufficient degree of between-group heterogeneity. Comparing the sequences of activities of participants in the same group, however, suggests rather low within-group homogeneity, as the participants show major differences in the proportion of activities conducted as well as their sequential order. For example, Angelina shows a long sequence of activities of exploration, developing a model in the very end (Fig. 2 top), while Jonathan shows activities of model development throughout, alternating with activities of exploration (Fig. 2 bottom). Consequently, a further differentiation of the groups seemed to be needed, and no further analysis of meaningful relationships and type construction based on the occurrence of activities was conducted (cf. Fig. 1; step 3).

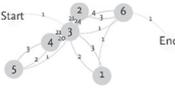
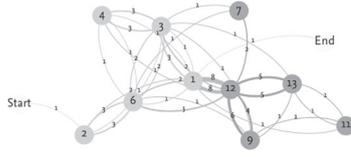
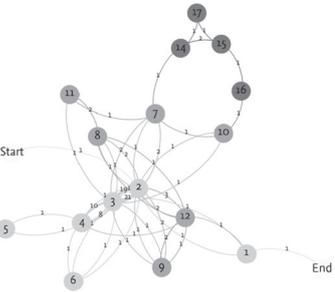
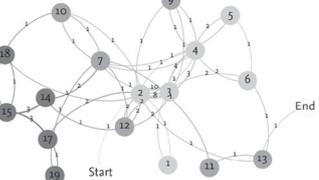
### **Cycle 2, Step 1: Transitions Between Activities as a Relevant Analysis Dimension**

Because a quantitative analysis such as clustering would only account for the amount and time the activities were conducted, the transitions between single modeling activities were selected as relevant attributes to take the sequential order of activities into account in a second cycle of developing the typology (cf. Sins et al. 2005). To visualize the transitions, the codelines depicting the sequence of activities performed by every participant were transformed into state transition graphs (Andrienko and Andrienko 2018). These state transition graphs show all nineteen identified activities as knots that are connected by directed edges of different weights accounting for the quantity of transitions between these knots. Additionally, two knots for “start” and “end” are appended. All state transition graphs are automatically laid out in Gephi (Bastian et al. 2009) using the force-directed graph layout algorithm ForceAtlas2 (Jacomy et al. 2014).

### **Cycle 2, Step 2: Grouping the Cases and Analysis of Empirical Regularities Based on the Transitions Between Activities**

The resulting graphs were then grouped by two raters in a first visual analysis. Six groups could be identified (Table 4), which further differentiate the three initial groups from step 1. Participants engaged in activities of modeling by the following:

**Table 4** Examples of state transition graphs for each identified group

Modeling strategy	Example	Pseudonym
(1) exploration only	 A state transition graph with 6 nodes (1-6) and a Start/End node. Node 1 is the Start node, and node 6 is the End node. Edges connect nodes 1-2, 1-3, 1-4, 1-5, 2-3, 3-4, 4-5, 5-6, and 6-End.	Carlo
(2a) exploration and development with a focus on development	 A state transition graph with 11 nodes (1-11) and a Start/End node. Node 1 is the Start node, and node 11 is the End node. Edges connect nodes 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, and 11-End.	Kara
(2b) exploration and development with a focus on exploration	 A state transition graph with 13 nodes (1-13) and a Start/End node. Node 1 is the Start node, and node 13 is the End node. Edges connect nodes 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12, 12-13, and 13-End.	Angelina
(2c) exploration and development, balanced	 A state transition graph with 13 nodes (1-13) and a Start/End node. Node 1 is the Start node, and node 13 is the End node. Edges connect nodes 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12, 12-13, and 13-End.	Jonathan
(3a) exploration, development and drawing predictions from a model once	 A state transition graph with 16 nodes (1-16) and a Start/End node. Node 1 is the Start node, and node 16 is the End node. Edges connect nodes 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12, 12-13, 13-14, 14-15, 15-16, and 16-End.	Ben
(3b) exploration, development and drawing predictions from a model repeatedly	 A state transition graph with 18 nodes (1-18) and a Start/End node. Node 1 is the Start node, and node 18 is the End node. Edges connect nodes 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12, 12-13, 13-14, 14-15, 15-16, 16-17, 17-18, and 18-End.	Raphael

The numbers in the circles illustrate the respective activities (Table 2), and the numbers on the lines show how often the activities were done sequentially. Activities of exploration, model development, or prediction are highlighted in different shades of gray

(1) Exploration only ( $n = 1$ )

(2a) Exploration and development with a focus on development ( $n = 1$ )

(2b) Exploration and development with a focus on exploration ( $n = 11$ )

- (2c) Exploration and development, balanced ( $n = 5$ )
- (3a) Exploration, development, and drawing predictions from a model once ( $n = 9$ )
- (3b) Exploration, development, and repeatedly drawing predictions from a model ( $n = 5$ )

Discussion of the results between both raters showed an agreement in the grouping of 23 out of 27 cases (85%), suggesting high between-group heterogeneity. Moreover, in the discussion of the grouping process, both raters considered size, complexity (number of knots and edges), the homogeneity of the edges, and edge weights in their analysis of the state transition graphs. Because complexity is mostly represented by the occurrence of the different activities, which was the attribute of analysis in cycle 1, homogeneity turned out to be the refining factor in cycle 2, which further differentiated the typology, improving within-group homogeneity.

### Cycle 2, Step 3: Analysis of Meaningful Relationships and Type Construction Based on the Transitions Between Activities

To evaluate the visual assessment, various quantifiable graph metrics were considered (cf. Andrienko and Andrienko 2018), representing the attributes complexity and homogeneity. State transition graphs with lower complexity show a higher number of communities as activities not conducted by the participant and, therefore, unconnected to the rest of the graph, are counted as one community each (Bastian et al. 2009). To represent complexity, the number of communities was inverted by subtraction with 14 (maximum number of communities—1). To quantify the homogeneity, the reciprocal centrality of the graph was chosen based on the sum of three centrality measures (closeness centrality, degree centrality, and betweenness centrality) to account for structural outliers (Ronqui and Travieso 2015). The mean values of each group in terms of complexity and homogeneity were consistent with the visual assessment (Table 5).

Participants in groups (3a) and (3b) show higher scores for complexity compared with the other groups, as they conduct a broader range of different activities because they use their models to predict the black box behavior. Participants in groups (2c) and (3b) show higher scores for homogeneity compared with groups (2a), (2b), and (3a), as they tend to conduct the different activities more equally. The single participants in groups (1) and (2a) show a complexity score of 0, as they carry out only a few different activities, resulting in the highest number of distinct communities, which were used to offset all calculations. Together with the

**Table 5** Mean values of the calculated complexity and homogeneity metrics for each group

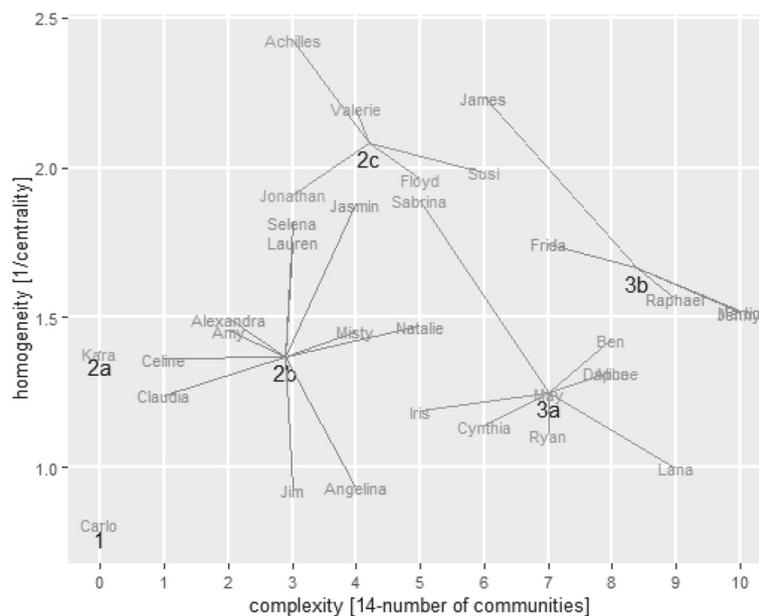
Group	$n$	Mean complexity	Mean homogeneity
(1) Exploration only	1	0.0	0.81
(2a) Exploration and development with a focus on development	1	0.0	1.39
(2b) Exploration and development with a focus on exploration	11	2.9	1.37
(2c) Exploration and development, balanced	5	4.2	2.08
(3a) Exploration, development, and drawing predictions from a model once	9	7.0	1.25
(3b) Exploration, development, and repeatedly drawing predictions from a model	5	8.4	1.67

participants' individual scores, defined areas can be observed that support the visual grouping and that are in line with expectations (Fig. 3).

However, two participants (Sabrina and James) showed a statistically higher homogeneity in comparison to their assigned groups and were therefore investigated further (cf. Fig. 1, step 3). Both participants invested less time in the investigation of the black box compared with their respective groups: while Sabrina and James engaged in the black box activity for about half an hour, the other participants in groups (3a) and (3b) invested an hour more on average. The shorter time of investigation then leads to less repetitive actions and a higher homogeneity score than for the other participants.

## Discussion and Conclusion

Modeling is seen as a central practice of scientific reasoning (e.g., Clement 2000; Gilbert 2004) and is highlighted in curricular documents as a core scientific practice (e.g., NGSS Lead States 2013). However, research on scientific modeling widely lacks detailed investigations of modeling processes (Nicolaou and Constantinou 2014). Therefore, this study aimed to contribute to filling this gap in science education research by developing a typology of preservice teachers' modeling strategies. This typology has the potential to inform and guide further research as well as the development of interventions aiming to foster competencies in scientific modeling. The proposed typology is based on the homogeneity and complexity of modeling processes (Fig. 3) and includes the following strategies: (1) exploration only; (2a) exploration and development with a focus on development; (2b) exploration and development with a focus on exploration; (2c) exploration and development, balanced; (3a) exploration, development,



**Fig. 3** Star plot illustrating the calculated metrics of each participant's state transition graph in conjunction with the mean values of his or her respective group (Table 4)

and drawing predictions from a model once; and (3b) exploration, development, and drawing predictions from a model repeatedly (Table 4).

Comparing the developed types of modeling strategies with strategies identified in previous studies of scientific reasoning processes, many parallels can be drawn, indicating that at least the strategies of modeling and experimental evaluation are quite similar: Participants of type (2b), who conduct activities of exploration and model development with a focus on exploration, show similarities to the participants of Schauble et al. (1991a), who utilize an engineering model of scientific investigation when conducting experiments. The participants carry out an investigation; in this case, the exploration of the black box, until a specific effect, for example, a repeating pattern, is reached. Then, the participants end the modeling process by following the initial task, drawing their model of the black box. In contrast, participants of type (2a) and (2c) perform activities of model development more frequently, expressing behavior similar to that previously described as model fitting (Sins et al. 2005). Moreover, in contrast to previous empirical studies suggesting that understanding the predictive use of models is challenging for most sample groups (Krell and Krüger 2016; Passmore et al. 2014), half of the cases we observed could be assigned to type (3a) or (3b) and included activities of prediction in their modeling processes. However, even participants of type (3b), who drew predictions from their model repeatedly, ended their modeling process when they found a possible solution. Participants did not further evaluate the developed models by trying to falsify the model or by systematically testing alternative hypotheses, although this is suggested in other studies as being important for scientific modeling (Grosslight et al. 1991; Louca and Zacharia 2015).

Methodologically, the identification of single activities of scientific reasoning, in this case modeling, built the foundation for a further analysis of scientific reasoning processes. Based on the occurrence of activities and the transitions between them, a typology of modeling strategies was successfully derived. The occurrence of activities alone was not regarded as a sufficient attribute, since the resulting groups did not show a sufficient degree of within-group homogeneity. By additionally considering the order of the sequence of activities in the form of their transitions, the typology was further refined, improving within-group homogeneity while maintaining between-group heterogeneity (Bailey 1994). Visualizing the modeling processes of each participant in different forms that emphasized the selected attributes, that is, codelines and state transitions graphs, allowed a fast, visual grouping process, helped reduce the complexity of the rich datasets, and improved their comparability. Grouping by multiple raters enabled the evaluation of interrater agreement and a reflection on the grouping process and the decisions involved. This led to the identification of more abstract attributes, such as the complexity and homogeneity of state transition graphs, that could be further quantified and that supported the visual assessment. Multiple cycles of typology development were necessary that were, in the present study, performed iteratively following the model of empirically grounded type construction (Kluge 2000). This model proved helpful in evaluating and refining the typology, as well as in avoiding errors in the assignment of a specific type to an individual case, as the model supported the systematic selection and evaluation of attributes and the resulting groups. We further argue that explicitly describing the different steps of empirically grounded type construction greatly improves transparency, as the selection of attributes and grouping processes have rarely been defined in other studies.

Regarding practical implications, we are aware that the black box setting and the sample size may limit the generalizability of the results. Also, participants' mental models (cf. Johnson-Laird 1983) were not considered in the analysis, as the operationalization of the present study focused on externalized models. Further research is needed that adopts the described methods to other settings and contexts to replicate the proposed typology.

Consequently, at this point, the integration of these findings into teaching environments and the development of interventions that foster competencies in scientific modeling still seem beyond reach. However, with the present findings extending the category system of Krell et al. (2019), preservice science teachers' individual modeling processes can be rapidly assigned one of the proposed types, highlighting the differences between preservice science teachers and reducing the complexity of the analysis. Identifying the type of modeling strategy, a preservice science teacher uses in a specific setting then might yield potential points of leverage on an individual level and could be used to guide this preservice science teacher to understand and use models as epistemic tools for scientific reasoning. Therefore, the proposed typology of modeling strategies might guide further research and theory development (cf. Klahr and Dunbar 1988) and potentially contribute to the development of specific educational interventions (cf. Schauble et al. 1991a, b; Klahr et al. 1993; Klahr and Nigam 2004). One example would be a learning progression similar to Schauble et al. (1991a, b), in which the proposed typology could be used as levels of the learning progression, which are encouraged through modeling tasks with increasing degrees of complexity. Herein, the black box setting might be considered rather complex, since participants are less likely to apply their background knowledge and the black box is nontransparent and dynamic (Betsch et al. 2011).

## Outlook

Our sample will be enlarged, and a possible improvement of within-group homogeneity and between-group heterogeneity will be further investigated. Also, aiming to automate the type assignment process, other statistical measures for calculating differences between sequences will be considered and possibly integrated into the typology as additional dimensions, including  $z$ -scores (cf. Sins et al. 2005) and Yule's  $Q$  (cf. Lämsä et al. 2020). To gain an understanding of how successful a specific type of modeling strategy is, we plan to investigate the relationship between a participant's type of modeling strategy and the quality of the participant's modeling product (i.e., the final model or models throughout the process), as specific strategies are assumed to be used by less successful participants (Schauble et al. 1991a; Sins et al. 2005). Finally, the constructed types will be characterized, including all assessed background variables and the quality of the developed models. We hope that this will help to develop an understanding of how preservice science teachers' model and which factors impact their modeling and, by doing so, provide valuable information for those who are aiming to develop preservice competencies of scientific modeling.

**Funding information** Open Access funding provided by Projekt DEAL. We thank Jonna Kirchhoff and Maike Barnebeck who contributed significantly to the qualitative part of the data analysis, and the German Research Foundation (DFG) for funding the project TypMoL (project No. 327507949). Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the DFG.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Acher, A., Arcá, M., & Sanmartí, N. (2007). Modeling as a teaching learning process for understanding materials: A case study in primary education. *Science Education*, 91(3), 398–418.
- Aizpurua, A., Lizaso, I., & Iturbe, I. (2018). Learning strategies and reasoning skills of university students. *Revista de Psicodidáctica (English ed.)*, 23(2), 110–116. <https://doi.org/10.1016/j.psicoe.2018.02.002>.
- Andrienko, N., & Andrienko, G. (2018). State transition graphs for semantic analysis of movement behaviours. *Information Visualization*, 17(1), 41–65. <https://doi.org/10.1177/1473871617692841>.
- Bailer-Jones, D. (2002). Scientists' thoughts on scientific models. *Perspectives on Science*, 10(3), 275–301. <https://doi.org/10.1162/106361402321899069>.
- Bailey, K. (1994). *Typologies and taxonomies: An introduction to classification techniques* (Vol. 102). Thousand Oaks, CA: Sage.
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge, NY: Cambridge University Press.
- Bastian M, Heymann S, Jacomy M (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 8(2009), 361–362.
- Betsch, T., Funke, J., & Plessner, H. (2011). *Denken–Urteilen, Entscheiden, Problemlösen*. Berlin/Heidelberg: Springer Nature.
- British Columbia Ministry of Education (BCMOE). (2019). Curriculum redesign. Retrieved from <https://curriculum.gov.bc.ca/rethinking-curriculum>
- Brown, D., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional Science*, 18(4), 237–261. <https://doi.org/10.1007/bf00118013>.
- Čančula, M. P., Planinšič, G., & Etkina, E. (2015). Analyzing patterns in experts' approaches to solving experimental problems. *American Journal of Physics*, 83(4), 366–374. <https://doi.org/10.1119/1.4913528>.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. <https://doi.org/10.1002/sce.10001>.
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9), 1041–1053.
- Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Springer Science & Business Media.
- Clement, J., & Rea-Ramirez, M. A. (2008). *Model based learning and instruction in science (model based learning and instruction in science)*. Dordrecht: Springer.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 2(3), 28–45.
- Giere, R. N. (1999). *Science without laws*. Chicago, IL: The University of Chicago Press.
- Giere, R. N., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning* (5th ed.). Belmont: Thomson Wadsworth.
- Gilbert, J. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education*, 2(2), 115–130. <https://doi.org/10.1007/s10763-004-3186-4>.
- Gilbert, J., & Osborne, R. J. (1980). The use of models in science and science teaching. *European Journal of Science Education*, 2(13), 3–13. <https://doi.org/10.1080/0140528800020103>.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Hammann, M., Phan, T. H., & Bayrhuber, H. (2008). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? [experimentation as problem-solving: Can the SDDS model be used to measure different dimensions of experimentation?]. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Eds.), *Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaft* (pp. 33–49). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartmann, S., Upmeier zu Belzen, A., Krüger, D., & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of preservice science teachers' competencies. *Zeitschrift Für Psychologie*, 223(1), 47–53. <https://doi.org/10.1027/2151-2604/a000199>.
- Harwood, W. S. (2004). A new model for inquiry: Is the scientific method dead? *Journal of College Science Teaching*, 33(7), 29–33.
- Hodson, D. (2014). Learning science, learning about science, doing science: Different goals demand different learning methods. *International Journal of Science Education*, 36, 2534–2553. <https://doi.org/10.1080/09500693.2014.899722>.

- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One*, *9*(6), e98679.
- Johnson-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge U.P.
- Justi, R., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education*, *25*, 1369–1386.
- Khan, S., & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, (19), 446–464. Retrieved from <https://link.springer.com/article/10.1007/s42330-019-00063-9>.
- Kind, P., & Osborne, J. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, *101*(1), 8–31. <https://doi.org/10.1002/sce.21251>.
- Klahr, D. (2002). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT press.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667. <https://doi.org/10.1111/j.0956-7976.2004.00737.x>.
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*(1), 111–146.
- Kluge, S. (2000). Empirically grounded construction of types and typologies in qualitative social research. *Forum, Qualitative Sozialforschung / Forum, Qualitative Sozialforschung*, *1*(1). <http://nbn-resolving.de/urn:nbn:de:0114-fqs0001145>.
- KMK. (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss [Biology education standards for the Mittlere Schulabschluss]* (5th ed.). München & Neuwied: Wolters Kluwer.
- KMK. (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung [Common federal requirements regarding scientific disciplines and science education in teacher preparation programs]*. Berlin: Sekretariat der Kultusministerkonferenz
- Koslowski, B. (2013). Scientific reasoning: Explanation, confirmation bias, and scientific practice. In G. J. Feist & M. E. Goman (Eds.), *Handbook of the psychology of science* (pp. 151–192). New York, NY: Springer.
- Krell, M. (2018). Schwierigkeitserzeugende Aufgabenmerkmale bei Multiple-Choice-Aufgaben zur Experimentierkompetenz im Biologieunterricht: Eine Replikationsstudie. *Zeitschrift Für Didaktik Der Naturwissenschaften*, *24*. <https://doi.org/10.1007/s40573-017-0069-0>.
- Krell, M., & Krüger, D. (2016). Testing models: A key aspect to promote teaching activities related to models and modelling in biology lessons? *Journal of Biological Education*, *50*(2), 160–173. <https://doi.org/10.1080/00219266.2015.1028570>.
- Krell, M., Upmeyer zu Belzen, A., & Krüger, D. (2014a). Context-Specificities in Students' Understanding of Models and Modelling: E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning. Nicosia, Cyprus: *European Science Education Research Association*.
- Krell, M., Upmeyer zu Belzen, A., & Krüger, D. (2014b). Students' levels of understanding models and modelling in biology: Global or aspect-dependent? *Research in Science Education*, *44*(1), 109–132. <https://doi.org/10.1007/s11165-013-9365-y>.
- Krell, M., Redman, C., Mathesius, S., Krüger, D., & van Driel, J. (2018). Assessing pre-service science teachers' scientific reasoning competencies. *Research in Science Education*. <https://doi.org/10.1007/s11165-018-9780-1>.
- Krell, M., Walzer, C., Hergert, S., & Krüger, D. (2019). Development and application of a category system to describe pre-service science teachers' activities in the process of scientific modelling. *Research in Science Education* *49*(5), 1319–1345. <https://doi.org/10.1007/s11165-017-9657-8>.
- Krüger, D., Hartmann, S., Nordmeier, V., & Upmeyer zu Belzen, A. (2020). Measuring scientific reasoning competencies. In O. Zlatkin-Troitschanskaia, H. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student learning in German higher education* (pp. 261–280). Wiesbaden: Springer.
- Lämsä, J., Hämmäläinen, R., Koskinen, P., Viiri, J., & Mannonen, J. (2020). The potential of temporal analysis: Combining log data and lag sequential analysis to investigate temporal differences between scaffolded and non-scaffolded group inquiry-based learning processes. *Computers & Education*, *143*, 103674. <https://doi.org/10.1016/j.compedu.2019.103674>.
- Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, *2*(3), 307–338.
- Lederman, N. G., & Abd-El-Khalick, F. (2002). Avoiding de-natured science: Activities that promote understandings of the nature of science. In W. F. M. (Ed.), *The Nature of Science in Science Education* (Vol. 5). Dordrecht: Science & Technology Education Library: Springer.

- Lehrer, R., & Schauble, L. (2015). The development of scientific thinking. *Handbook of child psychology and developmental science*, 2(7), 671–714.
- Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R: IST 2000 R* (2nd ed.). Göttingen: Hogrefe.
- Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471–492. <https://doi.org/10.1080/00131911.2011.628748>.
- Louca, L. T., & Zacharia, Z. C. (2015). Examining learning through modeling in k-6 science education. *Journal of Science Education and Technology*, 24(2), 192–215. <https://doi.org/10.1007/s10956-014-9533-5>.
- Mathesius, S., Hartmann, S., Upmeier zu Belzen, A., & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education: Assessing competencies using a paper-pencil test. In T. Tal & A. Yarden (Eds.), *The future of biology education research* (pp. 93–110). Israel: Haifa.
- Morris, B. J., Croker, S., Masnick, A. M., & Zimmerman, C. (2012). The emergence of scientific reasoning. In H. Kloos, B. J. Morris, & B. J. Amaral (Eds.), *Current topics in children's learning and cognition*. IntechOpen. <https://doi.org/10.5772/53885>.
- Nersessian, N. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere & H. Feigl (Eds.), *Cognitive Models of Science* (Vol. 15, pp. 3–45). Minneapolis, MN: University of Minnesota Press.
- Nersessian, N. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 133–153). New York: Cambridge University Press.
- Neumann, K., Härting, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki, & Hirano T. (Eds.), *Model science teacher preparation programs: An international comparison of what works*. Charlotte, NC: Information Age.
- Neumann, K., Schecker, H., & Theyßen, H. (2019). Assessing complex patterns of student resources and behavior in the large scale. *The Annals of the American Academy of Political and Social Science*, 683(1), 233–249. <https://doi.org/10.1177/0002716219844963>.
- NGSS Lead States. (2013). *Next generation science standards: For States, by States*. Washington, DC: The National Academies Press.
- Nicolaou, C. T., & Constantinou, C. P. (2014). Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13, 52–73. <https://doi.org/10.1016/j.edurev.2014.10.001>.
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265–279.
- Passmore, C. M., & Svoboda, J. (2012). Exploring opportunities for argumentation in modeling classrooms. *International Journal of Science Education*, 34(10), 1535–1554.
- Passmore, C., Stewart, J., & Cartier, J. (2009). Model-based inquiry and school science: Creating connections. *School Science and Mathematics*, 109(7), 394–402.
- Passmore, C., Gouvea, J. S., & Giere, R. (2014). Models in science and in learning science: Focusing scientific practice on sense-making. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 1171–1202). Dordrecht: Springer.
- Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. B. R. (Eds.), *Human Detection and Diagnosis of System Failures, NATO conference series* (Vol. 15, pp. 241–258). Boston: Springer.
- Reiff, R., Harwood, W. S., & Phillipson, T. (2002) A scientific method based upon research scientists' conceptions of scientific inquiry. In *Proceedings of the Annual International Conference of the Association for the Education of Teachers in Science*, Charlotte, NC: Association for the Education of Teachers in Science.
- 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>.
- Ronqui, J. R. F., & Travieso, G. (2015). Analyzing complex networks through correlations in centrality measurements. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(5), P05030.
- Ruppert, J., Duncan, R. G., & Chinn, C. A. (2019). Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education*, 49, 921–948. <https://doi.org/10.1007/s11165-017-9656-9>.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991a). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, 1(2), 201–238. [https://doi.org/10.1207/s15327809jls0102\\_3](https://doi.org/10.1207/s15327809jls0102_3).

- Schauble, L., Klopfer, L. E., & Raghavan, K. (1991b). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28(9), 859–882.
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: Sage Publications.
- Schwarz, C. V., Passmore, C., & Reiser, B. J. (Eds.). (2017). *Helping students make sense of the world using next generation science and engineering practices*. Arlington: NSTA Press.
- Sins, P. H., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The difficult process of scientific modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education*, 27(14), 1695–1721.
- van Driel, J. H., & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21, 1141–1153. <https://doi.org/10.1080/095006999290110> .
- VCAA [Victorian Curriculum and Assessment Authority]. (2016). Victorian Curriculum: F-10. Melbourne, VIC. Retrieved from <http://victoriancurriculum.vcaa.vic.edu.au/science/curriculum/f-10>.
- Windschitl, M., Thompson, J., & Braaten, M. (2008). How novice science teachers appropriate epistemic discourses around model-based inquiry for use in classrooms. *Cognition and Instruction*, 26(3), 310–378. <https://doi.org/10.1080/07370000802177193> .
- Zimmerman, C. (2005). The development of scientific reasoning skills. *Developmental Review*, 20(1), 99–149.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



# Was ist schwierig am Modellieren? Identifikation und Beschreibung von Hindernissen in Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer

Maximilian Göhner<sup>1</sup> · Moritz Krell<sup>1</sup>

Eingegangen: 8. September 2020 / Angenommen: 24. Mai 2021 / Online publiziert: 28. Juni 2021  
© Der/die Autor(en) 2021

## Zusammenfassung

Die Entwicklung von Modellkompetenz im naturwissenschaftlichen Unterricht kann potenziell zur Erreichung vielfältiger Ziele naturwissenschaftlicher Bildung beitragen. Studien deuten allerdings darauf hin, dass Modellierungsprozesse, in denen Modelle kritisch reflektiert oder als Werkzeuge zur Erkenntnisgewinnung eingesetzt werden, im Unterricht eher selten umgesetzt werden und dass Lehrkräfte in Bezug auf Fähigkeiten des naturwissenschaftlichen Modellierens weitere Förderung benötigen. Das Ziel der vorliegenden Studie ist die Identifikation und Beschreibung von Hindernissen, die in Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer auftreten. Hierzu wurden die Modellierungsprozesse von 36 Lehramtsstudierenden naturwissenschaftlicher Fächer bei der Untersuchung einer Black-box qualitativ-inhaltsanalytisch ausgewertet. Es konnten 13 verschiedene Hinderniskategorien identifiziert und beschrieben werden. Die identifizierten Hinderniskategorien weisen teils Parallelen zum Experimentieren und naturwissenschaftlichen Arbeiten allgemein auf: Spezifische Hinderniskategorien für das Modellieren ergeben sich dagegen beim Umgang mit Analogien und Erfahrungen und treten vor allem beim Testen des entwickelten Modells auf. Basierend auf vorherigen Arbeiten wurden zudem die Modellierungsprozesse der Lehramtsstudierenden analysiert und sechs typischen Modellierungsstrategien zugeordnet. Es scheint kein eindeutiger Zusammenhang zwischen den identifizierten Hindernissen und den Modellierungsstrategien vorzuliegen, da ProbandInnen, die ähnlichen Hindernissen begegnen, sich hinsichtlich ihrer Modellierungsstrategien teils deutlich unterscheiden. Es wird diskutiert, inwiefern die identifizierten Hinderniskategorien für die weitere Entwicklung von Diagnoseinstrumenten und zur gezielten Planung von Förderangeboten genutzt werden können.

**Schlüsselwörter** Modellkompetenz · Naturwissenschaftliches Modellieren · Hindernisse · Lehramtsstudierende · Qualitative Inhaltsanalyse

---

✉ Maximilian Göhner  
maximilian.goehner@fu-berlin.de

<sup>1</sup> Didaktik der Biologie, Freie Universität Berlin,  
Schwendenerstraße 1, 14195 Berlin, Deutschland

## What Is Difficult in Modeling? the Identification and Description of Challenges Pre-service Science Teachers Encounter in Modeling Processes

### Abstract

The development of modeling competence can potentially support the achievement of diverse goals in science education. However, studies suggest that modeling processes in which models are critically reflected or used as epistemic tools are rarely implemented in educational settings. Other studies propose a need to promote science teachers' abilities regarding scientific modeling. The present study aims to identify and describe challenges, which pre-service science teachers encounter while modeling. Therefore, the modeling practice of 36 pre-service science teachers engaging in a blackbox activity was analyzed using a qualitative content analytic approach. 13 different challenges were identified and described. Some of the identified challenges show parallels to previously identified challenges related to experimentation and scientific practices in general. In contrast, challenges, which seem to be specific for modeling are related to the use of analogies and occur mostly when the developed models are tested. Based on previous results, the pre-service science teachers' modeling practices were assigned to one of six types of modeling strategies. The identified challenges and the type of modeling strategy used by the pre-service science teachers seem not to be related, as pre-service science teachers encountering similar challenges can utilize very different modeling strategies. It is discussed, how the identified challenges can be used to support the development of diagnostic instruments as well as targeted interventions of modeling competence.

**Keywords** Modeling competence · Scientific modeling · Challenges · Pre-service science teachers · Qualitative content analysis

### Modellieren in den Naturwissenschaften

Die Entwicklung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung ist ein wichtiges Element der naturwissenschaftlichen Grundbildung (Bybee 2002; Osborne 2013), da diese Kompetenzen benötigt werden, um aktiv am naturwissenschafts- und technologieorientierten gesellschaftlichen Leben teilhaben zu können (Driver et al. 1996; Sharon und Baram-Tsabari 2020). Modellieren wird als zentrale Praxis der naturwissenschaftlichen Erkenntnisgewinnung betrachtet (Giere et al. 2006). Dabei wird der Prozess des naturwissenschaftlichen Modellierens zusammenfassend als Entwicklung und Evaluation von Repräsentationen natürlicher Systeme beschrieben (Godfrey-Smith 2006). Hierbei wird die Bezeichnung von Modellen als Repräsentationen allerdings auch kritisch diskutiert und alternativ *epistemic tools* (Knuuttila 2011) vorgeschlagen, um weniger den abbildenden und mehr den forschungsmethodischen Charakter von Modellen als Werkzeuge zur Erkenntnisgewinnung zu betonen (Gilbert und Justi 2016). Dem Verständnis von Modellen als Werkzeuge zur Erkenntnisgewinnung folgend, umfasst der Prozess des naturwissenschaftlichen Modellierens Untersuchungen in der empirischen *Erfahrungswelt*, die Entwicklung von Modellen als Erklärungen und Hypothesen in der theoretischen *Modellwelt*, sowie die Überprüfung dieser Erklärungen und Hypothesen durch weitere empirische Untersuchungen (Dounas-Frazer et al. 2018; Krell et al. 2016). Hierbei findet während des Modellierens ein Vergleich zwischen Modellwelt und Erfahrungswelt statt (Leisner-Bodenthin 2006), der den

Modellierenden im besten Fall bewusst ist und als solcher reflektiert wird (Mikelskis-Seifert und Leisner 2005).

Abb. 1 zeigt die dieser Studie zugrundeliegende Konzeptualisierung des Modellierens (Göhner und Krell 2018), welche auf der Basis naturwissenschaftstheoretischer und fachdidaktischer Arbeiten zum Modellieren entwickelt wurde (z. B. Clement 1989; Giere et al. 2006). Modellierungsprozesse werden darin als Sequenz von beobachtbaren Modellierungstätigkeiten aufgefasst, die wiederum drei Phasen des Modellierens zugeordnet werden können. Zur Phase der Exploration gehört die Beobachtung eines Phänomens in der Erfahrungswelt (Dounas-Frazer et al. 2018; Krell et al. 2016; Leisner-Bodenthin 2006; Mikelskis-Seifert und Leisner 2005). Es findet eine Exploration dieses Phänomens statt, welche sowohl unsystematisch als auch strukturiert und systematisch sein kann. Hierbei hat die Explorationsphase in der Regel das Ziel, Daten zu generieren und Regelmäßigkeiten oder Muster in diesen Daten zu entdecken. Eine Auswahl der in der Explorationsphase gewonnenen Daten und insbesondere deren Regelmäßigkeiten werden in die Modellwelt übertragen (Dounas-Frazer et al. 2018; Krell et al. 2016; Leisner-Bodenthin 2006; Mikelskis-Seifert und Leisner 2005). In der Herstellungsphase wird auf Grundlage von Analogien und Erfahrungen ein initiales Modell entwickelt, in dem die für das Phänomen als relevant erachteten Variablen des betrachteten Systems isoliert sind. Hierbei können zum Beispiel Analogien („[a] somewhat similar and familiar case in memory“; Clement 2008, S. 438) oder Erfahrungen, die eher mechanistischer Natur sind, zur Modellentwicklung herangezogen werden. Modelle können dabei vielfältige Repräsentationsformen aufweisen, zum Beispiel

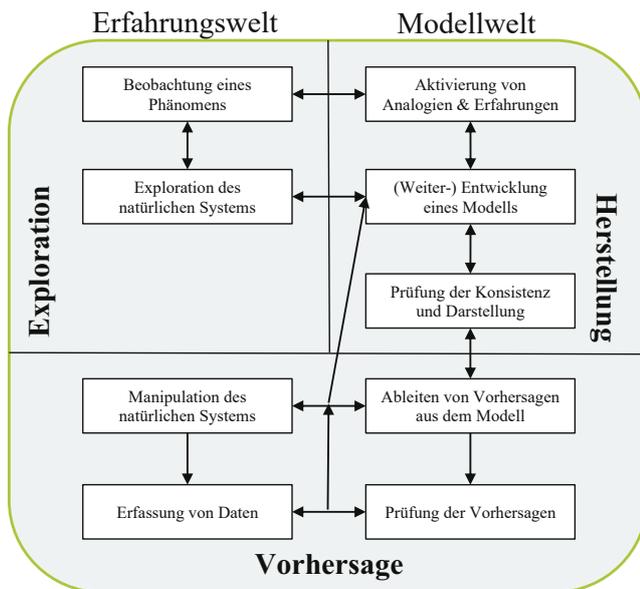


Abb. 1 Prozessdiagramm des Modellierens (Göhner und Krell 2018)

als Zeichnung (Chittleborough und Treagust 2007; Quillin und Thomas 2015) oder digital (Sins et al. 2005; Bo et al. 2018). Das entwickelte Modell kann dann auf interne Konsistenz sowie eine angemessene Darstellung des beobachteten Phänomens geprüft werden. Diese erkenntniserweiternde Entwicklung erklärender Modelle beruht auf abduktivem Schließen (Magnani 1999; Krüger und Upmeyer zu Belzen 2021). Darüber hinaus sind auch deduktive Schlüsse möglich, indem in der Phase der Vorhersage eine gedankliche oder materielle Manipulation des Modells stattfindet, die es erlaubt Hypothesen darüber abzuleiten, wie sich das System unter bestimmten Bedingungen verhalten wird (Giere et al. 2006). Können die abgeleiteten Hypothesen falsifiziert werden, kann indirekt auf eine mangelnde Passung zwischen Modell und System geschlossen werden. Das Modell sollte dann überarbeitet oder gänzlich verworfen werden. In beiden Fällen schließt sich eine erneute Testung im beschriebenen Sinne an (Krell et al. 2016). Das Modellieren wird somit als zyklischer, nicht-linearer Prozess verstanden. Konkrete Modellierungsprozesse können dabei eine vielfältige, nicht notwendigerweise vollständige Abfolge beobachtbarer Tätigkeiten des Modellierens umfassen, die wie beschrieben den drei Phasen des Modellierens zugeordnet werden können.

## Modellkompetenz als Ziel naturwissenschaftlichen Unterrichts

Modellkompetenz wird definiert als die Fähigkeiten, „mit Modellen zweckbezogen Erkenntnisse gewinnen zu können und über Modelle mit Bezug auf ihren Zweck urteilen zu

können, die Fähigkeiten, über den Prozess der Erkenntnisgewinnung durch Modelle und Modellierungen [...] zu reflektieren sowie die Bereitschaft, diese Fähigkeiten in problemhaltigen Situationen anzuwenden“ (Upmeyer zu Belzen und Krüger 2010). Insbesondere die Fähigkeiten, Modelle zur Vorhersage von Daten einzusetzen und sie als hypothetische Entitäten zu verstehen, werden als bedeutsame Elemente ausgeprägter Modellkompetenz betrachtet (Gouvea und Passmore 2017; Upmeyer zu Belzen und Krüger 2010). In der Literatur werden darüber hinaus verschiedene Dimensionen von Modellkompetenz unterschieden; hierzu gehören 1) das Verständnis über Modelle und den Prozess des Modellierens (Modellverstehen bzw. *metamodeling knowledge*), 2) Fähigkeiten des praktischen Modellierens sowie, in einzelnen Quellen, ebenfalls 3) die Qualität des entwickelten Modells (Chiu und Lin 2019; Nicolaou und Constantinou 2014; Nielsen und Nielsen 2019).

Die Entwicklung von Modellkompetenz im naturwissenschaftlichen Unterricht kann potenziell zum Erreichen vielfältiger Ziele naturwissenschaftlicher Bildung beitragen (Hodson 2014; Justi und Gilbert 2002). So können naturwissenschaftliche Inhalte mit Hilfe von Modellen erschlossen werden, wenn Modelle als Produkte der Wissenschaft verstanden werden (Acher et al. 2007; Vo et al. 2019). Darüber hinaus kann naturwissenschaftliche Praxis reflektiert werden, wenn Modellierungsprozesse umgesetzt und nachvollzogen werden (Fleige et al. 2012). Aufgrund der großen Bedeutung des Modellierens für die naturwissenschaftliche Erkenntnisgewinnung wird die Arbeit mit Modellen in Steuerungsdokumenten für den naturwissenschaftlichen (Schul-)Unterricht unterschiedlicher Länder explizit gefordert (ACARA 2015; BCMOE 2019; KMK 2005, 2020; NGSS Lead States 2013). In den Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss in Deutschland wird hervorgehoben, dass „gerade das Modellieren bzw. kritische Reflektieren des Modells bedeutsamer Teil der naturwissenschaftlichen Erkenntnisgewinnung“ (KMK 2005, S. 11) ist. So sollen SchülerInnen am Ende der Jahrgangsstufe 10 zum Beispiel in der Lage sein, „Modelle praktisch [zu] erstellen“ (Anforderungsbereich (AFB) I), „Sachverhalte mit Modellen [zu] erklären“ (AFB II) und „Hypothesen [zu] erstellen mit einem Modell“ (AFB III) (KMK 2005, S. 17).

Studien zum Einsatz von Modellen und dem Modellieren im Unterricht weisen allerdings darauf hin, dass diese curricularen Vorgaben oftmals nicht erreicht werden. So werden Modellierungsprozesse, in denen Modelle kritisch reflektiert oder als Werkzeuge zur Erkenntnisgewinnung eingesetzt werden, im Unterricht eher selten umgesetzt (Campbell et al. 2015). Im naturwissenschaftlichen Unterricht scheint das Entwickeln und Evaluieren von Modellen häufiger umgesetzt zu werden als die sich daran anschließende Weiterentwicklung von Modellen (Khan 2011).

Insbesondere das Testen von Modellen durch die Ableitung und Überprüfung von Hypothesen scheint im Unterricht eher selten umgesetzt zu werden (Krell und Krüger 2013); obwohl gerade ein Verständnis dieses Testens von Modellen als ein Schlüssel zur Förderung von Modellverstehen betrachtet wird (Krell und Krüger 2016; Lee und Kim 2014). In der Videostudie von Werner et al. (2019) konnten Aufgabenstellungen, die Modelle als Werkzeuge zur Erkenntnisgewinnung nutzen, sowie die kritische Reflexion von Modellen nur selten beobachtet werden. Insgesamt kann somit davon ausgegangen werden, dass Modelle im Unterricht entgegen der curricularen Vorgaben (ACARA 2015; BCMOE 2019; KMK 2005, 2020; NGSS Lead States 2013) und fachdidaktischen Zielsetzungen (Fleige et al. 2012) primär als Anschauungsmedien zur Erschließung naturwissenschaftlicher Inhalte eingesetzt werden. Das Potenzial, das Modelle und das Modellieren dabei für die Unterrichtsgestaltung zur Förderung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung mitbringen, wird somit offenbar eher selten genutzt.

Um den curricularen Vorgaben gerecht zu werden, sollen Lehrkräfte selbst über ausreichende Modellkompetenz verfügen (Günther et al. 2017, 2019), die sie befähigt im Unterricht eine Diagnose der Modellkompetenz von SchülerInnen, sowie die Planung und Umsetzung von entsprechenden Lerngelegenheiten vornehmen zu können (Elby und Hammer 2010; Sung und Oh 2018). Entsprechend wird auch im fachspezifischen Kompetenzprofil Biologie der ländergemeinsamen inhaltlichen Anforderungen in der Lehrerbildung gefordert, dass StudienabsolventInnen „vertraut [sind] mit basalen Arbeits- und Erkenntnismethoden der Biologie und über Kenntnisse und Fertigkeiten sowohl im hypothesengeleiteten Experimentieren und Modellieren [...] [verfügen]“ (KMK 2019). Diverse Studien weisen allerdings darauf hin, dass bei Lehrkräften in den verschiedenen Dimensionen der Modellkompetenz Defizite vorliegen (z.B. Krell und Krüger 2016; Nielsen und Nielsen 2019).

### **Bisherige Forschungsergebnisse zur Modellkompetenz von Lehrkräften naturwissenschaftlicher Fächer**

Mehrere Studien deuten darauf hin, dass angehende (z.B. Torres und Vasconcelos 2015; Yenilmez Turkoglu und Oztekin 2016) sowie praktizierende (z.B. Justi und Gilbert 2003; Krell und Krüger 2016) Lehrkräfte über ein eher eingeschränktes Modellverstehen verfügen und Modelle vor allem als Repräsentationen oder direkte Abbilder der Realität betrachten. Ein Verständnis über Modelle als Werkzeuge zur naturwissenschaftlichen Erkenntnisgewinnung ist bei

Lehrkräften in der Regel nur gering ausgeprägt (Gilbert und Justi 2016).

Es liegen, anders als in anderen Bereichen der naturwissenschaftlichen Erkenntnisgewinnung (z.B. Experimentieren: Baur 2018; Hammann 2004), nur eingeschränkt qualitative und prozessorientierte Studien zu Fähigkeiten und Vorgehensweisen von Lehrkräften beim Modellieren vor (Louca und Zacharia 2012; Nicolaou und Constantinou 2014). Insbesondere liegen kaum Studien vor, die gezielt Interventionen zur Förderung des Verstehens von Modellierungsprozessen untersuchen (Campbell et al. 2015).

Die vorliegenden Studien zu Modellierungsprozessen von Lehrkräften beobachten, dass Lehrkräfte eher zu einer produkt-orientierten Perspektive auf Modelle und das Modellieren tendieren (Nielsen und Nielsen 2019) und dass Lehrkräfte selten Modellierungsprozesse umsetzen, in denen Modelle zur Vorhersage von Daten eingesetzt werden (Göhner und Krell 2020a). In der Studie von Göhner und Krell (2020a) zum Modellieren an einer Blackbox setzten lediglich fünf von 32 angehenden Lehrkräften Modellierungsprozesse um, in denen wiederholt Vorhersagen aus entwickelten Modellen abgeleitet und überprüft wurden, während die Mehrzahl der ProbandInnen ausschließlich Tätigkeiten der Exploration und Herstellung zeigten (Abb. 1). Cheng et al. (2019) konnten zwar im Hinblick auf das Modellieren von Magnetismus beobachten, dass Lehrkräfte im Vergleich zu SchülerInnen elaboriertere Modelle entwickeln und eher wissenschaftliche Kriterien zur Evaluation der Modelle nutzen. Allerdings gelang dies auch nur etwa einem Drittel der dort teilnehmenden Lehrkräfte. Auch Windschitl und Thompson (2006) konnten in einer Interventionsstudie zeigen, dass es Lehrkräften schwerfällt, selbst Modelle zu entwickeln und als Werkzeuge zur Erkenntnisgewinnung zu nutzen. Zwar gelang es den Lehrkräften oftmals, über vorhandene wissenschaftliche Modelle elaboriert zu sprechen, beim selbstständigen Modellieren fiel es den teilnehmenden Lehrkräften allerdings schwer, theoretische Beziehungen in ihre Modelle zu integrieren, die ihnen nicht vollständig bekannt waren. Mehrere Studien beobachten zudem sogenanntes *model-fitting* Verhalten, bei dem der Fokus auf dem Output eines Modells liegt, anstatt auf den modellierten theoretischen Beziehungen; Modellparameter werden hierbei schrittweise angepasst, sodass der Output des Modells den tatsächlichen Beobachtungen entspricht, ohne tiefgehend über das Modell, die darin enthaltenen Elemente oder ihre Beziehungen zueinander nachzudenken (Hogan und Thomas 2001; Sins et al. 2005).

Neben einem mangelnden Modellverstehen wird auch mangelndes Fachwissen als ein Faktor vermutet, der Modellierungsprozesse erschwert (Ruppert et al. 2019). Demgegenüber beobachten Windschitl und Thompson (2006), dass es Lehrkräften oft nicht gelingt, richtig erkannte Zu-

sammenhänge und Konzepte über ein beobachtetes Phänomen in Modellierungsprozessen zu berücksichtigen und in die entwickelten Modelle zu integrieren. Selbst Lehrkräfte mit ausreichendem Fachwissen identifizieren zwar relevante Konzepte, aber entwickeln oft kein Modell, welches das vorhandene Wissen über das gegebene Phänomen akkurat repräsentiert. Diese unvollständigen Modelle wiederum behindern den weiteren Prozess des Modellierens, da sie nicht sinnvoll im Zusammenhang mit dem gegebenen Phänomen getestet werden können (Windschitl und Thompson 2006).

Es liegen keine Studien zur Entwicklung von Modellkompetenz bei angehenden Lehrkräften über den Studienverlauf vor, die systematisch untersuchen, inwieweit die curricularen Forderungen in Bezug auf das Modellieren (KMK 2019) in der universitären Lehrkräftebildung erreicht werden. Es ist demnach auch möglich, dass Lehrkräfte oftmals wenige Gelegenheiten haben, an Prozessen der naturwissenschaftlichen Erkenntnisgewinnung durch Modellieren teilzuhaben und diese explizit zu reflektieren. Studien zur Entwicklung von Kompetenzen wissenschaftlichen Denkens (inkl. der Dimension Modellieren) zeigen eine positive aber lediglich moderate Entwicklung dieser Kompetenzen bei Lehramtsstudierenden naturwissenschaftlicher Fächer im Verlauf des Lehramtsstudiums (Krüger et al. 2020). Hartmann et al. (2015) schlussfolgern, dass für die Entwicklung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung eine gezielte und explizite Förderung und Reflexion über die Prozesse naturwissenschaftlicher Erkenntnisgewinnung notwendig ist.

Zusammenfassend kann davon ausgegangen werden, dass die Förderung von Modellverstehen und Fachwissen nicht ausreicht, um bei Lehrkräften Fähigkeiten des naturwissenschaftlichen Modellierens zu entwickeln, sodass diese in der Lage sind, Modellierungsprozesse erfolgreich in die Unterrichtspraxis zu übertragen (Bo et al. 2018). Das bloße Nachahmen und Teilhaben an wissenschaftlichen Erkenntnisprozessen, beispielsweise während eines Lehramtsstudiums, reichen nicht zum Beherrschen ebenjener Prozesse aus (Bell et al. 2003). Eine gezielte und explizite Förderung ist nötig, um Modelle als Werkzeuge zur Erkenntnisgewinnung entwickeln und die Praxis des wissenschaftlichen Modellierens auch im schulischen Kontext umsetzen zu können (Clement 2000; Hartmann et al. 2015; Lehrer und Schauble 2006). Sung und Oh (2018) schlagen die Identifikation und Beschreibung von spezifischen Hindernissen vor, die in Modellierungsprozessen auftreten können. Naturwissenschaftliches Modellieren wird hierbei als Problemlösen verstanden (Mayer 2007). Ein Problem ist definiert durch das Vorhandensein eines Ziels, dessen Erreichen fraglich ist, und das Fehlen einer Routine, um Hindernisse auf dem Weg zu diesem Ziel zu beseitigen (Funke 2011; Mayer 2007). Ein Hindernis kann dabei verstanden werden als eine „Stelle im Bearbeitungsprozess,

in der rekonstruierbar ist, dass eine Person nichts oder etwas nicht selbstverständlich (im Sinne von nicht sicher, zweifelnd) ausführt und dabei auf nichts in der Aufgabensituation Anwendbares zugreifen möchte bzw. zugreifen kann“ (Lange, 2013, S. 32). Problemlösen bedeutet demnach „das Beseitigen eines Hindernisses oder das Schließen einer Lücke in einem Handlungsplan durch bewusste kognitive Aktivitäten, die das Erreichen eines beabsichtigten Ziels möglich machen sollen“ (Funke 2011, S. 138). Die Identifikation spezifischer Hindernisse in Modellierungsprozessen könnte daher sowohl zur problemtypenbasierten Modellierung und Diagnose von Modellkompetenz (Gut et al. 2014) als auch in Vermittlungssituationen genutzt werden, in denen hindernisspezifisch Hilfestellungen gegeben und die Entwicklung von Modellkompetenz gezielt unterstützt werden könnte (Baur 2018; Sins et al. 2005; Sung und Oh 2018).

## Ziel und Fragestellungen

Die vorliegende Studie hat das Ziel Hindernisse zu identifizieren und zu beschreiben, die in Modellierungsprozessen von Lehramtsstudierenden auftreten, sowie die identifizierten Hindernisse in den Modellierungsprozessen zu verorten. Hierbei werden drei Forschungsfragen verfolgt:

1. Welche Hindernisse lassen sich in den Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer identifizieren und beschreiben?
2. Inwiefern treten die identifizierten Hindernisse in einzelnen Phasen der Modellierungsprozesse von Lehramtsstudierenden naturwissenschaftlicher Fächer besonders häufig auf?
3. Inwiefern besteht ein Zusammenhang zwischen den identifizierten Hindernissen und den Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer?

## Methoden

### Studienkontext

Die vorliegende Studie ist im universitären Lehramtsstudium und damit in der ersten Phase der Lehrkräftebildung verortet. Lehrkräfte sollen bereits im Studium Teile des professionellen Wissens und der Kompetenzen entwickeln, die später im schulischen Alltag benötigt werden (Neumann et al. 2017); hierzu gehören auch Wissen und Kompetenzen im Bereich der Erkenntnisgewinnung (KMK 2019).

## Setting der Studie

Um Modellierungsprozesse anzuregen, wurde in der vorliegenden Studie ein Blackbox-Setting eingesetzt. Der Einsatz einer Blackbox stellt in der naturwissenschaftsdidaktischen Forschung einen etablierten Ansatz zur Untersuchung von Prozessen der Erkenntnisgewinnung und des wissenschaftlichen Denkens dar (Krell und Hergert 2019; Lederman und Abd-El-Khalick 2002; Passmore und Svoboda 2012). Die Blackbox repräsentiert hierbei ein natürliches Phänomen und die Untersuchung der Blackbox den Prozess der Erkenntnisgewinnung (Upmeier zu Belzen 2014). In der vorliegenden Studie wurde eine Wasser-Blackbox eingesetzt. Diese kann durch das Einfüllen von Wasser und die anschließende Beobachtung des resultierenden Outputs untersucht werden (Krell und Hergert 2019). Eine Darstellung des inneren Mechanismus der verwendeten Blackbox ist im Anhang enthalten (Abb. 6).

Die ProbandInnen wurden in einem Einführungsgespräch über das Ziel der Studie informiert, um ihr schriftliches Einverständnis zur Studienteilnahme gebeten und aufgefordert, einen kurzen Fragebogen zum Modellverstehen zu beantworten (Krell und Krüger 2016). Die nachfolgende Aufgabe der Untersuchung der Blackbox wurde dabei explizit als Modellierungsaufgabe angekündigt. Die ProbandInnen wurden anschließend aufgefordert, während der Untersuchung der Blackbox laut zu denken (Ericsson und Simon 1998), was ebenfalls im Rahmen des Einführungsgesprächs mit drei kurzen Aufgaben geübt wurde (Sandmann 2014). Im Anschluss an diese Vorbereitung wurden die ProbandInnen in einen Raum geführt, in welchem die Blackbox, Messbecher unterschiedlichen Fassungsvermögens, ein Wasserreservoir, eine Tafel mit Kreide und drei Kameras zur Dokumentation aufgestellt waren. Nach einer kurzen, standardisierten Erklärung der grundlegenden Funktionsweise der Blackbox wurde den ProbandInnen die folgende Aufgabenstellung verbal mitgeteilt: „Deine Aufgabe ist es, mit Hilfe der bereitgestellten Materialien das Verhalten der Blackbox zu untersuchen und ein Modell des Inneren der Blackbox zu entwickeln und fortlaufend auf einer Tafel zu dokumentieren. Die einzige Einschränkung ist: Du darfst die Blackbox nicht öffnen.“ Zudem war die Aufgabenstellung in Kurzform („Zeichnen Sie ein Modell des Inneren der Blackbox.“) zu jeder Zeit als sichtbarer Hinweis in Form eines laminierten Schildes an der Tafel angebracht. Alle ProbandInnen wurden zudem darüber informiert, dass keine zeitliche Beschränkung vorliegt (durchschnittliche Untersuchungszeit: 1 h 08 min; Anhang, Tab. 5). Während dieser Zeit verblieb der Erstautor oder eine geschulte studentische Hilfskraft im Raum, um die Technik zu überwachen und, falls nötig, die ProbandInnen mit standardisierten Impulsen dazu aufzufordern, weiterhin laut zu denken.

## Sampling

In der vorliegenden Studie wurde ein Sampling nach einem qualitativen Stichprobenplan mit zwei Samplingkriterien umgesetzt (Schreier 2010). Das erste Kriterium sind Kompetenzen des wissenschaftlichen Denkens (*scientific reasoning competencies*; Hartmann et al. 2015), da wissenschaftliches Denken und wissenschaftliches Modellieren nach theoretischen Überlegungen positiv zusammenhängen (Giere et al. 2006; Passmore et al. 2014). Als zweites Kriterium wurden allgemeine kognitive Fähigkeiten ausgewählt, welche neben verbalen und numerischen Fähigkeiten auch Fähigkeiten des räumlichen Denkens umfassen, welche elementare Voraussetzung für Modellierungsprozesse sind (Nersessian 2002). Zwischen allgemeinen kognitiven Fähigkeiten und Modellverstehen (Krell et al. 2014) sowie wissenschaftlichem Problemlösen (Nehring et al. 2015) konnte auch in empirischen Studien ein positiver Zusammenhang gezeigt werden. Die beiden Kriterien wurden mit zwei etablierten paper-pencil Instrumenten (Hartmann et al. 2015; Liepmann et al. 2007) bei  $N=666$  Lehramtsstudierenden naturwissenschaftlicher Fächer erhoben. Um die Wahrscheinlichkeit zu erhöhen, möglichst vielfältige Modellierungsprozesse beobachten zu können, wurden ausgehend von der jeweiligen Referenzstichprobe diejenigen Studierenden zum Hauptteil der vorliegenden Studie eingeladen, die für beide Kriterien eine (wissenschaftliches Denken) beziehungsweise eine halbe (kognitive Fähigkeiten) Standardabweichung besser oder schlechter abgeschnitten haben (heterogene Stichprobe; Schreier 2010). Dies traf auf 57 Studierende zu, von denen 36 letztlich teilnahmen (Anhang, Tab. 5). Eine Analyse der Modellierungsprozesse der ProbandInnen in Abhängigkeit der beiden Samplingkriterien ist andernorts publiziert (Göhner und Krell in Druck) und daher nicht Gegenstand der vorliegenden Arbeit.

## Analyse der Modellierungsprozesse

Die Tonspuren der Videos wurden vollständig transkribiert, dabei wurden neben den Verbalisierungen der ProbandInnen auch ausgewählte Informationen zu ihrem Verhalten im Transkript ergänzt (z. B. Einfüllen von Wasser, Zeichnen eines Modells an der Tafel). Die daraus resultierenden Transkripte wurden nach einer strukturierenden qualitativen Inhaltsanalyse und unter Zuhilfenahme der Videos untersucht (Schreier 2012). Innerhalb der Modellierungsprozesse der ProbandInnen wurden hierbei einzelne Tätigkeiten des Modellierens identifiziert, wofür auf ein Kategoriensystem mit 19 Kategorien (=Tätigkeiten) zurückgegriffen wurde (Tab. 1). Die Identifikation der einzelnen Tätigkeiten in den Modellierungsprozessen der ProbandInnen wurde von zwei Personen unabhängig voneinander vorgenommen (Göhner und Krell 2020b). Dabei wurde im Mittel eine

**Tab. 1** Kategoriensystem mit 19 Tätigkeiten des Modellierens (Krell et al. 2019) sowie deren Verortung in Phasen des Modellierungsprozesses

Phase	Oberkategorie	Kategorie (=Tätigkeit)	Beschreibung
E	Beobachtung eines Phänomens	1. Beobachtung eines Phänomens	ProbandInnen nehmen Verhalten von Blackbox als spontan nicht erklärbar wahr oder erkennen fehlende Passung zwischen angenommenem Muster der Blackbox und Verhalten von Blackbox
E	Exploration des Systems	2. Input/Output nicht hypothesengeleitet	ProbandInnen nehmen Eingabe von Input vor und/oder beobachten den Output (explorativ; nicht hypothesengeleitet)
		3. Zusammenfassung	ProbandInnen fassen das beobachtete Verhalten der Blackbox zusammen oder beschreiben es
		4. Input/Output hypothesengeleitet Daten	ProbandInnen nehmen Eingabe von Input vor und/oder beobachten den Output (Mustererkennung)
		5. Mustererkennung	ProbandInnen bestätigen/erkennen Muster
E,H,V <sup>a</sup>	Aktivierung von Analogien und Erfahrungen	6. Analogien und Erfahrungen	ProbandInnen formulieren Ideen über Mechanismus der Blackbox und/oder reflektieren Vor- und Nachteile ihrer Ideen und verwerfen ggf. Ideen (ohne zeichnerische Umsetzung)
H	(Weiter-) Entwicklung eines Modells	7. Modellentwicklung	ProbandInnen entwickeln auf Grundlage von Beobachtung, Analogien und/oder Erfahrungen zeichnerisch ein Modell der Blackbox
		8. Optimierung Modellobjekt	ProbandInnen entwickeln Modell weiter zur Optimierung von Funktionsfähigkeit, Ästhetik o. ä. (Modellobjekt)
		9. Optimierung Modell retrospektiv	ProbandInnen entwickeln Modell weiter aufgrund fehlerhafter Übereinstimmungen mit Beobachtungen an Blackbox (retrospektiv)
		10. Verwerfen Modell	ProbandInnen verwerfen Modell aufgrund fehlerhafter Übereinstimmungen mit Beobachtungen an Blackbox (retrospektiv; vgl. Kategorie 12) oder mangelnder Konsistenz des Modellobjekts (vgl. Kategorie 11)
H	Prüfung von Konsistenz und Darstellung	11. Prüfung Konsistenz Modellobjekt	ProbandInnen reflektieren/überprüfen/bewerten Konsistenz des Modells (Modellobjekt)
		12. Prüfung Konsistenz Modell/Daten	ProbandInnen vergleichen Eigenschaften des Modells mit Beobachtungen an Blackbox (retrospektiv)
H	Feststellung Konsistenz & Darstellung des Modells	13. Feststellen Konsistenz	ProbandInnen stellen Konsistenz und Angemessenheit des Modells fest
V	Ableiten von Vorhersage (Hypothese) aus Modell	14. Hypothese aus Modell	ProbandInnen nutzen Modell, um Vorhersage über Output bei bestimmtem Input zu treffen
V	Prüfung der Vorhersage (datenbasiert)	15. Input/Output hypothesengeleitet Modell	ProbandInnen nehmen Eingabe von Input vor und beobachten den Output (hypothesengeleitet aus Modell)
		16. Hypothese bestätigt	ProbandInnen bestätigen Hypothese aus Modell durch Output von Blackbox
		17. Hypothese falsifiziert	ProbandInnen falsifizieren Hypothese aus Modell durch Output von Blackbox
V	Ändern/Verwerfen des Modells (datenbasiert)	18. Optimierung Modell nach falsifizierter Hypothese	ProbandInnen weiterentwickeln Modell aufgrund von falsifizierter Hypothese (datenbasiert)
		19. Verwerfen Modell nach falsifizierter Hypothese	ProbandInnen verwerfen Modell aufgrund von falsifizierter Hypothese (datenbasiert)

E Exploration; H Herstellung; V Vorhersage (vgl. Abb. 1)

<sup>a</sup> Analogien und Erfahrungen können während des gesamten Modellierungsprozesses aktiviert werden. Diese Kategorie wird daher keiner spezifischen Phase zugeordnet

hohe Interrater-Übereinstimmung erreicht ( $0,45 \leq \kappa \leq 0,93$ ; Mittelwert $_{\kappa} = 0,77$ ; Landis und Koch 1977).

Auf der Grundlage der Komplexität und Homogenität der Tätigkeitsverläufe, das heißt dem Auftreten und der Abfolge der verschiedenen Tätigkeiten in den Modellierungsprozessen, wurden die ProbandInnen einer von sechs typischen Modellierungsstrategien zugeordnet (Tab. 2).

Diese Zuordnung der ProbandInnen basierend auf den konkret beobachteten Modellierungsprozessen zu typischen, idealisierten Modellierungsstrategien setzt keine aktiv von den ProbandInnen umgesetzte metakognitive Handlung voraus. Modellierungsprozesse werden als komplexer verstanden, umso mehr verschiedene Tätigkeiten des Modellierens sie umfassen (Göhner und Krell 2020a). Modellierungspro-

**Tab. 2** Typen von Modellierungsstrategien, deren Komplexität und Homogenität sowie die Häufigkeit des Auftretens in dieser Studie (Göhner und Krell 2020a)

Typ von Modellierungsstrategie	Kurzbeschreibung	Häufigkeit	Komplexität	Homogenität
1 Nur Exploration	Es werden ausschließlich Tätigkeiten der Exploration (1–5) durchgeführt. Eine Vermutung über den inneren Mechanismus der Blackbox wird lediglich verbal kommuniziert, es entsteht kein gezeichnetes Modell	1	Niedrig	Niedrig
2a Exploration & Herstellung; Fokus Exploration	Es werden Tätigkeiten der Exploration (1–5) und der Herstellung (7–13) durchgeführt. Dabei überwiegen <sup>a</sup> Tätigkeiten der Exploration. Das gezeichnete Modell entsteht oft erst am Ende der Modellierungsprozesse und dient primär der abschließenden Repräsentation der Vermutungen	12	Niedrig	Mittel
2b Exploration & Herstellung; Fokus Herstellung	Es werden Tätigkeiten der Exploration (1–5) und der Herstellung (7–13) durchgeführt. Dabei überwiegen <sup>a</sup> Tätigkeiten der Herstellung. Ein gezeichnetes Modell entsteht oft früh in Modellierungsprozessen und dient primär der fortlaufenden Entwicklung der Vermutungen	2	Mittel	Mittel
2c Exploration & Herstellung	Es werden Tätigkeiten der Exploration (1–5) und der Herstellung (7–13) durchgeführt, ohne dass ein Schwerpunkt auf einer der beiden Phasen erkennbar ist. Ein gezeichnetes Modell entsteht oft früh in Modellierungsprozessen und dient primär der retrospektiven Kontrastierung der entwickelten Vermutung und der beobachteten Daten	6	Mittel	Hoch
3a Exploration, Herstellung & Vorhersage	Es werden Tätigkeiten der Exploration (1–5), der Herstellung (7–13) und der Vorhersage (14–19) durchgeführt. Das gezeichnete Modell wird dabei einmalig zur Vorhersage von nachfolgend generierten Daten eingesetzt	9	Hoch	Mittel
3b Exploration, Herstellung & mehrmalige Vorhersage	Es werden Tätigkeiten der Exploration (1–5), der Herstellung (7–13) und der Vorhersage (14–19) durchgeführt. Das gezeichnete Modell wird dabei mehrmals zur Vorhersage von nachfolgend generierten Daten eingesetzt	6	Hoch	Hoch

<sup>a</sup> Das überwiegende Vorkommen von Tätigkeiten setzt voraus, dass mindestens 50 % der Tätigkeiten in einem Modellierungsprozess der entsprechenden Phasen zugeordnet werden können

zesse werden als homogener verstanden, umso gleichmäßiger die Tätigkeiten des Modellierens vorkommen (d. h. alle Tätigkeiten werden ähnlich häufig ausgeführt und auch die Wechsel zwischen Tätigkeiten sind gleichmäßig verteilt). Die Komplexität und die Homogenität der Modellierungsprozesse wurden mit Hilfe graphentheoretischer Maße quantifiziert (Göhner und Krell 2020a). In Bezug auf die Qualität der Modellierungsprozesse, erfüllen nur Modellierungsprozesse, die Modellierungsstrategie Typ 3a oder 3b zugeordnet werden, das curricular beschriebene Ideal der Modellierung zur naturwissenschaftlichen Erkenntnisgewinnung (ACARA 2015; BCMOE 2019; KMK 2005, 2020; NGSS Lead States 2013) und die Nutzung von Modellen zur Vorhersage von Daten (Gouvea und Passmore 2017; Upmeier zu Belzen und Krüger 2010). Diese beiden Modellierungsstrategien zeichnen sich durch das zyklische Durchlaufen vielfältiger Tätigkeiten des Modellierens aus, wobei wiederholt Daten gesammelt, Modelle entwickelt und anhand von Vorhersagen getestet werden.

### Identifikation der Hindernisse

Um Hindernisse in den Modellierungsprozessen der ProbandInnen zu identifizieren wurden alle 36 Transkripte im Rahmen einer weiteren strukturierenden qualitativen Inhaltsanalyse analysiert. Hierbei wurde initial eine induktive

Strategie der Kategorienbildung verfolgt, da bisher wenige Studien zu Modellierungsprozessen von Lehrkräften vorliegen (Louca und Zacharia 2012; Nicolaou und Constantinou 2014). Zur Pilotierung des Kategoriensystems wurden hierbei insgesamt neun der 36 Fälle ausgewählt, die sich hinsichtlich ihres Vorgehens und der verwendeten Modellierungsstrategie unterscheiden. Die Transkripte wurden event-basiert kodiert. Hindernisse sind hierbei nach Dörner (1976) operationalisiert als Stellen im Modellierungsprozess an denen die ProbandInnen innehalten und äußern, dass sie die Aufgabe als schwierig empfinden oder ratlos sind, sowie Stellen, an denen die ProbandInnen längere Pausen beim lauten Denken machen und dann mit einer anderen Aktivität fortfahren (vgl. „subjektive Barriere“; Dörner 1976). Zusätzlich wurden neben diesen Verbalisierungen, in denen die ProbandInnen von sich aus auf Hindernisse hinweisen oder längere Pausen machen, auch solche Textstellen markiert, in denen Verbalisierungen nicht der aktuell oder einer später ausgeführten Handlung entsprechen (vgl. „intersubjektive Barriere“; Dörner 1976). Basierend auf den markierten Textstellen von sechs der neun Transkripte wurden dann induktiv Kategorien von Hindernissen gebildet. Inhaltlich bestehen die Kodiereinheiten aus mindestens ganzen Sätzen, in denen entweder das Wahrnehmen eines Hindernisses verbalisiert wird, oder in denen Verbalisierungen enthalten sind, die sich später im Abgleich mit

den Handlungen der ProbandInnen als unstimmtig erkennen lassen. Zeitlich sind die Kodiereinheiten damit auf wenige Sekunden beschränkt, die zusätzlich in die Tätigkeiten des Modellierens als Kontexteinheit (wenige Sekunden bis mehrere Minuten) eingebettet sind. Der Erstautor und eine geschulte Mitarbeiterin haben in fünf Entwicklungszyklen abwechselnd Textstellen kategorisiert oder gegebenenfalls einzelne Kategorien weiter ausdifferenziert. Um das Kategoriensystem auch deduktiv zu verorten und Vergleichbarkeit sicherzustellen, wurden die zunächst induktiv gebildeten Kategorien mit vergleichbaren Studien, auch zu anderen naturwissenschaftlichen Arbeitsweisen und anderen Zielgruppen abgeglichen (Baur 2018; Hammann 2004; Hogan und Thomas 2001; Meister et al. 2020; Woolley et al. 2018). Dieser Entwicklungsprozess wurde begleitet durch die wiederholte Feststellung der Interrater-Übereinstimmung. Die Werte der Interrater-Übereinstimmung wurden dabei als Leitkriterium für das intersubjektive Verständnis über die Kategorien und deren Trennschärfe genutzt und bildeten die Basis für die Diskussion des Kategoriensystems und die Konsensfindung bei Nichtübereinstimmungen (Göhner und Krell 2020b). Die verbliebenen drei der neun zur Pilotierung des Kategoriensystems genutzten Fälle wurden vom Erstautor und der Mitarbeiterin analysiert, ohne weitere Änderungen am Kategoriensystem vorzunehmen. Die erneute Überprüfung der Interrater-Übereinstimmung über die drei Fälle hinweg zeigte mit  $\kappa=0,66$  eine substantielle Übereinstimmung (Landis und Koch 1977). Das so pilotierte Kategoriensystem wurde dann auf die weiteren 27 Fälle angewendet, wobei sich weitere kleine Anpassungen am Kategoriensystem ergaben (Schreier 2012).

### Auswertung der Daten in Bezug auf die Fragestellungen

Zur Beantwortung von Forschungsfragen 1 und 2 wurden die kodierten Hindernisse der ProbandInnen absolut und aufgeteilt nach den einzelnen Tätigkeiten und Phasen des Modellierens (Tab. 1) betrachtet. Zur Beantwortung von Forschungsfrage 3 wurde die Häufigkeit der einzelnen Hindernisse  $z$ -standardisiert und es wurde geprüft, inwiefern sich das Auftreten der identifizierten Hindernisse in Abhängigkeit der jeweils umgesetzten Modellierungsstrategie (Tab. 2) unterscheidet (ANOVA). Ergänzend wurde dies zwischen ProbandInnen, die Modelle zur Vorhersage genutzt haben (Strategien 3a, 3b;  $n=15$ ) und solchen, bei denen das nicht der Fall war (Strategien 1, 2a–c;  $n=21$ ), verglichen ( $t$ -Test). Alle statistischen Analysen wurden in der Softwareumgebung R vorgenommen. Um die Relevanz der identifizierten Hindernisse für die Modellierungsprozesse einschätzen zu können, wurde zudem eine kontrastierende Fallanalyse zweier ProbandInnen vorgenommen, bei denen ähnliche Hindernisse identifiziert wurden, die aber

verschiedene Modellierungsstrategien umsetzen (Schreier 2010).

## Ergebnisse

### Welche Hindernisse lassen sich in den Modellierungsprozessen identifizieren und beschreiben? (Fragestellung 1)

Innerhalb der Transkripte der ProbandInnen konnten 13 verschiedene Hinderniskategorien identifiziert werden (Tab. 3). 11 der 13 Hinderniskategorien (1–11) werden dabei von den ProbandInnen selbst verbalisiert. Bei den Hinderniskategorien 1 (*Fehlende relevante Analogie oder Erfahrung*) und 2 (*Fehlerhafte Dokumentation*) konnten zusätzlich Situationen identifiziert werden, in denen das Hindernis von außen zugeschrieben werden kann. Die Hinderniskategorien 12 (*Nichterkennen/Ignorieren von Mängeln am Modell*) und 13 (*Nichterkennen/Ignorieren von fehlender Passung*) werden nur von außen zugeschrieben und beruhen daher nicht auf den Verbalisierungen der ProbandInnen.

Insgesamt wurden 499 Hindernisse kodiert. Im Durchschnitt konnten damit je ProbandIn etwa 14 Hindernisse kodiert werden. Vier der identifizierten Hinderniskategorien kommen bei einzelnen ProbandInnen wiederholt vor. Je ProbandIn konnten drei bis zehn und durchschnittlich etwa sechs verschiedene Hinderniskategorien beobachtet werden. Zu den wiederholt auftretenden Hinderniskategorien gehören:

- *Fehlende relevante Analogie oder Erfahrung* ( $n=32$  ProbandInnen mit diesem Hindernis, insgesamt 132 Kodierungen, mittlere Häufigkeit pro ProbandIn  $M=3,7$ ),
- *Fehlerhafte Dokumentation* ( $n=30$ , insgesamt 89,  $M=2,5$ ),
- *Hinweise auf unsystematischen Input* ( $n=27$ , insgesamt 61,  $M=1,7$ ) und
- *Unzufriedenheit mit den Daten* ( $n=21$ , insgesamt 59,  $M=1,6$ ).

Andere Hinderniskategorien wie *Limitationen bei der Exploration* oder *Erwartungen an die Aufgabe* sind zwar ebenfalls häufig, treten bei den ProbandInnen aber kaum wiederholt auf. Die Hinderniskategorien *Randbeobachtung wird überbewertet*, *Mangelnde Zeichenfähigkeiten*, *Nichterkennen/Ignorieren von Mängeln am Modell*, *Nichterkennen/Ignorieren von fehlender Passung*, *Zustand des Systems unklar* und *Material unzureichend* wurden insgesamt seltener identifiziert.

**Tab. 3** Kategoriensystem der identifizierten Hindernisse in den Modellierungsprozessen von Biologie-Lehramtsstudierenden ( $N=36$ ) an einer Blackbox, absteigend sortiert nach Anzahl der ProbandInnen mit den Hindernissen

Nr	Hindernis	Beschreibung	Beispiel	Anzahl ProbandInnen mit diesem Hindernis	Anzahl Kodierungen
1	Fehlende relevante Analogie oder Erfahrung	ProbandInnen äußern, dass sie keinerlei Ideen haben oder nennen gewünschte Wirkung, für die sie jedoch keine passende Analogie oder Erfahrung nennen können. ProbandInnen fällt es schwer sich von bereits als unzureichend erklärter Idee zu lösen	„Ich habe schon eine Idee, ich komm nur nicht auf die, wie ich das, wie genau ich das mache, damit das so funktioniert, wie ich mir das vorstelle.“ (Achilles, 50)	32	132
2	Fehlerhafte Dokumentation	ProbandInnen äußern, dass Sie bei der Dokumentation einen Fehler gemacht oder etwas vergessen haben oder sind unsicher über die Qualität ihrer Dokumentation. ProbandInnen nehmen eine Beobachtung vor, aber dokumentieren diese gar nicht oder nicht richtig.	„Dann habe ich einen Fehler in meinen Notizen gemacht.“ (Cynthia, 18)	30	89
3	Hinweise auf unsystematischen Input	ProbandInnen äußern (auch retrospektiv) Probleme beim Abmessen/Einfüllen der gewünschten Wassermenge oder begründen die Wassermenge willkürlich, z. B. um bestimmte runde Zahlen zu erreichen.	„D. h. ich habe jetzt eine ungenaue Menge an Wasser rein gekippt“ (Floyd, 48)	27	61
4	Unzufriedenheit mit den Daten	ProbandInnen gehen von einem anderen Muster aus, welches z. B. durch Messfehler und verschiedene Daten entstanden ist und verfolgen dieses weiter, ProbandInnen erkennen kein Muster, da ihre Datenreihen keine Wiederholungen der Outputs aufweisen.	„Es ist so komisch, dass sich die Datensätze verändern. Das ist doch irgendwie/Ich kann einfach kein Muster erkennen.“ (Cynthia, 68)	21	59
5	Erwartungen an die Aufgabe	ProbandInnen äußern die Vermutung die Lösung der Aufgabe sei leichter oder komplexer als sie zu dem Zeitpunkt annehmen und/oder nennen andere Erwartungen, die sie nicht zufriedenstellend erfüllen können (z. B. mathematische Fähigkeiten).	„Ich hoffe das ist eigentlich kein Schülerversuch für die 5 Klasse oder so und ich beiße mir daran gerade die Zähne aus.“ (Angelina, 45)	17	32
6	Limitationen bei der Exploration	ProbandInnen kritisieren Beschränkungen im erwünschten Umgang mit der Blackbox (Öffnen, Bewegen, etc.).	„Darf man unten den Schlauch zuhalten und oben Wasser rein gießen? T: Nein, bitte nicht. P: Schade. [stöhnt]“ (Ben, 103)	15	18
7	Konkrete Analogie vorhanden, deren Funktionsweise unbekannt ist	ProbandInnen nennen konkrete Analogie, können diese in ihrer Funktionsweise jedoch nicht erklären oder zeichnen.	„So ein Überlaufschutz, der es dann komplett entleert. Wie ging sowas? Wie war denn das?“ (Jonathan, 41)	13	26
8	Zustand des Systems unklar	Der Zustand des Systems ist den ProbandInnen unklar, es werden oft Mutmaßungen über bereits vor Beginn der Untersuchung enthaltene Wassermengen aufgestellt.	„Das heißt, das muss irgendwie auf jeden Fall schon vorher Wasser drin gewesen sein.“ (Achilles, 28)	12	15
9	Mangelnde Zeichenfähigkeiten	ProbandInnen bekunden Unzufriedenheit mit ihrer Zeichnung oder bemängeln ihre Zeichenfähigkeit im Vorhinein einer Zeichnung und führen diese infolgedessen gar nicht erst durch.	„Oh ich kann sowas nicht zeichnen.“ (Frida, 13)	9	16
10	Randbeobachtung wird überbewertet	ProbandInnen interpretieren Beobachtungen wie die Fließgeschwindigkeit oder Geräusche im Inneren der BB als höchst relevant für das Lösen der Aufgabe und verwenden viel Zeit auf deren Untersuchung.	„Vielleicht hat das etwas mit Intensität zu tun.“ (Jim, 8)	8	10

**Tab. 3** (Fortsetzung)

Nr	Hindernis	Beschreibung	Beispiel	Anzahl ProbandInnen mit diesem Hindernis	Anzahl Kodierungen
11	Material unzureichend	ProbandInnen wünschen sich zusätzliche Materialien oder kritisieren vorhandenes Material.	„Um das jetzt zu testen, ohne das Gerät zu öffnen, würde ich sowas ähnliches bauen, aber ich sehe jetzt irgendwie nicht so Materialien, die sich eignen.“ (James, 25)	8	19
12	Nichterkennen/ Ignorieren von Mängeln am Modell	ProbandInnen erkennen physikalische Mängel nicht oder ignorieren diese und behalten sie aufgrund fehlender Alternativideen bei.	„da müssten die Behälter ja irgendwann voll sein Mist!“ (ignoriert und macht weiter; Jenny, 49)	4	8
13	Nichterkennen/ Ignorieren von fehlender Passung	Physikalische oder datenbasierte Nichtpassung werden von den ProbandInnen nicht erkannt oder bewusst aufgrund fehlender Alternative beibehalten.	„Eigentlich kann ich das verwerfen, weil dazu passt mein Modell nicht.“ (verwirft Modell jedoch nicht; Cynthia, 33)	4	10

### Inwiefern treten die identifizierten Hindernisse in einzelnen Phasen der Modellierungsprozesse besonders häufig auf? (Fragestellung 2)

Abb. 3 zeigt, wie viele Kodierungen je Hinderniskategorie mit den identifizierten Tätigkeiten (Tab. 1) zusammen auftreten. Die meisten Hindernisse treten zusammen mit der Tätigkeit *Zusammenfassung* auf, wobei insbesondere die Hindernisse *Fehlerhafte Dokumentation*, *Hinweise auf unsystematischen Input*, *Unzufriedenheit mit den Daten* und *Fehlende relevante Analogie oder Erfahrung* hier häufig identifiziert werden. Das Hindernis *Fehlende relevante Analogie oder Erfahrung* tritt zudem mit allen Tätigkeiten zusammen auf, die der Exploration und Herstellung zugeordnet werden, und darüber hinaus mit der Tätigkeit *Hypothese falsifiziert*, die der Phase der Vorhersage zugeordnet wird. Besonders häufig tritt das Hindernis *Fehlende relevante Analogie oder Erfahrung* dabei zusammen mit den Tätigkeiten *Beobachtung eines Phänomens*, *Prüfung Konsistenz Modell/Daten*, *Zusammenfassung* und *Analogien und Erfahrungen* auf. Alle identifizierten Hindernisse treten zudem mindestens einmal während der Tätigkeit *Prüfung Konsistenz Modell/Daten* auf.

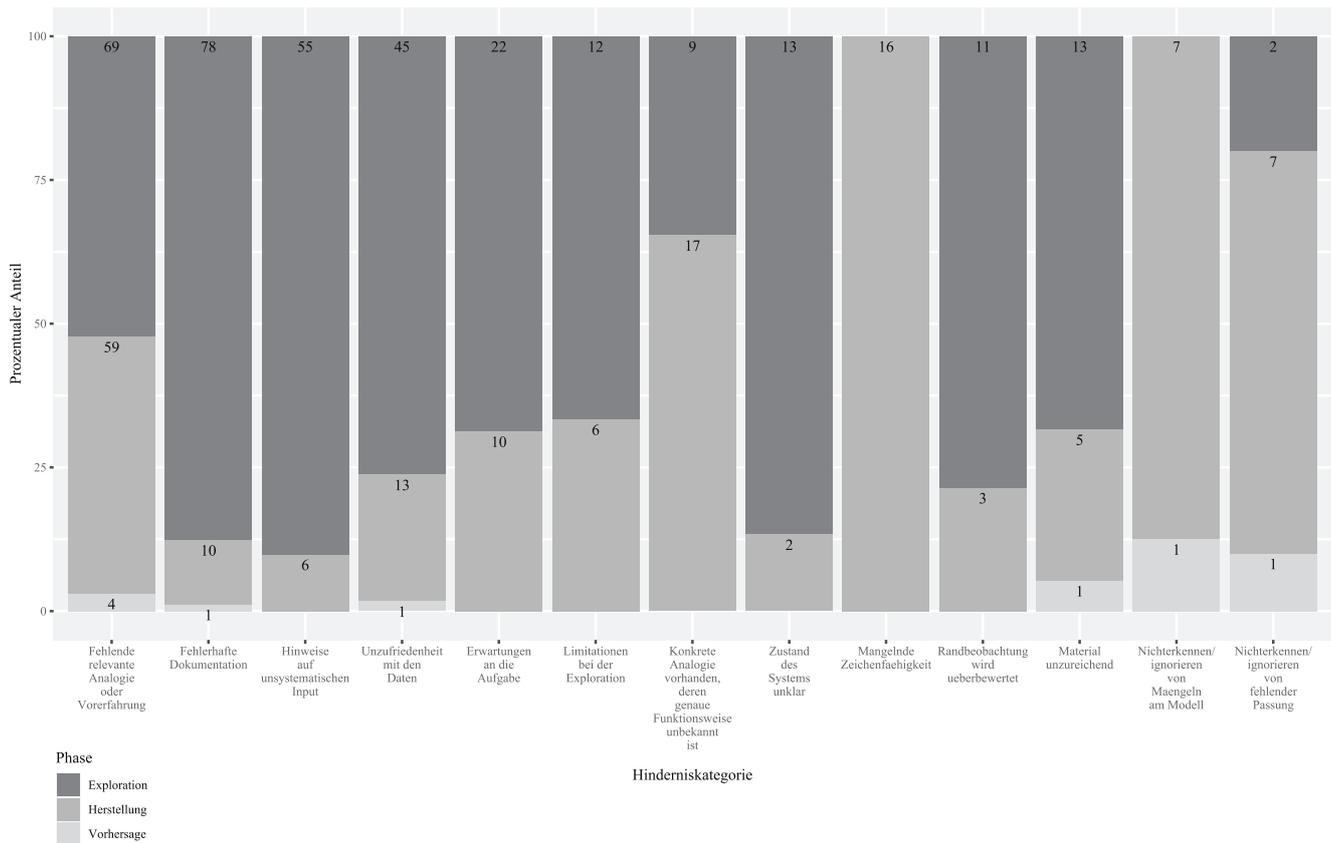
Abb. 2 zeigt die relative Häufigkeit (prozentual) der identifizierten Hindernisse in den drei Phasen des Modellierungsprozesses: Exploration, Herstellung und Vorhersage. Innerhalb der Modellierungsprozesse wurden die meisten Hindernisse in den Phasen der Exploration (299) oder der Herstellung (131) identifiziert, die auch einen Großteil des beobachtbaren Verhaltens ausmachen (68 % Tätigkeiten der Exploration, 28 % Tätigkeiten der Herstellung). Selten wurden Hindernisse dagegen in der Phase der Vorhersage (21) identifiziert, die insgesamt allerdings auch selten vorkommen. Dabei zeigen sich hindernisspezifische Unter-

schiede über die Phasen der Modellierungsprozesse hinweg. So treten zum Beispiel die Hinderniskategorien *Hinweise auf unsystematischen Input* und *Fehlerhafte Dokumentation* primär in Phasen der Exploration auf, wohingegen die Hinderniskategorien *konkrete Analogie vorhanden*, *deren genaue Funktionsweise unbekannt ist*, *Nichterkennen/ignorieren von Mängeln am Modell* und *Nichterkennen/ignorieren von fehlender Passung* vor allem in Phasen der Herstellung auftreten. Die Hinderniskategorie *Mangelnde Zeichenfähigkeit* tritt ausschließlich in der Phase der Herstellung auf. Keine der Hinderniskategorien tritt anteilig am häufigsten in der Phase der Vorhersage auf. Grundsätzlich ergeben sich die gleichen Befunde, wenn auch die zeitliche Dauer der Phasen berücksichtigt wird (Anhang, Abb. 7).

### Inwiefern besteht ein Zusammenhang zwischen den identifizierten Hindernissen und den umgesetzten Modellierungsprozessen? (Fragestellung 3)

Abb. 4 zeigt die Häufigkeit der Hindernisse in Abhängigkeit der umgesetzten Modellierungsstrategie. Für die häufigste Hinderniskategorie *Fehlende relevante Analogie oder Erfahrung* liegt der Median der ( $z$ -standardisierten) Häufigkeit beispielsweise bei  $Med = -0,62$  (Strategie 2a),  $Med = 0,31$  (Strategie 2b),  $Med = 0,12$  (Strategie 2c),  $Med = 0,49$  (Strategie 3a),  $Med = 0,12$  (Strategie 3b). Dies ist die einzige Hinderniskategorie, bei der sich nach Zusammenfassung von ProbandInnen, die keine Tätigkeiten der Vorhersage umsetzen ( $Med = -0,62$ ;  $n = 21$ ) und solchen, die Tätigkeiten der Vorhersage umsetzen ( $Med = 0,49$ ;  $n = 15$ ), augenscheinlich ein Häufigkeitsunterschied zwischen den Gruppen ergibt (Abb. 4, unten).

Statistisch zeigen sich keine signifikanten Unterschiede hinsichtlich der Häufigkeit der Hindernisse zwischen den



**Abb. 2** Gestapeltes Säulen-Diagramm aller Hindernisse unterteilt nach ihrem Auftreten in den Phasen des Modellierungsprozesses. Zur besseren Sichtbarkeit der Verhältnisse wurden die relativen (prozentualen) Häufigkeiten dargestellt

sechs Modellierungsstrategien (ANOVAs;  $p > 0,05$  für alle Hindernisse). Auch der Vergleich von ProbandInnen, die Tätigkeiten der Vorhersage umsetzen und solchen, die keine Tätigkeiten der Vorhersage umsetzen, zeigt keine signifikanten Unterschiede in der Häufigkeit der Hindernisse ( $t$ -Test;  $p > 0,05$ ).

Über die statistische Analyse hinaus liefert die Fallanalyse Evidenz für die Relevanz der identifizierten Hindernisse für die Modellierungsprozesse der ProbandInnen (Schreier 2010). Im Folgenden werden daher die ProbandInnen James und Jasmin im Detail betrachtet. Diese zeigen ähnliche Hindernisse, allerdings sehr unterschiedliche Modellierungsprozesse (Tab. 4).

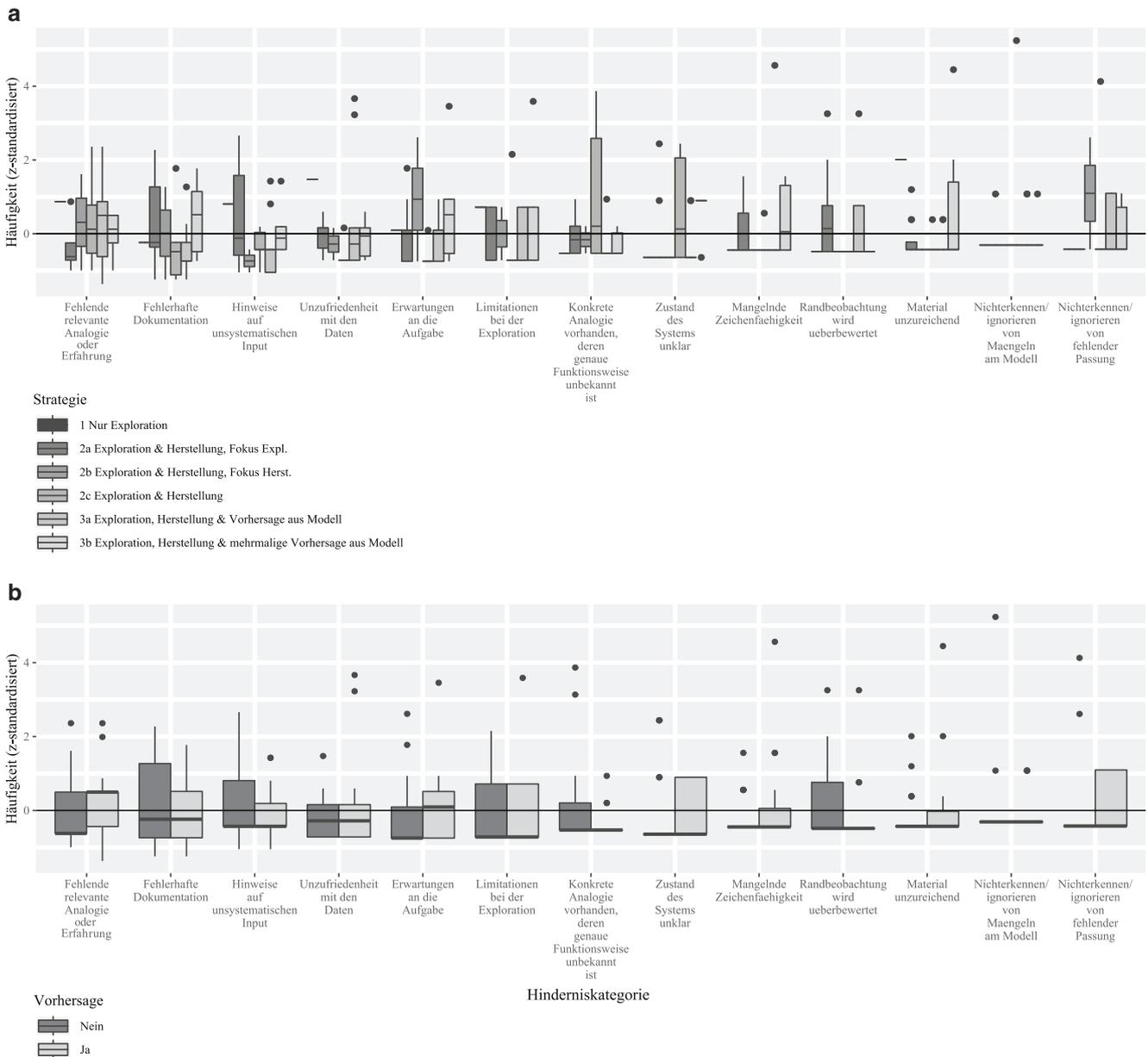
Die Modellierungsprozesse von James und Jasmin sind in Abb. 5 als Codelines illustriert. Diese Codelines zeigen die event-basiert kodierte Tätigkeiten des Modellierens. Hierbei wird die zeitliche Abfolge der Tätigkeiten als Events auf der x-Achse und die jeweils kodierte Tätigkeit auf der y-Achse dargestellt. Die kodierte Hindernisse und ihr Auftreten eingebettet in die Tätigkeiten sind als Sprechblasen dargestellt, sodass ersichtlich wird, wo in den beiden Modellierprozessen, welche Hindernisse aufgetreten sind. Der Modellierungsprozess von James besteht aus 52 kodierte Tätigkeiten und dauert etwa 27 min (Abb. 5,

**Tab. 4** Übersicht über die Häufigkeit der Hindernisse bei James und Jasmin

Nr	Problemkategorie	James	Jasmin
1	Fehlende relevante Analogie oder Erfahrung	3	1
2	Fehlerhafte Dokumentation	1	1
3	Hinweise auf unsystematischen Input	2	3
4	Unzufriedenheit mit den Daten	–	2
5	Erwartungen an die Aufgabe	–	1
6	Limitationen bei der Exploration	–	–
7	Konkrete Analogie vorhanden, deren Funktionsweise unbekannt ist	–	–
9	Mangelnde Zeichenfähigkeiten	–	–
10	Randbeobachtung wird überbewertet	–	–
10	Zustand des Systems unklar	1	–
11	Material unzureichend	3	2
12	Nichterkennen/Ignorieren von Mängeln am Modell	–	–
13	Nichterkennen/Ignorieren von fehlender Passung	–	–
–	Total	10	10

	1. Beobachtung eines Phänomens	2. Input/Output nicht hypothesengeleitet	3. Zusammenfassung	4. Input/Output hypothesengeleitet Daten	5. Mustererkennung	6. Analogien und Erfahrungen	7. Modellentwicklung	8. Optimierung Modellobjekt	9. Optimierung Modell retrospektiv	10. Verwerfen Modell	11. Prüfung Konsistenz Modellobjekt	12. Prüfung Konsistenz Modell/Daten	13. Feststellen Konsistenz	14. Hypothese aus Modell	15. Input/Output hypothesengeleitet Modell	16. Hypothese bestätigt	17. Hypothese falsifiziert	18. Optimierung Modell nach falsifizierter Hypothese	19. Verwerfen Modell nach falsifizierter Hypothese	Total
1 Fehlende relevante Analogie oder Erfahrung	30	4	19	3	1	16	8	2	4	3	12	25	2	0	0	0	3	0	0	132
2 Fehlerhafte Dokumentation	10	14	42	11	0	2	0	0	1	1	2	4	0	1	1	0	0	0	0	89
3 Hinweise auf unsystematischen Input	6	12	21	13	0	1	0	0	0	2	1	3	0	0	2	0	0	0	0	61
4 Unzufriedenheit mit den Daten	18	3	19	4	0	1	1	0	1	2	0	9	0	0	0	0	1	0	0	59
5 Erwartungen an die Aufgabe	7	3	8	3	0	4	0	0	1	3	0	2	1	0	0	0	0	0	0	32
6 Limitationen bei der Exploration	1	1	7	1	0	1	1	0	0	2	0	3	0	0	1	0	0	0	0	18
7 Konkrete Analogie vorhanden, deren Funktionsweise unbekannt ist	1	0	0	0	0	16	2	0	2	0	2	3	0	0	0	0	0	0	0	26
8 Zustand des Systems unklar	1	3	7	1	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	15
9 Mangelnde Zeichenfähigkeiten	0	0	0	0	0	2	10	1	0	0	2	1	0	0	0	0	0	0	0	16
10 Randbeobachtung wird überbewertet	0	3	6	1	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	14
11 Material unzureichend	1	4	7	1	0	1	0	0	0	0	1	3	0	0	0	1	0	0	0	19
12 Nichterkennen/Ignorieren von Mängeln am Modell	0	0	0	0	0	0	1	0	0	0	1	5	0	0	0	0	0	1	0	8
13 Nichterkennen/Ignorieren von fehlender Passung	1	0	1	0	0	1	0	0	0	0	0	4	2	0	0	0	0	1	0	10
Total	76	47	137	38	1	48	23	3	9	13	21	66	5	1	4	1	4	2	0	499

**Abb. 3** Kreuztabelle der Hinderniskategorien und Tätigkeiten aller 36 ProbandInnen, die zeigt, wie viele Kodierungen je Hinderniskategorie mit den identifizierten Tätigkeiten gemeinsam auftreten

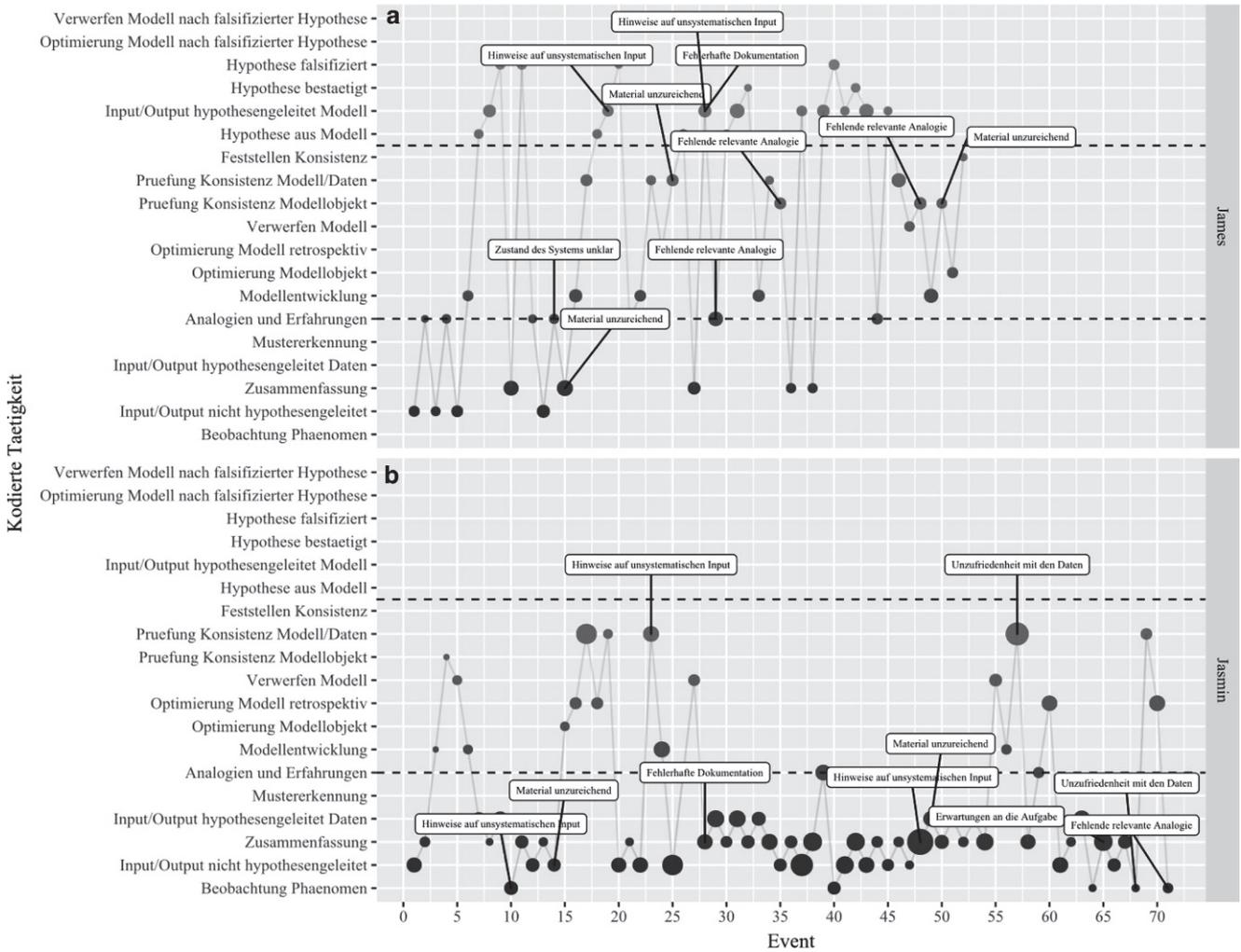


**Abb. 4** Boxplots der z-standardisierten Häufigkeit der Problemkategorien. **a** gruppiert nach Modellierungsstrategie (vgl. Tab. 2); **b** gruppiert basierend auf dem (Nicht-)Auftreten der Tätigkeit 14: „Probanden nutzen Modell, um Vorhersage über Output bei bestimmtem Input zu treffen“

oben). Hierbei durchläuft James in seinem Modellierungsprozess alternierend Phasen der Exploration, Herstellung und Vorhersage (Modellierungsstrategie 3b; vgl. Tab. 2). In James' Modellierungsprozess treten zehn Hindernisse aus sechs Hinderniskategorien auf, die sich relativ gleichmäßig auf die drei Phasen der Modellierung verteilen. Der Modellierungsprozess von Jasmin dauert 1 h und 28 min und besteht aus 71 kodierten Tätigkeiten (Abb. 5, unten). Jasmins Modellierungsprozess setzt sich hauptsächlich aus Tätigkeiten der Exploration zusammen, sowie fünf kürzeren Phasen der Herstellung, in denen sie ihr Modell entwickelt (Modellierungsstrategie 2a; vgl. Tab. 2). In Jasmins Model-

lierungsprozess wurden zehn Hindernisse aus sechs Hinderniskategorien identifiziert.

**James** Die ersten Hindernisse treten bei James auf als er, ohne einen spezifischen Grund zu nennen, die Menge des Inputs von vorher 400 ml auf 200 ml verringert (Anhang, Tab. 6; Hinderniskategorie 3) und einen Output von 350 ml beobachtet (Events 13, 14). Diese Beobachtung eines Outputs, der größer ist als die Input-Wassermenge, veranlasst ihn dazu, sich Gedanken über den Ausgangszustand der Blackbox zu machen („Also insgesamt kann natürlich im gesamten System schon vorher mehr Flüssigkeit drin ge-



**Abb. 5** Tätigkeitssequenzen der ProbandInnen James (a) und Jasmin (b), die das jeweilige Vorgehen im Modellierungsprozess sowie die darin identifizierten Hindernisse zeigen. Die Hinderniskategorie „Fehlende relevante Analogie oder Vorerfahrung“ ist aus Gründen der besseren Lesbarkeit als „Fehlende relevante Analogie“ abgekürzt

wesen sein.“; Hinderniskategorie 8). Zudem äußert er den Wunsch, nicht nur ein „theoretisches“, sondern ein gegenständliches Modell entwickeln zu können, für welches allerdings keine Materialien vorliegen (Hinderniskategorie 11). Stattdessen hantiert er mit den verfügbaren Wassergläsern, um sich vor Augen zu führen, „was passiert, wenn zwei verschiedene Gefäße mit unterschiedlicher Menge Wasser gefüllt werden“, wobei er sich in seiner Annahme bestätigt, dass „auch immer nur so viel rauskommen [kann], wie eigentlich auch reinkommt.“ James zeichnet nun ein weiteres Modell (Event 16), in welchem er die bisherige Zeichnung zweier Gefäße um eine Wippe erweitert, die er retrospektiv anhand von auditiven Beobachtungen evaluiert („Das hörte sich eher an, wie so ein Schlauch, der befüllt wird.“). Um diese Hypothese zu prüfen, gießt er anschließend eine größere Menge Wasser ein, die er nicht genauer abmisst („Jetzt ist die Menge des Wassers erstmal egal, das ich hineinfülle.“; Hinderniskategorie 3) und verwirft die Analogie einer

Wippe auf Basis seiner Beobachtung. James zieht nun ausgehend von der Beobachtung, dass bei jedem Output das Wasser eher schwallartig herausfließt, weitere Analogien heran, aus denen er ein weiteres Modell entwickelt (Events 22–25). Hierbei integriert er zuerst eine Membran, die langsames Diffundieren des Wassers in ein weiteres Gefäß mit einem Kippmechanismus erlaubt. Zur Überprüfung dieses Modells möchte er erneut ein gegenständliches Modell entwickeln, stellt aber fest: „Ich sehe jetzt irgendwie nicht so Materialien, die sich eignen.“ (Hinderniskategorie 11). Bei der weiteren Exploration der Blackbox variiert James nun weiterhin die Inputmengen (Hinderniskategorie 3), um beispielsweise zu sehen, „was passiert, wenn ich nur eine sehr geringe Menge Wasser hineingebe“ und reflektiert, dass er nicht dokumentiert oder abmisst, wie viel Wasser er hineingibt (Hinderniskategorie 2). Außerdem erwartet er, dass es zum Inneren der Blackbox eine „Entsprechung in der Natur [gibt], die [ihm] jetzt gerade noch nicht einfällt“ (Hinder-

niskategorie 1). In der Evaluation seines finalen Modells, welches erneut einen Wippenmechanismus beinhaltet (Anhang, Abb. 8), sagt er: „Leider fällt mir kein anderer Mechanismus ein [...], da bin ich jetzt auch kein Ingenieur, Techniker oder Hydrauliker, da fehlt mir einfach ein bisschen Hintergrundwissen.“ (Hinderniskategorie 1).

**Jasmin** In Jasmins Modellierungsprozess wurden 10 Hindernisse identifiziert, die vornehmlich in Situationen auftreten, in denen sie ihre Beobachtungen zusammenfasst (Tätigkeit 3, vgl. Tab. 1) oder ihr Modell retrospektiv mit beobachteten Daten vergleicht (Tätigkeit 13). Die ersten Hindernisse treten bei Jasmin in der zweiten Explorationsphase (Events 7–14) auf, als sie einen Output beobachtet, den sie, basierend auf ihren vorherigen Beobachtungen, nicht erwartet hat. Hierbei nimmt sie den beobachteten Output nicht als valides Ergebnis ihrer Untersuchung wahr, sondern schließt darauf, dass sie selbst unsystematisch vorgegangen ist („Also habe ich das falsch abgemessen.“; Hinderniskategorie 3) oder die Blackbox nicht funktionsfähig ist („Ich glaube das Gerät ist kaputt“; Hinderniskategorie 11). Tatsächlich jedoch beginnt sie ab dieser Situation, in ihrem Inputverhalten unsystematisch zu werden und variiert die nächsten Inputmengen mit dem Ziel, bestimmte, von ihr erwartete Outputs zu beobachten (Anhang, Tab. 6; Inputs 4–10). In Jasmins Modellierungsprozess können ab hier weitere Hindernisse beobachtet werden: Teilweise werden Beobachtungen nicht dokumentiert (Hinderniskategorie 2), sie äußert wiederholt, beim Abmessen ungenau vorgegangen zu sein (Hinderniskategorie 3) und begründet dies mit unzureichendem Material („Ich hätte jetzt liebend gern eine Pipette.“; Hinderniskategorie 11). Am Ende ihres Modellierungsprozesses ist sie entsprechend unzufrieden mit den von ihr aufgenommenen Daten (Hinderniskategorie 4), obwohl sie von sich selbst erwartet, die Aufgabe lösen zu können: „Ich denke für so einen Kram viel zu kompliziert. Meistens ist das so eine banale Billiglösung am Ende.“ Die Grundidee aufeinanderfolgender Gefäße behält sie über ihren gesamten Modellierungsprozess hin bei, wobei sie die Darstellungsweise sowie Anzahl und Volumina der Gefäße ändert, bis sie bei ihrem finalen Modell angekommen ist (Anhang, Abb. 8). Letztlich beschließt sie, dass ihr eine passende Idee fehlt (Hinderniskategorie 1), und beendet frustriert ihren Modellierungsprozess („Ich hab[e] keinen Plan [und] auch keinen Bock mehr. [...] [Ich] komm[e] nicht zu einer Lösung [...]. Den Rest überlasse ich dann den Profis.“).

Zusammenfassend zeigt eine qualitative Betrachtung der beiden ProbandInnen James und Jasmin, dass diese zwar durchaus ähnliche Hindernisse in ihren Modellierungsprozessen erfahren und verbalisieren, beide jedoch unterschiedlich mit diesen Hindernissen umgehen. Während James beim Auftreten von Hindernissen sein Modell verändert und wiederholt testet, konzentriert sich Jasmin

auf die weitere Exploration der Blackbox, in der Hoffnung, die erwarteten Daten zu generieren.

## Diskussion

Die vorliegende Studie hatte zum Ziel, Hindernisse zu identifizieren und zu beschreiben, die in Modellierungsprozessen von Lehramtsstudierenden bei der Untersuchung einer Blackbox auftreten. Im Folgenden werden die Ergebnisse in Bezug auf die drei Fragestellungen diskutiert.

### Welche Hindernisse lassen sich in den Modellierungsprozessen identifizieren und beschreiben? (Fragestellung 1)

Es wurden 13 Hinderniskategorien in den Modellierungsprozessen der 36 ProbandInnen identifiziert (Tab. 3). Hierbei konnten teilweise Hinderniskategorien repliziert werden, die bereits in anderen Forschungsarbeiten zum Modellieren (Hogan und Thomas 2001; Meister et al. 2020; Sins et al. 2005; Windschitl und Thompson 2006), Experimentieren (Baur 2018; Hammann 2004), sowie wissenschaftlichen Arbeiten allgemein (Woolley et al. 2018) identifiziert worden waren (Hinderniskategorien 2 (*Fehlerhafte Dokumentation*), 3 (*Hinweise auf unsystematischen Input*), 4 (*Unzufriedenheit mit den Daten*) und 10 (*Randbeobachtung wird überbewertet*)). Parallelen zum Experimentieren und wissenschaftlichen Arbeiten im Allgemeinen lassen sich insbesondere im Hinblick auf die explorativen Tätigkeiten des Modellierens und den Umgang mit den generierten Daten erkennen: Unsystematisches Vorgehen (Baur 2018; Hammann 2004; Woolley et al. 2018), die Auswahl unvollständiger oder unrelevanter Variablen (Hammann 2004; Woolley et al. 2018) und die fehlerhafte Dokumentation von Ergebnissen (Woolley et al. 2018) können arbeitsweisenunabhängig und über verschiedene Zielgruppen hinweg beobachtet werden.

Die Hinderniskategorien 1 (*Fehlende relevante Analogie oder Erfahrung*), 7 (*Konkrete Analogie vorhanden, deren Funktionsweise unbekannt ist*), 12 (*Nichterkennen/Ignorieren von Mängeln am Modell*) und 13 (*Nichterkennen/Ignorieren von fehlender Passung*) scheinen dagegen spezifisch für das Modellieren zu sein, da sie nicht in der zum Vergleich herangezogenen Studien zum Experimentieren (Baur 2018; Hammann 2004) oder wissenschaftlichen Arbeiten allgemein (Woolley et al. 2018) beschrieben werden. Die zentrale Rolle von Analogien für Modellierungen (Clement 2008; Krell et al. 2019) wird insbesondere durch das häufige Auftreten der Hinderniskategorie 1 deutlich. Gleichzeitig zeigt Hinderniskategorie 7, dass einige ProbandInnen, analog zu den Beobachtungen von Windschitl und Thompson (2006), Probleme haben,

richtig erkannte Zusammenhänge in ihre Modelle zu integrieren. Beide Hinderniskategorien stützen hierbei die Vermutung, dass Modellieren und deklaratives Fachwissen in einem *korreliert-abhängigen* Verhältnis stehen (vgl. Nehring und Schwichow 2020). Dies wird exemplarisch in James' Verbalisierungen bezüglich seines finalen Modells deutlich („Leider fällt mir kein anderer Mechanismus ein [...], da bin ich jetzt auch kein Ingenieur, Techniker oder Hydrauliker, da fehlt mir einfach ein bisschen Hintergrundwissen.“). Das Verhältnis zwischen Hinderniskategorie 1 und 7 könnte im weitgehenden Fehlen von relevantem, deklarativem Fachwissen zur Blackbox begründet sein. Mangelndes Fachwissen, fehlende relevante Analogien oder Erfahrungen oder geringe Kreativität kann das für Modellierungen zentrale abduktive Schließen behindern, also die erkenntniserweiternde Entwicklung erklärender Modelle (Magnani 1999; Krüger und Upmeyer zu Belzen 2021). Die Hinderniskategorien 4 (*Unzufriedenheit mit den Daten*), 12 (*Nichterkennen/Ignorieren von Mängeln am Modell*) und 13 (*Nichterkennen/Ignorieren von fehlender Passung*) deuten zudem darauf hin, dass der Wechsel zwischen Modell- und Erfahrungswelt, das heißt die Vermittlung zwischen Modell und generierten Daten (Dounas-Frazer et al. 2018), problematisch wird, wenn die ProbandInnen die beobachteten Daten nicht interpretieren können oder als unlogisch erachten (Hammann 2004). Solche unerwarteten, anomalen Daten, welche die Blackbox durch ihren inneren Mechanismus provoziert (insb. wenn der Output das eingefüllte Volumen eines Inputs überschreitet), werden oftmals nicht anhand eines gezeichneten Modells erklärt, sondern vielmehr methodisch unter anderem mit mangelnden eigenen Fähigkeiten begründet (Meister et al. 2020).

Obwohl Hinderniskategorie 9 (*Mangelnde Zeichenfähigkeit*) in keiner der zum Vergleich herangezogenen Studien zum Modellieren (Hogan und Thomas 2001; Meister et al. 2020; Sins et al. 2005; Windschitl und Thompson 2006), Experimentieren (Baur 2018; Hammann 2004) oder wissenschaftlichen Arbeiten allgemein (Woolley et al. 2018) beschrieben wird, weisen andere Studien durchaus auf die Relevanz des Zeichnens sowie damit verbundenen Hindernissen im Kontext naturwissenschaftlichen Unterrichts hin (Chittleborough und Treagust 2007; Quillin und Thomas 2015). Bei einem Probanden (Carlo) ist *Mangelnde Zeichenfähigkeit* als Hindernis besonders auffällig: Er zeichnet überhaupt kein Modell und begründet dies sowohl mit mangelnden Fähigkeiten als auch geringer Motivation, wodurch sein Modellierungsprozess nur aus Tätigkeiten der Exploration besteht (Strategie 1; Tab. 2). Studien aus der Repräsentationsforschung bieten hier zwei mögliche Erklärungen: Zum einen ist der Proband eventuell kein „Visualisierer“ (Koć-Januchta et al. 2017), sodass ihm das Zeichnen als nicht hilfreich zur Lösung der Aufgabe erscheint. Zum

anderen könnte er das Zeichnen aktiv vermeiden um kein „fehlerhaftes“ Modell als Endergebnis präsentieren zu müssen. Eine weitere Probandin (Iris) verwirft beispielsweise ihr finales Modell, da sie es als falsch einschätzt und nicht als Ergebnis präsentieren möchte. Bei beiden ProbandInnen wird damit zusätzlich sichtbar, dass volitionale Aspekte (z. B. Selbstregulation, Selbstwirksamkeitserwartung; Bandura 1994) bei Modellierungsprozessen eine große Rolle spielen (Hogan und Thomas 2001; Nehring et al. 2015; Nordheimer 2019). Motivationale und volitionale Aspekte sind zwar Teil gängiger Kompetenzdefinitionen (Weinert 2001), werden jedoch in der Kompetenzforschung oft explizit ausgeklammert (Klieme et al. 2007). Insbesondere zu Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung und des Modellierens liegen keine prozessorientierten Studien vor, die motivationale und volitionale Aspekte als Einflussfaktoren auf die Performanz (d. h. das Modellieren) untersuchen. Zukünftige Studien könnten hier anknüpfen und beispielsweise die Selbstregulation oder Selbstwirksamkeitserwartung während des Modellierens untersuchen (Bandura 1994; Nordheimer 2019).

Die weiteren Hinderniskategorien und 5 (*Erwartungen an die Aufgabe*), 6 (*Limitationen bei der Exploration*), 8 (*Zustand des Systems unklar*) und 11 (*Material unzureichend*) scheinen dagegen im konkreten Setting begründet zu sein. Das Auftreten der Hinderniskategorien 6 (*Limitationen bei der Exploration*) und 8 (*Zustand des Systems unklar*) könnte gerade im Zusammenhang mit den Hinderniskategorien 1 (*Fehlende relevante Analogie oder Erfahrung*) und 4 (*Unzufriedenheit mit den Daten*) dazu führen, dass die Erfahrungswelt nicht als ausreichend fruchtbar wahrgenommen wird, um einen Transfer in die Modellwelt zu ermöglichen, wenn zum Beispiel das Erkennen einer Regelmäßigkeit in den Daten ausbleibt (Leisner-Bodenthin 2006). Im Hinblick auf Hinderniskategorien 6 (*Limitationen bei der Exploration*) und 8 (*Zustand des Systems unklar*) und im Kontext von Modellieren als Problemlösen kann die Blackbox als komplexes, intransparentes Problem verstanden werden (Funke 2011). Es ist somit erwartbar, dass ähnliche Hindernisse auch bei anderen, vergleichbar intransparenten Problemen beobachtet werden können. Insgesamt stützen solche spezifisch im Setting begründeten Hindernisse theoretische Annahmen zur Abhängigkeit der Performanz von der Aufgabenstellung und von situationspezifischen Fähigkeiten (Blömeke et al. 2015). Dies wird auch in den Hinderniskategorien 11 (*Material unzureichend*) und 13 (*Erwartungen an die Aufgabe*) deutlich. Insbesondere Hinderniskategorie 13 zeigt, dass die Wahrnehmung der Situation (vgl. Blömeke et al. 2015) den Modellierungsprozess beeinflussen kann: Einige ProbandInnen (Boris, Kara; Anhang, Tab. 5) verbalisieren bereits nach wenigen Minuten, die Blackbox „gelöst“ zu haben und möchten daraufhin aufhören. Das Setting wird als kur-

ze, vollständig lösbare Aufgabe und nicht als komplexes Problem betrachtet (Hogan und Thomas 2001; Schauble et al. 1991). Berland et al. (2016) argumentieren hierzu, dass wissenschaftliche Praxis (hier: Modellierungsprozesse) und das entwickelte Wissensprodukt (hier: Modell) von verschiedenen epistemischen Überlegungen abhängen. In diesen Überlegungen sollte die Art der Antwort, die das Modell in Bezug auf das zu lösende Problem darstellt, sowie die Generalisierbarkeit, Begründbarkeit und Zielgruppe des Modells bedacht werden (Berland et al. 2016). Bennett et al. (2020) zeigten zudem bei der Untersuchung von Modellierungsprozessen in verschiedenen Kontexten, dass SchülerInnen zum Teil aus einzelnen Worten in der Aufgabenstellung (z. B. „Population“ oder „Missbildung“) spezifische Erwartungen an die Lösung der Aufgabe ableiten und dementsprechend beim Modellieren vorgehen (z. B. aktivieren sie beim Wort „Population“ Analogien oder Erfahrungen zur Reproduktion, während bei „Missbildung“ Analogien oder Erfahrungen zum Gen-zu-Protein-Schema aktiviert werden). Es liegen bislang jedoch nur vereinzelt Studien bezüglich des Einflusses epistemischer Überlegungen und spezifischer Aufgabenmerkmale in Modellierungen vor (Krell 2019; Ke und Schwarz 2020).

### **Inwiefern treten die identifizierten Hindernisse in einzelnen Phasen der Modellierungsprozesse besonders häufig auf? (Fragestellung 2)**

Hinderniskategorien, die eher explorativ und datenorientiert sind (z. B. *Fehlerhafte Dokumentation, Hinweise auf unsystematischen Input, Unzufriedenheit mit den Daten*), können eher in Phasen der Exploration beobachtet werden (Abb. 4). Demgegenüber treten modellierungsspezifische Hinderniskategorien auch eher in Phasen der Herstellung auf (z. B. *Mangelnde Zeichenfähigkeit, Nichterkennen/Ignorieren von Mängeln am Modell, Nichterkennen/Ignorieren von fehlender Passung*). Hinderniskategorie 1 (*Fehlende relevante Analogie oder Erfahrung*) tritt allerdings mit nahezu allen Tätigkeiten und damit über den gesamten Modellierungsprozess hinweg auf, was erneut die zentrale Rolle der Analogiebildung für Modellierungen unterstreicht (Clement 2008; Krell et al. 2019). Das Auftreten von eher explorativen und datenorientierten Hindernissen in Phasen der Exploration und modellierungsspezifischen Hindernissen in Phasen der Herstellung macht deutlich, dass die ProbandInnen in der Lage sind eine Mehrzahl der auftretenden Hindernisse während des Auftretens zu erkennen und zu verbalisieren. Das weniger häufig beobachtete Auftreten von Hindernissen, die eigentlich eng mit einer anderen Phase des Modellierens zusammenhängen, kann dagegen die retrospektive Reflexion des eigenen Modellierungsprozesses oder sogar eine metakognitive, zielgerichtete Steuerung des eigenen Modellierungsprozesses anzeigen (vgl. Selbstregulation; Nordheimer 2019).

Es wurde jedoch kein Fall beobachtet, in dem ein Hindernis antizipiert wird, das noch gar nicht aufgetreten ist, um es aktiv zu vermeiden. Es ist davon auszugehen, dass die retrospektive Reflexion der Hindernisse im eigenen Modellierungsprozesses auf eine größere Wirkung des jeweiligen Hindernisses hindeutet, der eine Verbalisierung und Fehlersuche nötig macht, da der Problemlöseprozess nachhaltig behindert ist. Für die Phase der Vorhersage wurden keine spezifischen Hindernisse beobachtet (Abb. 2). Hierfür sind mehrere Gründe denkbar: Entweder haben ProbandInnen, die in der Lage sind, Vorhersagen aus ihrem Modell abzuleiten, bereits alle aufkommenden Hindernisse überwunden, um auf naturwissenschaftlich adäquatem Niveau modellieren zu können, oder Hindernisse in Bezug auf die Vorhersage äußern sich eher im Fehlen bestimmter Tätigkeiten (z. B. fehlende Hypothesenbildung; vgl. Baur 2018). In Anbetracht der bisherigen Forschungsergebnisse zur Modellkompetenz von Lehrkräften (Krell und Krüger 2016; Nielsen und Nielsen 2019; Windschitl und Thompson 2006) erscheint letzteres wahrscheinlicher; gerade da die hier beobachteten Modellierungsprozesse durchaus variabel sind (Quigley et al. 2017) und selten Tätigkeiten der Vorhersage enthalten. Allerdings muss auch beachtet werden, dass das seltene Auftreten vorhersagender Tätigkeiten (vgl. Tab. 2) die Wahrscheinlichkeit verringert, dort typischerweise auftretende Hindernisse auch tatsächlich beobachten zu können.

Auf Ebene der Tätigkeiten (Tab. 1) können darüber hinaus die *Zusammenfassung* und die *Prüfung der Konsistenz zwischen dem Modell und den Daten* als Schlüssel-tätigkeiten identifiziert werden: ProbandInnen verbalisieren am häufigsten Hindernisse während der Tätigkeit *Zusammenfassung*, die analog zu den Ergebnissen von Sins et al. (2005) auch die häufigste Tätigkeit darstellt (Tab. 3). ProbandInnen versuchen bei der *Zusammenfassung* ihre Beobachtungen zu interpretieren, ohne sie in Bezug zum Modell zu setzen, wobei besonders die explorativen und datenorientierten Hinderniskategorien auftreten. Während der Tätigkeit *Prüfung der Konsistenz zwischen dem Modell und den Daten* treten dagegen alle identifizierten Hinderniskategorien mindestens einmal auf (Tab. 3). Die vorliegenden Ergebnisse stützen damit die Annahme, dass das Testen von Modellen einen Schlüsselaspekt zur Förderung der Modellkompetenz von Lehrkräften darstellt (Krell und Krüger 2016; Lee und Kim 2014), da während des Testens von Modellen potenziell alle identifizierten Hinderniskategorien diskutiert werden können.

### Inwiefern besteht ein Zusammenhang zwischen den identifizierten Hindernissen und den umgesetzten Modellierungsprozessen? (Fragestellung 3)

ProbandInnen mit ähnlichen Hindernissen können sich hinsichtlich ihrer Modellierungsprozesse deutlich unterscheiden, sodass kein eindeutiger Zusammenhang zwischen den identifizierten Hindernissen und den beobachteten Modellierungsprozessen vorzuliegen scheint (Abb. 4). Dies wird insbesondere in der kontrastierenden Fallanalyse deutlich, die aufzeigt, dass James und Jasmin zwar ähnliche Hindernisse verbalisieren, deren Umgang mit den Hindernissen aber unterschiedlich ausfällt (Abb. 5). Während James beim Auftreten von Hindernissen zuerst sein Modell verändert, konzentriert sich Jasmin auf die weitere Exploration der Blackbox, in der Hoffnung, bessere Daten zu generieren. Bei der absichtlichen Variation des Inputs geht sie jedoch eher unsystematisch vor, und möchte, wie viele der ProbandInnen, einen bestimmten Output erzielen. Dieses Vorgehen weist Parallelen zum *model-fitting* auf, wobei allerdings nicht das Modell, sondern die Inputmenge schrittweise variiert wird (vgl. Hogan und Thomas 2001; Sins et al. 2005). Jasmin ist in dieser Situation zwar in der Lage, das eigene Modell zu reflektieren (vgl. Windschitl und Thompson 2006) und Hindernisse zu erkennen, schafft es aber nicht, alternative Handlungen auszuführen. James dagegen nutzt sein Modell, um Hindernisse zu überwinden, indem er wiederholt Hypothesen von seinem Modell ableitet und testet. Ähnliche Muster konnten auch bei weiteren, hier nicht näher dargestellten ProbandInnen beobachtet werden. Durch das prinzipielle Auftreten ähnlicher Hinderniskategorien bei ProbandInnen, die unterschiedliche Modellierungsprozesse umsetzen, ist somit davon auszugehen, dass das Auftreten unterschiedlicher Hindernisse die Modellierung im Sinne der naturwissenschaftlichen Erkenntnisgewinnung nicht verhindert – vielmehr erscheint der Umgang mit auftretenden Hindernissen während des Modellierungsprozesses zentral für das erfolgreiche Problemlösen (Funke 2011; Mayer 2007). Hierbei sollten eine systematische Exploration und die Entwicklung kreativer Ideen und Analogien, welche das Potenzial haben, diese Beobachtungen zu erklären, als sich ergänzende Voraussetzungen gesehen werden (Krell et al. 2019).

### Limitationen

Die Blackbox wurde in der vorliegenden Studie einem authentischen naturwissenschaftlichen Kontext vorgezogen, um die inhaltliche Komplexität des Problems zu reduzieren. Hierdurch ist für die Untersuchung der Blackbox kein ausgeprägtes (naturwissenschaftliches) Vorwissen notwendig und es kann auf den Problemlöseprozess fokussiert werden („könnenszentriertes Problemlösen“; Frieger und Lind

2003). Zusätzlich konnte kein systematischer Einfluss der im Samplingprozess berücksichtigten Kompetenzen des wissenschaftlichen Denkens und allgemeinen kognitiven Fähigkeiten der ProbandInnen auf die Modellierungsprozesse und beobachteten Hindernisse festgestellt werden (Göhner und Krell in Druck). Dies deckt sich mit Beobachtungen anderer Studien zu Problemlöseprozessen (vgl. Mayer 2007). Für die Modellierung der hier verwendeten Blackbox kann allerdings physikalisches Wissen hilfreich sein (z.B. im Bereich Hydraulik). Hierdurch kann eine Übertragbarkeit der vorliegenden Ergebnisse auf die Modellierung authentischer naturwissenschaftlicher Probleme limitiert sein, da wie bereits dargestellt, inhaltliches Wissen einen wichtigen Faktor für die Umsetzung von Modellierungsprozessen darstellt (Ruppert et al. 2019; Hinderniskategorie 1). Zudem können trotz der Aufforderung zum lauten Denken Teile der Denkprozesse von der Verbalisierung der ProbandInnen ausgeschlossen sein. Dabei bedingt die Beschränkung auf direkte Verbalisierungen (*concurrent thinking aloud*) und die beobachtbaren Handlungen auch, dass zum Beispiel mentale Modelle der ProbandInnen nicht berücksichtigt werden konnten. So wurden nur die an der Tafel gezeichneten Modelle auch als Modelle gewertet, während reine Verbalisierungen als Analogiebildung (Tab. 1) interpretiert wurden. Schließlich könnte das seltene Auftreten vorhersagender Tätigkeiten die Wahrscheinlichkeit verringert haben, entsprechende Hindernisse beobachten zu können. Eine Erweiterung der Stichprobe könnte die hierdurch angedeutete Problematik einer mangelnden Sättigung prüfen.

### Implikationen

Im Rahmen der Kompetenzmodellierung und -erfassung können die identifizierten Hindernisse langfristig zur Unterscheidung von Problemtypen des Modellierens beitragen (Gut et al. 2014; Hild 2020), indem Modellierprobleme zum Beispiel auf der Basis typischer Hindernisprofile unterschieden werden. Gut et al. (2014) nutzen den Problemtypenansatz beispielsweise, um die Vielfalt experimenteller Aktivitäten in vier Problemtypen einzuteilen und für jeden Problemtyp eine spezifische Kompetenzprogression zu modellieren. Dies erlaubt die problemtypenbasierte Entwicklung von Aufgaben, die zu Messinstrumenten gebündelt werden können (z.B. Hild 2020). In Bezug auf Modellkompetenz könnte die Identifikation der hier beschriebenen Hindernisse in verschiedenen Modellierungsaufgaben damit zur Differenzierung vergleichbarer Problemtypen beitragen.

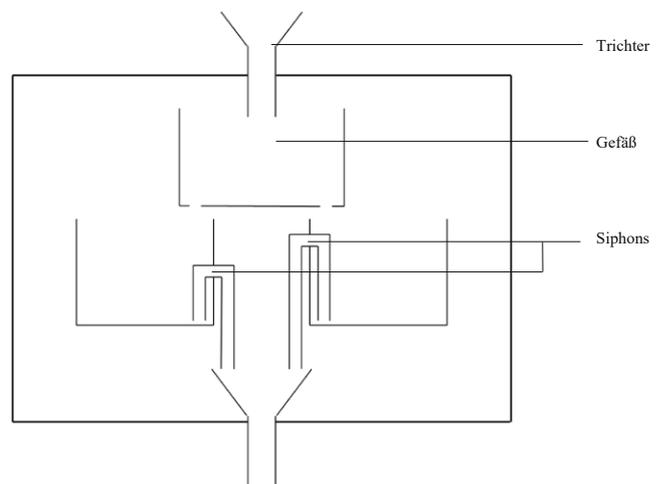
In Interventionen können die identifizierten und beschriebenen Hinderniskategorien genutzt werden, um angemessene Hilfestellungen und Impulse zu planen, die Lehrkräften bei Hindernissen im Modellierungsprozess

spontan präsentiert werden, sodass eine nachhaltige Behinderung des Modellierungsprozesses vermieden wird (Baur 2018; Sung und Oh 2018). So sind beispielsweise bezüglich der explorativen Hinderniskategorien (z. B. *Fehlerhafte Dokumentation, Hinweise auf unsystematischen Input, Unzufriedenheit mit den Daten*) Hilfestellungen in Form von externen, gesicherten Datenquellen denkbar, während beim Auftreten von Hinderniskategorie 1 (*fehlende relevante Analogie oder Erfahrung*) eine Auswahl von Beispielmechanismen oder zu überprüfenden Modellen vorgelegt werden könnte. Im hier präsentierten Setting könnte bei Auftreten von Hinderniskategorie 8 (*Zustand des Systems unklar*) der aktuelle Inhalt der Blackbox vom Versuchsleiter mitgeteilt werden, ohne darüber hinaus Hinweise zum inneren Mechanismus der Blackbox zu geben.

## Fazit

Die vorliegenden Ergebnisse können einen Ausgangspunkt für die Entwicklung problemtypenbasierter Kompetenzmodellierungen darstellen und so beispielsweise für die weitere Entwicklung von Diagnoseinstrumenten verwendet werden (Baur 2018; Gut et al. 2014). Gleichzeitig können mit Rücksicht auf die identifizierten Hindernisse gezielte und explizite Förderangebote konstruiert werden, die Modelle als Werkzeuge zur Erkenntnisgewinnung wahrnehmbar und die Praxis des wissenschaftlichen Modellierens auch im schulischen Kontext umsetzbar machen (Clement 2000; Lehrer und Schauble 2006). Weitere Forschung ist nötig, um die Generalisierbarkeit der gezeigten Hindernisse sicherzustellen und den Einfluss epistemischer Überlegungen und spezifischer Aufgabenmerkmale auf das Modellieren systematisch zu untersuchen.

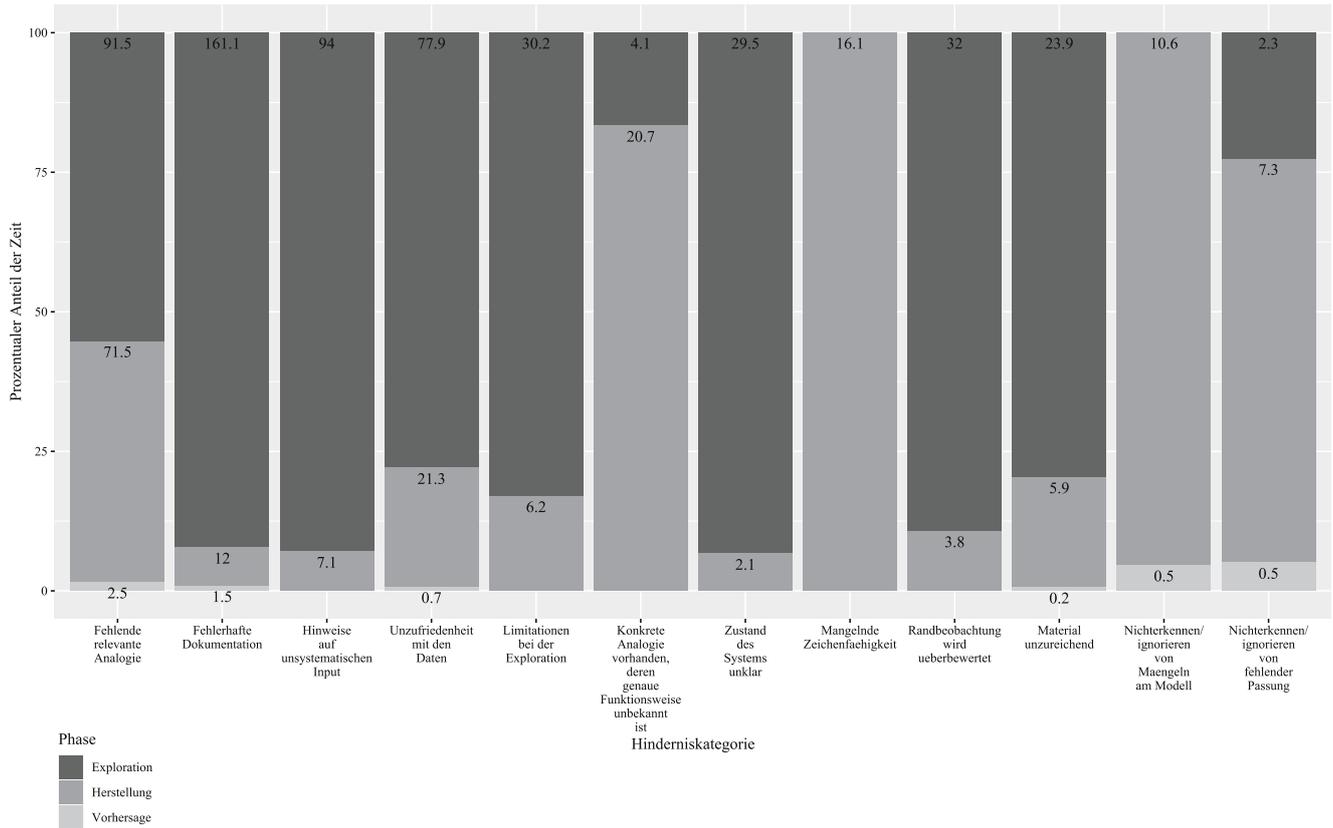
## Anhang



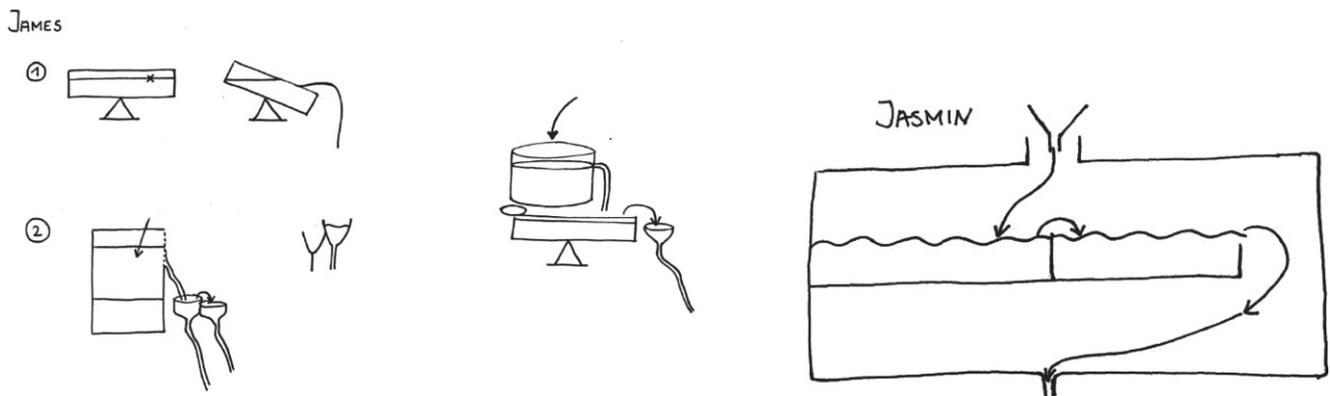
**Abb. 6** Darstellung des inneren Mechanismus der Black. Wasser wird durch einen Trichter in ein erstes Gefäß geleitet, das dieses gleichmäßig auf zwei weitere Gefäße verteilt. Wird ein bestimmter Wasserstand in einem der beiden unteren Gefäße erreicht entleert sich das jeweilige Gefäß vollständig durch den Siphon. Da die beiden Siphons in unterschiedlichen Höhen angebracht sind, unterscheidet sich der eine Entleerung auslösende Wasserstand bei beiden Gefäßen (*links 350 ml, rechts 550 ml*)

**Tab. 5** Übersicht über die teilnehmenden Lehramtsstudierenden, alphabetisch geordnet nach Pseudonym

Pseudonym	Alter	Studiengang	Studienfach 1	Studienfach 2	Untersuchungszeit
Achilles	25	Bachelor	Biologie	Politik	01:32:02
Alexandra	20	Bachelor	Ernährungswissenschaften	Biologie	01:38:26
Alice	22	Bachelor	Biologie	Physik	01:38:22
Amy	21	Bachelor	Ernährungswissenschaften	Biologie	01:22:13
Angelina	21	Bachelor	Biologie	Mathematik	01:01:17
Ben	19	Bachelor	Biologie	Politik	01:47:20
Boris	19	Bachelor	Biologie	Geschichte	00:11:06
Carlo	33	Master	Biologie	Englisch	01:15:29
Celine	17	Bachelor	Ernährungswissenschaften	Biologie	00:50:52
Claire	21	Bachelor	Biologie	Deutsch	01:32:42
Claudia	30	Bachelor	Biologie	Politik	00:33:30
Cynthia	29	Master	Sozialkunde	Biologie	01:40:43
Daphne	18	Bachelor	Biologie	Spanisch.	00:50:39
Donna	44	Bachelor	Biologie	Philosophie	01:23:46
Floyd	24	Master	Biologie	Informatik	00:33:55
Frida	29	Master	Kunst	Biologie	01:49:12
Iris	28	Bachelor	Biologie	Mathematik	01:25:16
James	39	Bachelor	Agrarwissenschaften	Biologie	00:26:37
Jasmin	25	Master	Biologie	Geschichte	01:28:25
Jenny	21	Bachelor	Deutsch	Biologie	01:52:03
Jim	24	Master	Biologie	Mathematik	01:04:14
Jonathan	25	Bachelor	Biologie	Geschichte	00:50:13
Kara	25	Master	Biologie	Mathematik	00:08:07
Lana	23	Bachelor	Englisch	Biologie	00:48:33
Lauren	19	Bachelor	Deutsche Philologie	Biologie	01:16:53
Martin	33	Master	Biologie	Chemie	01:52:22
May	27	Master	Biologie	Mathematik	01:27:16
Misty	24	Master	Biologie	Mathematik	01:25:20
Natalie	21	Bachelor	Biologie	Geschichte	01:23:55
Raphael	28	Bachelor	Biologie	Chemie	01:05:58
Rocco	24	Master	Grundschulpädagogik	NAWI	00:29:28
Ryan	30	Master	Englisch	Biologie	01:05:03
Sabrina	18	Bachelor	Biologie	Geschichte	00:34:05
Selena	24	Master	Biologie	Deutsch	00:43:56
Susi	33	Master	Biologie	Politik	01:34:54
Valerie	21	Bachelor	Deutsch	Biologie	00:46:21



**Abb. 7** Gestapeltes Säulen-Diagramm des zeitlichen Anteils aller Hindernisse unterteilt nach ihrem Auftreten in den Phasen des Modellierungsprozesses. Zur besseren Sichtbarkeit der Verhältnisse wurde der relative (prozentuale) zeitliche Anteil dargestellt. Die Säulen sind zusätzlich mit den absoluten Werten beschriftet, die die Minuten angeben, in denen das Hindernis in der jeweiligen Phase auftritt. Im Vergleich zur in Abbildung 2 dargestellten Häufigkeit der Hindernisse ergeben sich nur geringfügige Unterschiede



**Abb. 8** Finale Modelle von James und Jasmin

**Tab. 6** Die vorgenommenen Inputs und beobachteten Outputs der beiden ProbandInnen James und Jasmin (?=Der In- oder Output wurde durch die ProbandInnen nicht verbalisiert und ist auch videografisch nicht rekonstruierbar)

James			Jasmin		
Nr.	Input	Output	Nr.	Input	Output
1	400	0	1	400	0
2	400	400	2	400	400
3	400	600	3	400	640
4	200	350	4	200	0
5	?	?	5	40	75
6	?	?	6	400	?
7	100	0	7	200	0
8	100	0	8	100	0
9	100	0	9	100	?
10	100	0	10	?	?
11	100	0	11	400	0
12	100	?	12	400	400
13	400	1000	13	400	620
			14	400	400
			15	400	0
			16	400	880
			17	400	0
			18	400	400
			19	400	600
			20	400	400
			21	400	0
			22	400	1000
			23	1000	500
			24	100	?
			25	600	300

**Danksagung** Die Autoren danken der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung des Projekts TypMoL (Projekt-nummer 327507949).

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**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

- ACARA (2015). The Australian curriculum: science. Australian curriculum, assessment and reporting authority. <http://www.australiancurriculum.edu.au/>. Zugegriffen: 11. Aug. 2020.
- Acher, A., Arcá, M., & Sanmartí, N. (2007). Modeling as a teaching learning process for understanding materials: a case study in primary education. *Science Education*, 91(3), 398–418.
- Bandura, A. (1994). Self-Efficacy. In V. S. Ramachandran (Hrsg.), *Encyclopedia of human behavior* (S. 71–81). New York: Academic Press.
- Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 24, 115–129.
- BCMOE (2019). Curriculum redesign. British columbia ministry of education. <https://curriculum.gov.bc.ca/rethinking-curriculum>. Zugegriffen: 12. Sept. 2020.
- Bell, R. L., Blair, L. M., Crawford, B. A., & Lederman, N. G. (2003). Just do it? Impact of a science apprenticeship program on high school students' understandings of the nature of science and scientific inquiry. *Journal of Research in Science Teaching*, 40(5), 487–509.
- Bennett, S., Gotwals, A. W., & Long, T. M. (2020). Assessing students' approaches to modelling in undergraduate biology. *International Journal of Science Education*, 42(10), 1697–1714.
- Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: making scientific practices meaningful for students. *Journal of Research in Science Teaching*, 53(7), 1082–1112.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift Für Psychologie*, 223(1), 3–13.
- Bo, W. V., Fulmer, G. W., Lee, C. K.-E., & Chen, V. D.-T. (2018). How do secondary science teachers perceive the use of interactive simulations? The affordance in Singapore context. *Journal of Science Education and Technology*, 27, 550–565.
- Bybee, R. W. (2002). Scientific Literacy: Mythos oder Realität? In W. Gräber, P. Nentwig, T. R. Koballa & R. Evans (Hrsg.), *Scientific literacy* (S. 21–43). Wiesbaden: Springer VS.
- Campbell, T., Oh, P. S., Maughn, M., Kiriazis, N., & Zuwallack, R. (2015). A review of modeling pedagogies: pedagogical functions, discursive acts, and technology in modeling instruction. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(1), 159–176.
- Cheng, M.-F., Wu, T.-Y., & Lin, S.-F. (2019). Investigating the relationship between views of scientific models and modeling practice. *Research in Science Education*. <https://doi.org/10.1007/s11165-019-09880-2>.
- Chittleborough, G., & Treagust, D. F. (2007). The modelling ability of non-major chemistry students and their understanding of the sub-microscopic level. *Chemistry Education Research and Practice*, 8(3), 274–292.
- Chiu, M.-H., & Lin, J.-W. (2019). Modelling competence in science education. Position Paper. *Disciplinary and Interdisciplinary Science Education Research*. <https://doi.org/10.1186/s43031-019-0012-y>.
- Clement, J. (1989). Learning via model construction and criticism. In J. A. Glover, R. R. Ronning & C. R. Reynolds (Hrsg.), *Handbook of Creativity* (S. 341–381). Boston: Springer US.
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9), 1041–1053.
- Clement, J. (2008). *Creative model construction in scientists and students: the role of imagery, analogy, and mental simulation*. Dordrecht: Springer.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung* (2. Aufl.). Stuttgart: Kohlhammer.

- Dounas-Frazer, D. R., Ríos, L., Pollard, B., Stanley, J. T., & Lewandowski, H. (2018). Characterizing lab instructors' self-reported learning goals to inform development of an experimental modeling skills assessment. *Physical Review Physics Education Research*, 14(2), 20118.
- Driver, R., John, L., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University Press.
- Elby, A., & Hammer, D. (2010). Epistemological resources and framing: a cognitive framework for helping teachers interpret and respond to their students' epistemologies. In L. D. Bendixen & F. C. Feucht (Hrsg.), *Personal epistemology in the classroom: theory, research, and implications for practice* (S. 409–434).
- Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3), 178–186.
- Fleige, J., Seegers, A., Upmeyer zu Belzen, A., & Krüger, D. (2012). Förderung von Modellkompetenz im Biologieunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, 65(1), 19–28.
- Friege, G., & Lind, G. (2003). Allgemeine und fachspezifische Problemlösekompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 63–74.
- Funke, J. (2011). Problemlösen. In T. Betsch, J. Funke & H. Plessner (Hrsg.), *Denken – Urteilen, Entscheiden, Problemlösen* (S. 136–199). Berlin, Heidelberg: Springer.
- Giere, R. N., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning*. Belmont: Thomson Wadsworth.
- Gilbert, J. K., & Justi, R. (2016). *Modelling-based teaching in science education*. Cham: Springer.
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and philosophy*, 21(5), 725–740.
- Göhner, M., & Krell, M. (2018). Modellierungsprozesse von Lehramtsstudierenden der Biologie. *Erkenntnisweg Biologiedidaktik*, 17, 45–63.
- Göhner, M., & Krell, M. (2020a). Preservice science teachers' strategies in scientific reasoning: the case of modeling. *Research in Science Education*. <https://doi.org/10.1007/s11165-020-09945-7>.
- Göhner, M., & Krell, M. (2020b). Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review. *Zeitschrift Für Didaktik Der Naturwissenschaften*. <https://doi.org/10.1007/s40573-020-00111-0>.
- Göhner, M., & Krell, M. (in Druck). Analyzing the relationships between pre-service biology teachers' modeling processes, scientific reasoning competencies and general cognitive abilities. In M. Ergazaki & K. Kampourakis (Hrsg.), *Contributions from Biology Education Research*. Berlin Heidelberg: Springer.
- Gouvea, J., & Passmore, C. M. (2017). "Models of" versus "Models for". *Science & Education*, 26(1-2), 49–63.
- Günther, S. L., Fleige, J., Upmeyer zu Belzen, A., & Krüger, D. (2017). Interventionsstudie mit angehenden Lehrkräften zur Förderung von Modellkompetenz im Unterrichtsfach Biologie. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals: Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 215–236). Wiesbaden: Springer VS.
- Günther, S. L., Fleige, J., Upmeyer zu Belzen, A., & Krüger, D. (2019). Using the case method to foster preservice biology teachers' content knowledge and pedagogical content knowledge related to models and modeling. *Journal of Science Teacher Education*, 30(4), 321–343.
- Gut, C., Metzger, S., Hild, P., & Tardent, J. (2014). Problemtypenbasierte Modellierung und Messung experimenteller Kompetenzen von 12- bis 15-jährigen Jugendlichen. *Didaktik der Physik Frühjahrstagung*. [https://phzh.ch/MAPortrait\\_Data/158541/11/Gut\\_et\\_al\\_2014\\_Problemtypenbasierte\\_Modellierung.pdf](https://phzh.ch/MAPortrait_Data/158541/11/Gut_et_al_2014_Problemtypenbasierte_Modellierung.pdf). Zugegriffen: 12. Aug. 2020.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle: Merkmale und ihre Bedeutung - dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 57(4), 196–203.
- Hartmann, S., Upmeyer zu Belzen, A., Krüger, D., & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of preservice science teachers' competencies. *Zeitschrift für Psychologie*, 223(1), 47–53.
- Hild, P. (2020). *Problemtypenbasierte Kompetenzmodellierung beim praktisch-naturwissenschaftlichen Arbeiten: Design, Validierung und Einsatz von Aufgaben zum effektbasierten Vergleichen (Dissertation)*. Kiel: Christian-Albrechts-Universität.
- 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.
- Hogan, K., & Thomas, D. (2001). Cognitive comparisons of students' systems modeling in ecology. *Journal of Science Education and Technology*, 10(4), 319–345.
- Justi, R., & Gilbert, J. K. (2002). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. *International Journal of Science Education*, 24(12), 1273–1292.
- Justi, R., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education*. 25(11), 1369–1386. <https://doi.org/10.1080/0950069032000070324>.
- Ke, L., & Schwarz, C. V. (2020). Supporting students' meaningful engagement in scientific modeling through epistemological messages: a case study of contrasting teaching approaches. *Journal of Research in Science Teaching*. <https://doi.org/10.1002/tea.21662>.
- Khan, S. (2011). What's missing in model-based teaching. *Journal of Science Teacher Education*, 22(6), 535–560.
- 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). Bonn & Berlin: BMBF.
- KMK (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. München & Newied: Wolters Kluwer.
- KMK (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_16-Fachprofile-Lehrerbildung.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf). Zugegriffen: 12. Aug. 2020.
- KMK (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife*. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2020/2020\\_06\\_18-BildungsstandardsAHR\\_Biologie.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf). Zugegriffen: 16. Febr. 2021.
- Knuutila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, 42(2), 262–271.
- Koč-Januchta, M., Höffler, T., Thoma, G.-B., Precht, H., & Leutner, D. (2017). Visualizers versus verbalizers: effects of cognitive style on learning with texts and pictures. *Computers in Human Behavior*, 68, 170–179.
- Krell, M. (2019). Assessment of meta-modeling knowledge: learning from triadic concepts of models in the philosophy of science. *Advance online*. <https://doi.org/10.18452/19813>.
- Krell, M., & Hergert, S. (2019). The blackbox approach: Analyzing modeling strategies. In A. Upmeyer zu Belzen, D. Krüger & J. H. van Driel (Hrsg.), *Towards a competence-based view on models and modeling in science education* (S. 147–160). Cham: Springer.
- Krell, M., & Krüger, D. (2013). Wie werden Modelle im Biologieunterricht eingesetzt? Ergebnisse einer Fragebogenstudie. *Erkenntnisweg Biologiedidaktik*, 12, 9–26.
- Krell, M., & Krüger, D. (2016). Testing models: a key aspect to promote teaching activities related to models and modelling in biology lessons? *Journal of Biological Education*, 50(2), 160–173. <https://doi.org/10.1080/00219266.2015.1028570>.

- Krell, M., Upmeier zu Belzen, A., & Krüger, D. (2014). Students' levels of understanding models and modelling in biology: global or aspect-dependent? *Research in Science Education*, 44(1), 109–132. <https://doi.org/10.1007/s11165-013-9365-y>.
- Krell, M., Upmeier zu Belzen, A., & Krüger, D. (2016). Modellkompetenz im Biologieunterricht. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologiedidaktische Forschung: Schwerpunkte und Forschungsgegenstände* (S. 83–102). Berlin: Logos.
- Krell, M., Walzer, C., Hergert, S., & Krüger, D. (2019). Development and application of a category system to describe pre-service science teachers' activities in the process of scientific modelling. *Research in Science Education*, 49(5), 1319–1345. <https://doi.org/10.1007/s11165-017-9657-8>.
- Krüger, D., Hartmann, S., Nordmeier, V., & Upmeier zu Belzen, A. (2020). Measuring scientific reasoning competencies: multiple aspects of validity. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper & C. Lautenbach (Hrsg.), *Student learning in German higher education* (S. 261–280). Wiesbaden: Springer VS.
- Krüger, D., & Upmeier zu Belzen, A. (2021). Kompetenzmodell der Modellierkompetenz: Die Rolle abduktiven Schließens beim Modellieren. *Zeitschrift für Didaktik der Naturwissenschaften*. <https://doi.org/10.1007/s40573-021-00129-y>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lange, D. (2013). *Inhaltsanalytische Untersuchung zur Kooperation beim Bearbeiten mathematischer Problemaufgaben*. Münster: Waxmann.
- Lederman, N. G., & Abd-El-Khalick, F. (2002). Avoiding de-natured science: Activities that promote understandings of the nature of science. In W. McComas (Hrsg.), *The Nature of Science in Science Education*. Dordrecht: Springer.
- Lee, S., & Kim, H.-B. (2014). Exploring secondary students' epistemological features depending on the evaluation levels of the group model on blood circulation. *Science & Education*, 23(5), 1075–1099.
- Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. *Handbook of child psychology*, 6, 153–196.
- Leisner-Bodenthin, A. (2006). Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift Für Didaktik der Naturwissenschaften*, 12, 91–109.
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000R: IST 2000R*. Göttingen: Hogrefe.
- Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471–492.
- Magnani, L. (1999). Model-based creative abduction. In L. Magnani, N. J. Nersessian & P. Thagard (Hrsg.), *Model-based reasoning in scientific discovery* (S. 219–238). : Kluwer.
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung* (S. 177–186). Berlin, Heidelberg: Springer.
- Meister, S., Krell, M., Göhner, M., & Upmeier zu Belzen, A. (2020). Pre-service biology teachers' responses to first-hand anomalous data during modelling Processes. *Research in Science Education*. <https://doi.org/10.1007/s11165-020-09929-7>.
- Mikelskis-Seifert, S., & Leisner, A. (2005). Investigation of effects and stability in teaching model competence. In K. Boersma, M. Goedhart, O. de Jong & H. Eijkelhof (Hrsg.), *Research and the Quality of Science Education* (S. 337–351). Dordrecht: Springer.
- Nehring, A., & Schwichow, M. (2020). Was ist Wissen, was ist Können? Deutungen des Kompetenzbegriffs und deren psychometrische Konsequenzen im Kontext von Fachwissen und Variablenkontrollstrategie. *Zeitschrift für Didaktik der Naturwissenschaften*. <https://doi.org/10.1007/s40573-020-00113-y>.
- Nehring, A., Nowak, K. H., Upmeier zu Belzen, A., & Tiemann, R. (2015). Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic variables. *International Journal of Science Education*, 37(9), 1343–1363.
- Nersessian, N. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich & M. Siegal (Hrsg.), *The cognitive basis of science* (S. 133–153). New York: Cambridge University Press.
- Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki & T. Hirano (Hrsg.), *Model science teacher preparation programs: An international comparison of what works*. Charlotte: Information Age.
- NGSS Lead States (2013). Next generation science standards: for states, by states: appendix f - science and engineering practices in the ngss. In NGSS Lead States (Hrsg.), *Next generation science standards: for states, by states*. Washington, DC: The National Academies Press.
- Nicolaou, C. T., & Constantinou, C. P. (2014). Assessment of the modeling competence: a systematic review and synthesis of empirical research. *Educational Research Review*, 13, 52–73.
- Nielsen, S. S., & Nielsen, J. A. (2019). A competence-oriented approach to models and modelling in lower secondary science education: Practices and rationales among danish teachers. *Research in Science Education*. <https://doi.org/10.1007/s11165-019-09900-1>.
- Nordheimer, R. (2019). *Volition in Modellierungsprozessen von Biologie-Lehramtsstudierenden: Entwicklung eines Kategoriensystems zur Erfassung der Selbststeuerung* (unveröffentlichte Master Thesis). Freie Universität, Berlin.
- Osborne, J. (2013). The 21st century challenge for science education: assessing scientific reasoning. *Thinking skills and creativity*, 10, 265–279.
- Passmore, C. M., & Svoboda, J. (2012). Exploring opportunities for argumentation in modeling classrooms. *International Journal of Science Education*, 34(10), 1535–1554.
- Passmore, C. M., Gouvea, J. S., & Giere, R. (2014). Models in science and in learning science: Focusing scientific practice on sense-making. In M. R. Matthews (Hrsg.), *International handbook of research in history, philosophy and science teaching* (S. 1171–1202). Dordrecht: Springer.
- Quigley, D., Ostwald, J., & Sumner, T. (2017). *Scientific modeling: using learning analytics to examine student practices and classroom variation*. Proceedings of the Seventh International Learning Analytics & Knowledge Conference. (S. 329–338).
- Quillin, K., & Thomas, S. (2015). Drawing-to-learn: a framework for using drawings to promote model-based reasoning in biology. *CBE - Life Sciences Education*. <https://doi.org/10.1187/cbe.14-08-0128>.
- Ruppert, J., Duncan, R. G., & Chinn, C. A. (2019). Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education*, 49(3), 921–948.
- Sandmann, A. (2014). Lautes Denken - die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179–188). Berlin, Heidelberg: Springer.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, 1(2), 201–238.
- Schreier, M. (2010). Fallauswahl. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 238–251). Wiesbaden: VS.
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: SAGE.
- Sharon, A. J., & Baram-Tsabari, A. (2020). Can science literacy help individuals identify misinformation in everyday life? *Science Education*. <https://doi.org/10.1002/sce.21581>.
- Sins, P. H., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The difficult process of scientific modelling: an analysis of novices'

- reasoning during computer-based modelling. *International Journal of Science Education*, 27(14), 1695–1721.
- Sung, J. Y., & Oh, P. S. (2018). Sixth grade students' content-specific competencies and challenges in learning the seasons through modeling. *Research in Science Education*, 48(4), 839–864.
- Torres, J., & Vasconcelos, C. (2015). Nature of science and models: comparing Portuguese prospective teachers' views. *EURASIA Journal of Mathematics, Science and Technology Education*, 11(6), 1473–1494.
- Upmeyer zu Belzen, A. (2014). Black Box: Modellierung von Prozessen naturwissenschaftlicher Erkenntnisgewinnung. In D. Ludwig, C. Weber & O. Zauzig (Hrsg.), *Das materielle Modell. Objektgeschichten aus der wissenschaftlichen Praxis* (S. 99–106). Paderborn: Wilhelm Fink.
- Upmeyer zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- Vo, T., Forbes, C., Zangori, L., & Schwarz, C. V. (2019). Longitudinal investigation of primary inservice teachers' modelling the hydrological phenomena. *International Journal of Science Education*, 41(18), 2788–2807.
- Weinert, F. (2001). Vergleichende Leistungsmessungen an Schulen. In F. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim, Basel: Beltz.
- Werner, S., Förtsch, C., Boone, W., von Kotzebue, L., & Neuhaus, B. J. (2019). Investigating how German biology teachers use three-dimensional physical models in classroom instruction: a video study. *Research in Science Education*, 49(2), 437–463.
- Windschitl, M., & Thompson, J. (2006). Transcending simple forms of school science investigation: the impact of preservice instruction on teachers' understanding of model-based inquiry. *American Educational Research Journal*, 43(4), 783–835.
- Woolley, J. S., Deal, A. M., Green, J., Hathenbruck, F., Kurtz, S. A., Park, T. K. H., et al. (2018). Undergraduate students demonstrate common false scientific reasoning strategies. *Thinking skills and creativity*, 27, 101–113.
- Yenilmez Turkoglu, A., & Oztekin, C. (2016). Science teacher candidates' perceptions about roles and nature of scientific models. *Research in Science & Technological Education*, 34(2), 219–236.

Publikation 5 (Seite LXXXVI bis XCIX) wurde aus urheberrechtlichen Gründen entnommen und ist abrufbar unter [https://doi.org/10.1007/978-3-030-89480-1\\_18](https://doi.org/10.1007/978-3-030-89480-1_18).

# Investigating the dimensions of modeling competence among preservice science teachers: Meta-modeling knowledge, modeling practice, and modeling product

Maximilian Felix Göhner<sup>1</sup>  | Tom Bielik<sup>1</sup>  | Moritz Krell<sup>2</sup> 

<sup>1</sup>Faculty of Biology Education, Freie Universität Berlin, Berlin, Germany

<sup>2</sup>Department Biology Education, Leibniz Institute for Science and Mathematics Education, Kiel, Germany

## Correspondence

Maximilian Felix Göhner, Faculty of Biology Education, Freie Universität Berlin, Berlin, Germany.  
Email: [maximilian.goehner@fu-berlin.de](mailto:maximilian.goehner@fu-berlin.de)

## Abstract

Worldwide, teachers are expected to engage their students in authentic practices, like scientific modeling. Research suggests that teachers experience challenges when integrating modeling in their classroom instruction, with one explanation that teachers themselves lack the necessary modeling competence. Currently, theoretical conceptualizations structure the modeling competence into three dimensions: meta-modeling knowledge, modeling practice, and modeling products. While each of these dimensions is well researched on its own and the three dimensions are commonly expected to be highly positively related, studies investigating their specific relationships are widely lacking. Aiming to fill this gap, the present study investigated the meta-modeling knowledge, modeling practice, and modeling products of 35 secondary preservice biology teachers engaging in a black box modeling task. Data were collected with an established pen-and-paper questionnaire consisting of five constructed response items assessing meta-modeling knowledge and by videotaping the participants engaging in the black box modeling task. Herein, the three dimensions of

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Research in Science Teaching* published by Wiley Periodicals LLC on behalf of National Association for Research in Science Teaching.

modeling competence were operationalized as five variables including decontextualized and contextualized meta-modeling knowledge, complexity, and homogeneity of the modeling processes and a modeling product score. In contrast to our expectations and common assumptions in the literature, significant relationships between the five variables were widely lacking. Only the complexity of the modeling processes correlated significantly with the quality of the modeling products. To investigate this relationship further, a qualitative in-depth analysis of two cases is presented. Implications for biology teacher education are discussed.

#### KEYWORDS

preservice science teachers, modeling practice, modeling competence, meta-modeling knowledge, modeling product

## 1 | INTRODUCTION

Influential science education standard documents, such as the Next Generation Science Standards in the United States (NGSS Lead States, 2013), call upon teachers to engage their students in authentic scientific practices (Osborne, 2014). Since modeling stands at the heart of any authentic scientific endeavor (Giere, 1999), it is seen as an essential part of science teaching and learning (Gilbert & Justi, 2016; Upmeier zu Belzen, Krüger, et al., 2019). Consequently, scientific modeling is proposed to stand in the center of science curricula, incorporating other scientific practices as well (Lehrer & Schauble, 2006; Meister et al., 2021; Passmore et al., 2014; Windschitl et al., 2008). However, this focus on scientific practices demands preservice science teachers to develop the necessary prerequisites (“professional competences”; Baumert & Kunter, 2013) needed to be able to plan lessons, teach, and reflect upon the teaching-learning processes professionally. Hence, next to general pedagogical knowledge and competences, preservice science teachers need to develop competences related to scientific practices—such as modeling—as part of their professional competences (Osborne, 2014); with modeling being one of such competences.

Empirical research has shown that teachers’ epistemic ideas about models and modeling (i.e., meta-modeling knowledge) impacts their classroom instruction (Harlow et al., 2013; Vo et al., 2015). However, meta-modeling knowledge is only one dimension in the current theoretical conceptualization of modeling competence, which also encompasses the teachers’ abilities to engage in the modeling practices (Nicolaou & Constantinou, 2014) and their abilities to develop a high-quality modeling product (Chiu & Lin, 2019). Although, each of those dimensions of modeling competence is addressed by empirical research, the relationships between preservice science teachers’ meta-modeling knowledge, their modeling practices, and modeling product are to date rarely studied (Cheng et al., 2021; Chiu & Lin, 2019; Louca & Zacharia, 2012; Nielsen & Nielsen, 2021a) and was consequently emphasized as “[o]ne of the most pressing needs for future research” (Louca & Zacharia, 2012, p. 486).

## 1.1 | Models and modeling in science education

Models are central tools for communicating and reasoning in science and essential to scientists for explaining phenomena and for predicting possible outcomes (Giere et al., 2006; Godfrey-Smith, 2006; Harré, 1970). Consequently, modeling competence is emphasized in standards and curricula in many countries (ACARA, 2015; BCMOE, 2019; KMK, 2020; NGSS Lead States, 2013; NRC, 2012; VCAA, 2016). Based on theoretical approaches in the philosophy of science, scientific models can be defined as epistemic tools for sense-making (Knuuttila, 2011). Accordingly, scientific modeling is the iterative and cyclic process of developing and using models in science, aiming at investigating, representing, explaining, and predicting phenomena (Giere et al., 2006; Passmore et al., 2014).

## 1.2 | Science teachers' competences related to models and modeling in science

One goal of teacher education is to equip future teachers with the prerequisites needed to plan lessons, teach, and reflect upon the teaching-learning processes professionally (Baumert & Kunter, 2013; Carlson & Daehler, 2019). Science teachers need to have meta-modeling knowledge as well as the abilities to engage in modeling practices and to develop high-quality modeling products (Chiu & Lin, 2019; Nicolaou & Constantinou, 2014). Furthermore, science teachers need related pedagogical content knowledge (PCK), including knowledge about teaching with and about models and about how to conduct modeling activities in science classrooms (Justi & Van Driel, 2006).

A multitude of research has been conducted on investigating pre- and in-service science teachers' professional competences related to models and modeling, including, but not limited to, teachers' instructional practice regarding models and modeling in classrooms, as well as their own meta-modeling knowledge and modeling practice (Khan, 2011; Krell & Krüger, 2016; Oh & Oh, 2011; Shi et al., 2021; Torres & Vasconcelos, 2015; Vo et al., 2015, 2019).

Most studies indicate that classroom practice typically gives students few opportunities to meaningfully engage with models (Campbell et al., 2015; Khan, 2011). Studies on the instructional practice of teachers in science classrooms suggest that teachers encounter a multitude of challenges and mainly focus on knowledge aspects represented in models, disregarding the predictive nature of models (Harlow et al., 2013; Nielsen & Nielsen, 2021b; Shi et al., 2021; Vo et al., 2015, 2019). However, it is commonly assumed, that teachers' own ideas about models and modeling shape their instructional classroom practice (Harlow et al., 2013; Vo et al., 2015, 2019). For example, Vo et al. (2015) observed that those teachers, who more strongly emphasized otherwise uncommon epistemic ideas (*evidence*, *mechanism*, and *audience*) and modeling practices (*evaluate* and *revise*), employed better instructions during classroom practice. However a later longitudinal investigation by the same authors, suggests this transfer is likely delayed and takes place over vast timescales, often leading to changes in instructional practice only after a year (Vo et al., 2019).

Keeping in mind that teachers' own epistemic ideas can impact their classroom instruction, it is alarming that most studies focusing on pre- and in-service science teachers' meta-modeling knowledge unravel rather uninformed views of teachers on models and modeling in science, including naïve realist views of models as simple copies of reality (e.g., Krell & Krüger, 2016; Torres & Vasconcelos, 2015). Moreover, only a fraction of science teachers expresses that

models, as well as the underlying ideas represented in models, may be tested by deducing predictions (Krell & Krüger, 2016).

Studies investigating teachers' modeling practice are rare, but the existing studies come to similar results. It is, for example, consistently observed that some aspects of modeling practice are challenging for teachers, especially the evaluation of models (Khan, 2011; Vo et al., 2015), and, again, the predictive use of models (Göhner & Krell, 2020a). Göhner and Krell (2020a) investigated the modeling practice of 32 secondary preservice biology teachers investigating a black box. They observed that only 14 of these 32 secondary preservice biology teachers used their developed models to predict the investigated system's behavior. Moreover, even among these 14 secondary preservice biology teachers, the developed models were rarely evaluated based on predictions in a repeated and systematic manner.

In summary, studies addressing teachers' instructional practice regarding modeling their epistemic ideas about models and modeling, and their modeling practice, suggest that the potential of models as epistemic tools in teaching scientific practices is left untapped (Harlow et al., 2013; Nielsen & Nielsen, 2021b). The existing body of empirical research, theoretical considerations, and curricula highlight the epistemic nature of models as tools for scientific reasoning. However, in classroom settings models are still predominantly used by teachers for the purpose of communicating scientific content, who are likely not fully aware of the potential, models and modeling may have. While teachers' epistemic ideas of models and their relation to classroom instruction are quite well researched, studies connecting teachers' modeling practice with their classroom instruction and connecting teachers' epistemic ideas of models with their modeling practice are widely lacking. Our study aims to fill this gap in science education research. In the next section, our theoretical conceptualization of what constitutes modeling competence will be briefly explained.

### 1.3 | Modeling competence in science education

In general, modeling competence is seen as one necessary element of teachers' professional competences (Günther et al., 2019; Osborne, 2014). Modeling competence is defined as a system of the knowledge, skills, and abilities necessary to engage in the process of developing and using models for reasoning in science, including motivational dispositions to apply these capabilities for problem-solving in specific situations (Upmeier zu Belzen, van Driel, et al., 2019). When achieving modeling competence, a person should better understand scientific concepts, develop an appreciation of the nature of science, and advance in their mastery of the scientific process (Gilbert & Justi, 2016).

As suggested in recent literature (Chiu & Lin, 2019; Nicolaou & Constantinou, 2014; Nielsen & Nielsen, 2021a), modeling competence can be divided into three dimensions: *meta-modeling knowledge*, *modeling practice*, and the *modeling product*. In the following, we will expand on each of these dimensions, highlighting the theoretical scope of this article.

#### 1.3.1 | Meta-modeling knowledge

Meta-modeling knowledge is a term commonly used to describe epistemic ideas about models and modeling. Schwarz et al. (2009) defined meta-modeling knowledge as knowledge about "how models are used, why they are used, and what their strengths and limitations are"

TABLE 1 Theoretical framework for meta-modeling knowledge (Upmeier zu Belzen, van Driel, et al., 2019)

Aspect	Level I	Level II	Level III
Nature of models	Replication of the phenomenon	Idealized representation of the phenomenon	Theoretical reconstruction of the phenomenon
Multiple models	Different model objects	Different foci on the phenomenon	Different hypotheses about the phenomenon
Purpose of models	Describing the phenomenon	Explaining the phenomenon	Predicting something about the phenomenon
Testing models	Testing the model object	Compare the model and the phenomenon	Testing hypotheses about the phenomenon
Changing models	Correcting defects in the model object	Revising due to new insights	Revising due to the falsification of hypotheses about the phenomenon

Note: This framework uses the term *model object* referring to the work of Mahr (2011). In the present article, the term modeling product is used instead, because this term is more established in science education literature (e.g., Chiu & Lin, 2019).

(pp. 634–635). There are different approaches to conceptualize meta-modeling knowledge, each of them defining related aspects (Schwarz et al., 2009) or dimensions (Crawford & Cullin, 2005) as part of meta-modeling knowledge. One of the more common frameworks for meta-modeling knowledge, which will be used throughout this study, was proposed by Upmeier zu Belzen, van Driel, et al. (2019). They propose five aspects of meta-modeling knowledge: *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*. For each, three levels of understanding are distinguished (Table 1). Level I is related to naïve views, understanding models as direct copies of reality and focusing on features of the modeling product itself, rather than on the representational and predictive function of models. Level II is related to more advanced views, understanding models mainly as idealized representations or media to visualize and explain something, while level III adds the appreciation of the predictive power of models as research tools (Upmeier zu Belzen, van Driel, et al., 2019).

Ke and Schwarz (2020, p. 5) distinguish between meta-modeling knowledge independent of the specific learning context (i.e., decontextualized meta-modeling knowledge) and “epistemological knowledge about models and modeling in action” (i.e., contextualized meta-modeling knowledge). The context-dependency of meta-modeling knowledge has been suggested to be an issue of critical importance for assessing and teaching meta-modeling knowledge, which, therefore, should be further investigated in science education (Krell et al., 2014; Sikorski, 2019).

### 1.3.2 | Modeling practice

Science education literature addresses modeling practice in several theoretical frameworks that use overlapping terms such as modeling activities (Fretz et al., 2002; Göhner & Krell, 2020a), model-based learning practices (Louca & Zacharia, 2012), or modeling phases (Constantinou, 1999). In this article, we will refer to *modeling practice* as an umbrella term encompassing any modeling behavior or cognitive operation while being engaged in modeling. To the specific modeling practices carried out by the participants in this study, however, we will refer to as *modeling processes*, which are

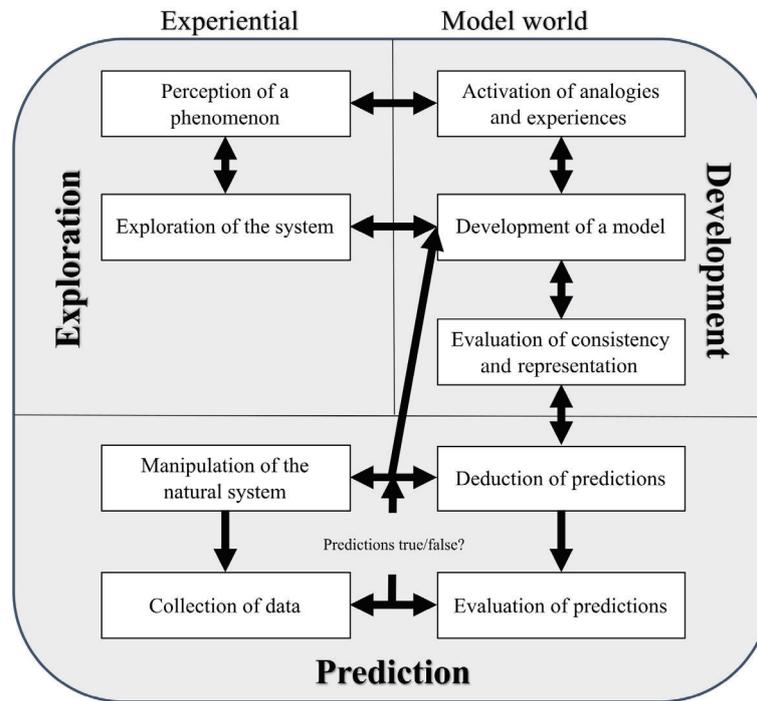


FIGURE 1 Model of modeling (adapted from Krell et al., 2019)

specific sequences of observable and distinguishable operations, in turn, referred to as *modeling activities* (Göhner & Krell, 2020a).

These modeling activities are incorporated in the model of modeling (Figure 1), which is the theoretical framework used for the description of modeling processes in this article (see Krell et al. (2019) for a detailed description).

The model of modeling distinguishes between the experiential world and the modeling world (Göhner & Krell, 2020a; Krell et al., 2019) and aligns with the model of scientific activity (NRC, 2012). Collecting data, making observations, and conducting experiments are part of the real world “investigating” sphere (i.e., experiential world), while models are part of the theoretical “developing explanations and solutions” sphere (i.e., modeling world). Additionally, the role of models as epistemic tools (Knuuttila, 2011) is emphasized in the model of modeling, when new hypotheses are deduced or predictions are drawn from the model, which are then also empirically tested (Dounas-Frazer & Lewandowski, 2018). Practically, the model of modeling can be used to operationalize the assessment and description of modeling processes.

### 1.3.3 | Modeling product

The main outcome of any modeling process is the development of a tangible, visible, and communicable artifact (i.e., modeling product) that demonstrates the modeler’s understanding and that can be evaluated by specific criteria for its quality (e.g., epistemic criteria; Pluta et al., 2011). In learning contexts, modeling products are understood to be of high quality, if they externalize and express learners’ thoughts and help them visualize and examine

components of their theories (Jonassen & Ionas, 2008). Commonly in science education research, modeling products are evaluated in a more content-related approach based on the integration of specific components or relationships between these components (Chang et al., 2020). Chiu and Lin (2019) identified “a lack of deep discussions on the topic of modeling products” (p. 2). Most available studies analyze students’ modeling products as indicators for evaluating students’ modeling practices and meta-modeling knowledge (Bamberger & Davis, 2013; Cheng & Lin, 2015; Ergazaki et al., 2007; Schwarz et al., 2009).

Relevant to this study, modeling products can be evaluated for their quality, that is including the components and relationships that are required to accurately explain and predict the phenomenon, and evaluated for their complexity, that is the number of components and relationships in the model (Göhner & Krell, 2020a).

### 1.3.4 | Connecting meta-modeling knowledge, modeling practice, and modeling product

Many researchers in science education propose modeling practices and meta-modeling knowledge as the two broad constituent dimensions of modeling competence (e.g., Nicolaou & Constantinou, 2014; Upmeier zu Belzen, van Driel, et al., 2019). Chiu and Lin (2019) propose to add the modeling product as a third dimension as competent modelers are assumed to develop the ability to construct high quality models. Some researchers claim, that meta-modeling knowledge shapes or guides the practice of modeling (Lee & Kim, 2014; Nicolaou & Constantinou, 2014; Schwarz et al., 2009). Others conceptualize it the other way around, proposing that engagement in the modeling practice might contribute to the development of meta-modeling knowledge (Gobert & Pallant, 2004; Schwarz & White, 2005). Empirical findings propose that engagement in the modeling practices alone is not sufficient to foster meta-modeling knowledge but that associated reflections on the practices are necessary to reach this goal (Schwarz & White, 2005). This was also found in studies focusing on the broader construct of scientific meta-knowledge (i.e., nature of science; Abd-El-Khalick & Lederman, 2000). However, only a few studies had put the suggested relationships between the dimensions of modeling competence to the test (Chiu & Lin, 2019; Sins et al., 2009).

Summarizing, meta-modeling knowledge, and modeling products have been studied by many researchers and defined in a variety of approaches. However, process-oriented studies of students’ or teachers’ engagement in modeling, opposed to meta-modeling knowledge and modeling products, are still widely lacking in science education (Louca & Zacharia, 2012; Nicolaou & Constantinou, 2014). Therefore, we set to identify and characterize meta-modeling knowledge, modeling practices, and quality of modeling products in a sample of secondary preservice biology teachers and investigate the relationships between these three dimensions of the modeling competence. This will provide valuable theoretical knowledge, which is also important for understanding how to promote these dimensions in science teacher education (Nicolaou & Constantinou, 2014).

## 1.4 | Aims and research question

The following research question will be addressed in this study: What are the relationships between secondary preservice biology teachers’ meta-modeling knowledge (contextualized and

decontextualized), their modeling practices, and the quality of their modeling products while engaging in a modeling task? Following common assumptions in the literature, we expect all dimensions of modeling competence to be positively related.

## 2 | METHODS

### 2.1 | Context of the study

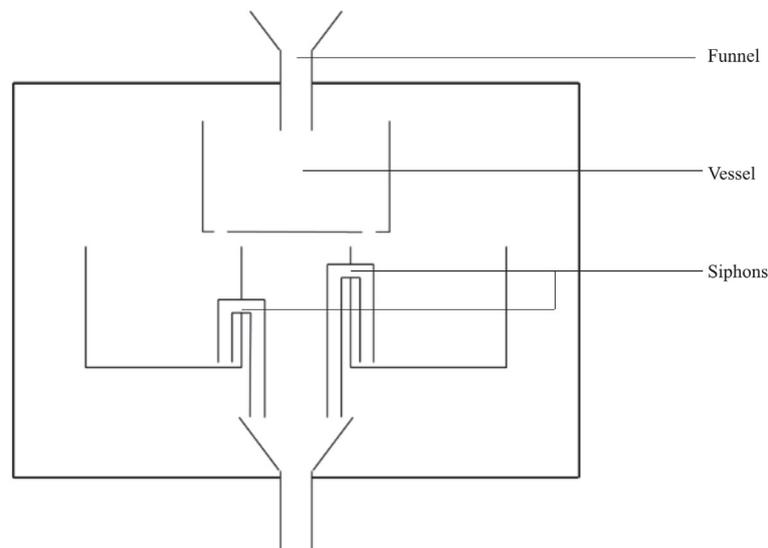
This study is situated in the first phase of secondary preservice biology teacher education in Germany. Secondary preservice biology teachers in Germany usually study two subjects (with one of them being biology) in a six-semester bachelor's program, followed by a four-semester master's program (concurrent teacher education programs). At the end of their studies, secondary preservice biology teachers are requested to having developed basic professional knowledge and competences needed for their profession (Neumann et al., 2017). These include knowledge and competences regarding inquiry and reasoning in science (KMK, 2019). A significant positive development of German secondary preservice science teachers' scientific reasoning competences over their course of studies has been described in empirical studies (Krüger et al., 2020).

### 2.2 | Sample

The sample population consisted of secondary preservice biology teachers, enrolled in the bachelor or master teacher education program in one of two involved German universities. To increase the likelihood of observing a variety of different modeling processes, a theoretical sampling strategy—heterogeneous sampling (Patton, 1990)—was used in the present study. Previous studies suggest that the quality of scientific modeling is positively related to scientific reasoning competences (e.g., Cheng & Lin, 2015) and general cognitive abilities (e.g., Nehring et al., 2015). Therefore, these two variables were used as screening variables, employing established pen-and-paper instruments (Krüger et al., 2020; Liepmann et al., 2007). Fifty-seven secondary preservice biology teachers, who had extreme scores (one half standard deviation higher or lower than the mean scores of the respective norm sample) in both assessments, were invited to participate in the study. Thirty-five secondary preservice biology teachers agreed to participate, aged between 17 and 39, with an average age of 24 years. Twenty-two participants were enrolled in the bachelor's program when participating in the study and 13 were enrolled in the master's program. Eight of the secondary preservice biology teachers were additionally enrolled in another scientific subject (including chemistry, physics, food science, agricultural science, and computer science). The study was not mandatory for any university courses or obligatory parts of the curriculum; participation was voluntary. Researchers and participants had no formal relationships to one another.

### 2.3 | Black box modeling task

A black box modeling task was applied in this study. The black box approach is established in science education research to study processes of scientific thinking and modeling (Lederman & Abd-El-Khalick, 2002; Passmore & Svoboda, 2012). In this study, a water black box was used



**FIGURE 2** Model of the inner mechanism of the water black box used. Water is funneled into a first vessel and then directed equally into two further vessels. If a specific volume of water is reached in each of these vessels, the water flows out the black box through a siphon. As the siphons are installed on different heights, the specific volume of water for each vessel differs (Krell et al., 2019)

(Figure 2). It can be explored by filling the black box with water (input), which then results in measurable outputs of water (see Krell et al., 2019 for a detailed description of the black box). It has been shown that black box approaches are suitable to elicit modeling processes, in which models are used as epistemic tools to investigate the black box (Krell et al., 2019; Passmore & Svoboda, 2012), including all steps considered in the model of modeling (Figure 1).

Giving an example, repeating an input of 400 ml six times, produces a pattern of 0, 400, 600, 400, 0, 1000 ml outputs. Typically, the third output cannot readily be explained by the participants. In the model of modeling this resembles the perception of a phenomenon, which is a starting point for modeling processes as described above, including the exploration of the system (i.e., the black box), the activation of analogies and experiences (of what might be inside the black box), the development of a (drawn) model and its evaluation regarding consistency and representation. Using imagistic simulation, hypotheses (e.g., about the next output) can be deduced from the drawn model and the model can be tested by making another input, leading to the model's confirmation or rejection. In the latter case, the drawn model should be modified.

To gain additional insights into the participants' reasoning processes, they were asked to concurrently think aloud (Leighton & Gierl, 2007). This was practiced with three short exercises as part of the introduction to the study, in which the participants also answered the questionnaire on decontextualized meta-modeling knowledge (see below), were informed that participation was voluntary, and signed an informed consent form. After these preparations, the participants were brought into a room equipped with three video cameras, the black box, some prefilled beakers of water, a bucket as water reservoir, and a chalkboard. The first author briefly explained the basic functionality of the black box using a prepared script and provided the following task: "Draw a model of the inside of the black box." Participants were informed, that

there are no time constraints. The first author stayed in the room to prevent any technical errors and, if necessary, to remind the participant to concurrently think aloud; otherwise, he did not intervene.

## 2.4 | Data collection and analysis

All qualitative data analyses were done within the methodological frame of qualitative content analysis, that is including category systems guiding analysis, coding by different persons, finding consensus coding after discussion, and calculating Cohen's Kappa ( $\kappa$ ) as a measure of intrarater- and interrater-agreement (Schreier, 2012). In addition, the following procedures of data collection and analysis have been conducted in this study.

### 2.4.1 | Decontextualized meta-modeling knowledge

The participants' meta-modeling knowledge was assessed based on the framework described above (Upmeier zu Belzen, van Driel, et al., 2019), using an established pen-and-paper questionnaire which consists of five constructed response items (Krell & Krüger, 2016). These items are related to the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models* (Table 2). The questionnaire was given to each participant immediately before engaging in the black box modeling task. In the questions, the respondents were asked to provide their understanding of the five aspects related to the scientific discipline of biology; however, as no more specific context was provided in the tasks and the questions have not been answered during any kind of modeling activity, we will refer to what has been assessed in the questionnaire as *decontextualized meta-modeling knowledge* (Ke & Schwarz, 2020). In total, 166 out of the possible 175 responses were analyzed from the 35 participants, as some participants did not respond to all five questions or responses could not be assigned to any level. All responses to the questionnaire were qualitatively analyzed using an already established category system, by which a specific level can be identified (ranging from I to III; Table 1) in the responses for each of the five aspects (Krell & Krüger, 2016). Statements including multiple response levels were always given the highest identified level. Each statement was coded twice by the first author and additionally by a trained student assistant, achieving substantial intrarater-agreements ( $\kappa = 0.68$ ), and substantial interrater-agreements ( $\kappa = 0.71$ ).

### 2.4.2 | Contextualized meta-modeling knowledge

The transcripts of the participants' modeling processes were analyzed by a trained student assistant based on the same category system that has been used for the analysis of the questionnaire (Table 2). Statements related to the five aspects of meta-modeling knowledge, verbalized by the participants throughout their modeling process, were identified and coded accordingly. Hence, this procedure aimed to assess the participants' *contextualized meta-modeling knowledge* or their meta-modeling knowledge "in action," respectively (Ke & Schwarz, 2020). In line with the evaluation of the questionnaire, participants expressing statements of varying levels throughout their modeling process were given the highest level observed for each aspect. To secure the quality of the analysis, 20% of the material (7 of the 35 transcripts) were randomly selected and

**TABLE 2** The constructed response items for the five aspects of meta-modeling knowledge, with excerpts of response examples provided for each level and aspect

Item	Level I	Level II	Level III
<b>Nature of models</b>			
Explain the extent to which biological models are equivalent to their biological phenomenon.	A model should be as close to the original as possible and represent it well. (Daphne)	A model resembles the original strongly to little. Depending on the purpose of the model, it has structural or functional similarities. (James)	Models are just theoretical constructs created by people. (Selina)
<b>Alternative models</b>			
Why are there different models of one biological phenomenon?	As soon as a model is equivalent to the original, multiple perspectives are possible in model development and representation. (Kara)	As some models represent only a specific function or process, there are multiple possibilities to resemble an original object. (Celine)	There can be multiple models for one original object if different theories describing that model exist or if there are various hypotheses, which have neither been falsified nor verified yet. (Cynthia)
<b>Purpose of models</b>			
What is the purpose of models in biology?	Models in biology serve various purposes; in my opinion, however, they have a superficial purpose: to allow observations of the initial object that would not be possible under normal circumstances. (Claudia)	[Models] explain the structures of the source object. (May)	[The purpose of models is the] simulation of biological processes to gain knowledge or to solve problems. (Jonathan)
<b>Testing models</b>			
How do we test whether a biological model serves its purpose?	By testing the model, that is, using it in biology classes and testing it. (Frida)	By comparing with the original and see if the model shows/explains everything. (Achilles)	The model can be used to test hypotheses/make predictions that can be evaluated. (Ryan)

TABLE 2 (Continued)

Item	Level I	Level II	Level III
Changing models			
What are some reasons for changing a biological model?	Depending on its use, that is, if the model is understood in class or not, it is important to further adapt the model and improve it. [...] It must be a refinement or correction. (Frida)	There are new data that indicate errors in the previous model. (Carlo)	Predictions made by the model are not correct. (Floyd)

*Note:* Names in brackets are pseudonyms. Note that the questionnaire has been developed and administered in the German language. The full German questionnaire including the category system is available upon request from the third author.

coded twice over an interval of two weeks by the trained student assistant and once by the first author. Cohen's Kappa indicated substantial intrarater- ( $\kappa = 0.69$ ) and substantial interrater-agreement ( $\kappa = 0.72$ ).

### 2.4.3 | Modeling practice

For analyzing modeling practices, the behavior and verbalizations of the participants engaging in the black box modeling task were videotaped. The verbalizations were then transcribed verbatim, including selected behavioral aspects (e.g., making an input or observing an output). These transcripts were initially analyzed qualitatively to identify single modeling activities. For this, an established category system was used in the analysis. It consists of 9 main categories and 19 sub-categories representing the modeling activities (Table 3). The nine main categories of the category system stem from the model of modeling described earlier in this study (Figure 1; for a detailed description see Krell et al., 2019). If necessary, the videos were considered as additional data sources in the analysis. Each transcript was coded twice over an interval of two weeks by the first author and additionally by a trained student assistant. Cohen's Kappa indicated almost perfect intrarater-agreement ( $\kappa = 0.83$ ) and substantial interrater-agreement ( $\kappa = 0.77$ ).

From the resulting sequences of modeling activities, state transition graphs were built to visualize each participant's modeling process (Andrienko & Andrienko, 2018). Each participant's state

TABLE 3 Activities of the modeling process (Krell et al., 2019)

Phase	Category	Sub-category (activity)
Exploration	1. Perception of a phenomenon	
	2. Exploration of the system	2.1. Input/output (exploratory)
		2.2. Summarizing/describing observations
		2.3. Input/output (pattern detection)
		2.4. Recognizing patterns
	3. Activation of analogies and experiences	
Development	4. Development of model	4.1. Graphically develop model
		4.2. Change model to optimize consistency
		4.3. Change model to optimize representation
		4.4. Reject model due to poor consistency/representation
	5. Evaluation of consistency and representation	5.1. Evaluate consistency
		5.2. Evaluate representation
6. Finding of consistency and representation		
Prediction	7. Deduction of predictions	
	8. Evaluation of predictions	8.1. Input/output (to test predictions)
		8.2. Confirmation of prediction
		8.3. Falsification of prediction
	9. Modification/rejection of model	9.1. Change model due to falsified predictions
		9.2. Reject model due to falsified predictions

transition graph consists of knots for each state (i.e., modeling activity) and edges depicting the transitions between two modeling activities. The edges are assigned the number of transitions observed between each modeling activity as weight.

All modeling processes were qualitatively evaluated based on the modeling activities shown during the process and their sequential order visualized as state transition graphs. Additionally, the state transition graphs were used to quantify the modeling practices for further analysis. Two variables, the complexity and homogeneity of the secondary preservice biology teachers' modeling processes, were estimated as described in Göhner and Krell (2020a, 2020b). This is shortly described in the following.

Modeling processes are understood to be more complex, if they include various modeling activities (Göhner & Krell, 2020a). Complexity was therefore estimated using the graph metric known as “communities,” which involves counting all subgraphs within each state transition graph (Porter et al., 2009). Participants may show a limited range of activities, resulting in state transition graphs with more subgraphs, as some activities are not addressed and thus, not connected to others. The basic communities score was further normalized by subtracting the communities score of each participant from the maximum communities score achieved in this study.

Modeling processes are understood to be more homogenous, if the observed modeling activities and the transitions between them are more equally distributed, leading to the state transition graph being more independent from a single knot (Göhner & Krell, 2020a). To estimate homogeneity, the centrality of every state transition graph was determined. The centrality score was reciprocally transformed, as centrality describes the dependence of a graph on a single knot (Newman, 2010). To account for structural outliers, homogeneity was estimated using the sum of the three measures of centrality: closeness centrality, degree centrality, and betweenness centrality (Ronqui & Travieso, 2015).

#### 2.4.4 | Modeling product

To assess the quality of the participants' modeling products, a novel category system was inductively developed, based on the previous work of the authors. As suggested by Bielik et al. (2018), modeling products were considered of high quality when they included the components and relationships that are required to accurately explain the black box phenomenon (Figure 2). Here, modeling products of high quality were expected to include three concepts, which were found to be necessary to explain the water distribution inside the black box and the input and output patterns (Krell et al., 2019): (1) The modeling product should contain water reservoirs filling up with water, (2) the water reservoirs should be embedded in a parallel system of diverging paths, and (3) water should be fully emptied at a specific fill level. The specific realization of these three concepts can vary, incorporating alternative ideas; water reservoirs, for example, could be realized through vessels, cisterns, or even sponges. As some modeling products could not be evaluated based on their drawn appearance alone, the participants' verbal statements during the development of their modeling products were considered as an additional data source.

The participants' final modeling products were photographed, digitally reconstructed, and coded. However, in two cases, participants presented multiple final modeling products as possible solution for the given black box task. To code the final modeling products, it was noted if each concept was integrated, by being either drawn or described verbally. Participants' scores range from zero to three, integrating neither of these concepts (=0), one (=1), two (=2), or all of these concepts (=3) into their drawn modeling product or verbal descriptions thereof. Further attributes

of the modeling products like their aesthetic appearance or fit to the obtained data have not been evaluated. For participants with multiple final modeling products, only the highest scoring modeling product was considered in the statistical analysis. Every modeling product ( $N = 42$ ) was coded twice over an interval of two weeks by the first and once by the second author. Cohen's Kappa indicated almost perfect intrarater- ( $\kappa = 0.84$ ) and almost perfect interrater-agreement ( $\kappa = 0.82$ ).

### 2.4.5 | Relationships between meta-modeling knowledge, modeling process, and modeling product

Spearman's rank correlations were calculated to analyze the relationship between decontextualized and contextualized meta-modeling knowledge for each aspect, complexity, and homogeneity of the modeling process, and the modeling product score. To obtain a measure of decontextualized and contextualized meta-modeling knowledge, the participants' mean score across the five aspects were calculated. Furthermore, due to the rather small sample size, nonparametric tests were carried out. Effect size measures have been calculated based on Lenhard and Lenhard (2016).

To provide in-depth insights into the potential nature and direction of the relationships between secondary preservice biology teachers' meta-modeling knowledge, modeling practices, and modeling products, sample cases will be qualitatively described. Hereby, the cases have been selected based on the statistical relationships found in the quantitative analysis. The selected cases were analyzed by reconstructing their individual modeling processes from the transcripts, which are illustrated as codelines, showing the sequential order of the individual modeling activities.

## 3 | RESULTS

Aggregated scores for meta-modeling knowledge, modeling processes, and modeling products will be provided in the following section. The full table with the individual scores can be found in Table S1.

### 3.1 | Decontextualized meta-modeling knowledge

Table 4 shows the distribution of response levels regarding the five aspects of meta-modeling knowledge assessed with the questionnaire. Across the five aspects, the mean response levels range from 1.97 for the aspect *nature of models* to 2.31 for the aspect *multiple models*. The mean

**TABLE 4** Distribution of response levels regarding the five aspects of decontextualized meta-modeling knowledge

Aspect	<i>N</i>	<i>n</i> responses level I	<i>n</i> responses level II	<i>n</i> responses level III	Mean response level ( $\pm SD$ )
Nature of models	34	4	27	3	1.97 ( $\pm 0.45$ )
Multiple models	35	1	22	12	2.31 ( $\pm 0.52$ )
Purpose of models	33	7	13	13	2.18 ( $\pm 0.76$ )
Testing models	30	2	18	10	2.27 ( $\pm 0.57$ )
Changing models	34	3	24	7	2.12 ( $\pm 0.53$ )

**TABLE 5** Distribution of response levels regarding the five aspects of contextualized meta-modeling knowledge identified within the modeling processes

Aspect	N	n responses level I	n responses level II	n responses level III	Mean response level ( $\pm$ SD)
Nature of models	21	4	2	15	2.43 ( $\pm$ 0.79)
Multiple models	7	3	0	4	2.14 ( $\pm$ 1.07)
Purpose of models	16	2	10	4	2.13 ( $\pm$ 0.62)
Testing models	15	0	5	10	2.67 ( $\pm$ 0.48)
Changing models	11	2	4	5	2.27 ( $\pm$ 0.79)

level assigned to the 166 responses from all participants is 2.17, suggesting, that the participants understand models mainly as idealized representations or media to visualize and explain something (Upmeier zu Belzen, van Driel, et al., 2019).

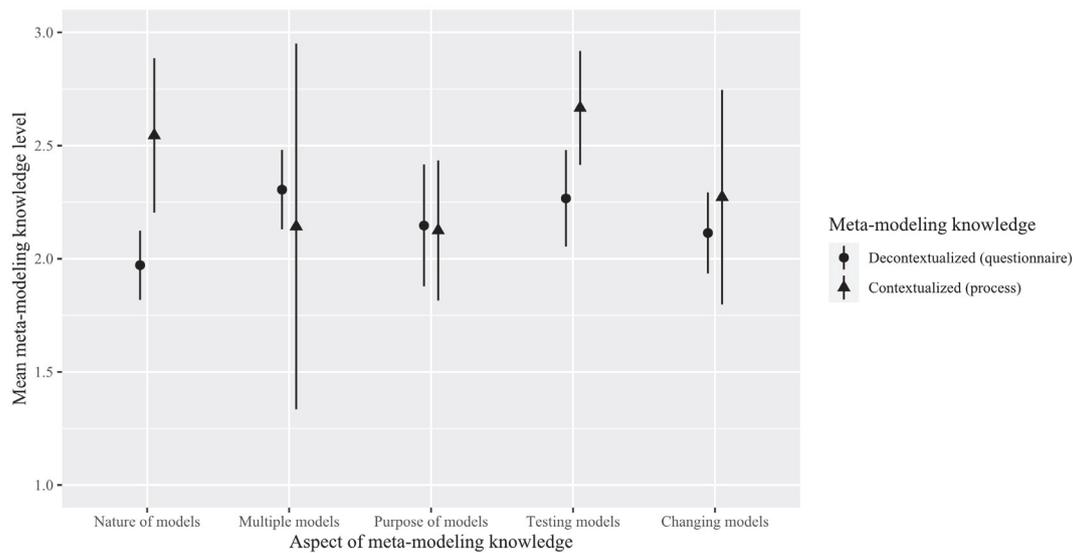
### 3.2 | Contextualized meta-modeling knowledge

One-hundred twenty-five statements related to meta-modeling knowledge were identified in the modeling processes of the 35 participants, which could be assigned to response levels of the category system (Table 5). It should be noticed that the number of statements related to meta-modeling knowledge in the modeling processes varies between the participants and the aspects mentioned, depending on the length, and complexity of the modeling processes. For some of the participants, up to sixteen statements related to meta-modeling knowledge were identified, whereas 19 participants rarely verbalized anything related to meta-modeling knowledge, leading to the identification of less than 3 statements for each of these participants. Six participants made no statements related to meta-modeling knowledge during their modeling processes at all. Also, the number of statements differs between the aspects of meta-modeling knowledge (Table S1).

Across the five aspects of meta-modeling knowledge, the mean response levels ranged from 2.13 for the aspect *purpose of models* to 2.67 for the aspect *testing models* (Table 5). The mean level of contextualized meta-modeling knowledge was 2.39, indicating a slightly higher understanding compared to the decontextualized meta-modeling knowledge. However, this still suggests that most participants understand models as idealized representations or media for visualization and explanation.

Figure 3 compares the mean response levels of decontextualized and contextualized meta-modeling knowledge for each aspect. A Wilcoxon signed-rank test revealed a significant difference between the participants' level of decontextualized and contextualized meta-modeling knowledge for the aspect *nature of models*, with decontextualized meta-modeling knowledge being significantly lower ( $z = -2.84$ ,  $p = 0.005$ ;  $d = 1.09$ , large effect size measure). For the other aspects, no significant differences were found (i.e.,  $p > 0.05$ ).

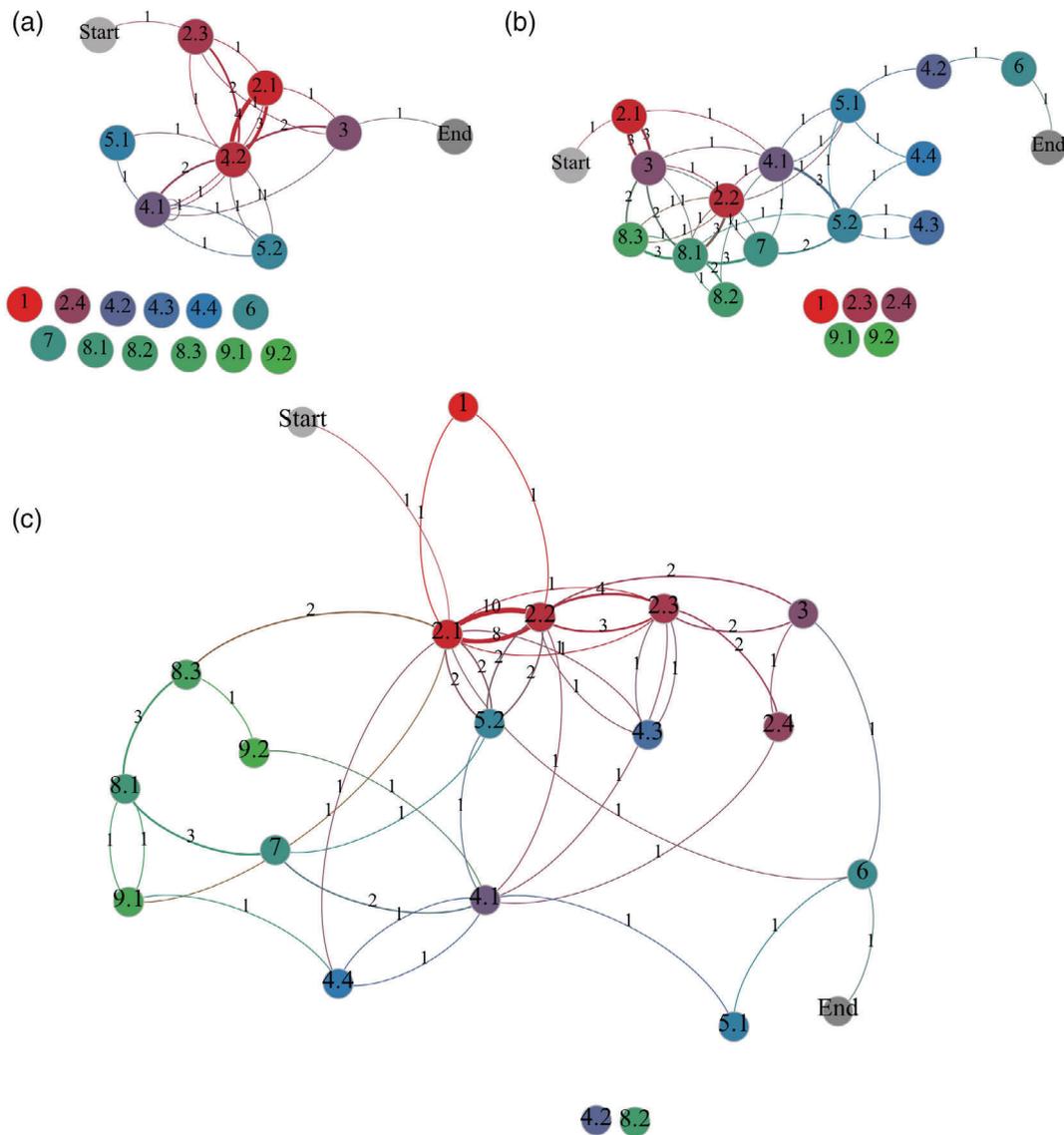
In summary, the meta-modeling knowledge of the participants, independent of being decontextualized or contextualized, indicates an understanding of models as idealized representations or media for visualization and explanation as the predominant perspective on models in our sample.



**FIGURE 3** Comparison of mean meta-modeling knowledge levels across the five aspects of meta-modeling knowledge and with regard to the context of assessment, either decontextualized within the questionnaire or contextualized as identified within the modeling processes. Points and triangles indicate mean values and lines indicate the range of the mean value  $\pm$  two times standard error

### 3.3 | Modeling practice

As described above, the specific modeling practices carried out by the participants in this study are referred to as modeling processes, which consist of different modeling activities. The length of the participants' modeling processes varied between eight minutes to almost 2 h (mean length: 1 h 9 min). In these modeling processes, the participants conducted between six and eighteen (mean: 12) different modeling activities. The qualitative analysis revealed, that the modeling processes of 20 participants included modeling activities of exploration and model development. The modeling processes of 14 participants additionally included activities of prediction. One participant's modeling process showed only modeling activities of exploration (i.e., no model development). By transforming these modeling processes into state transition graphs, the complexity (mean: 4.94) and homogeneity (mean: 1.50) of the modeling processes were quantified. Figure 4 illustrates three examples of state transition graphs. The state transition graph of Claudia's modeling process (Figure 4a) is characterized by rather low complexity, as she included only a limited range of modeling activities into her modeling process. In comparison to all transitions, the transitions between modeling activities 2.1 (input/output, exploratory) and 2.2 (summarizing/describing observations) occur more often, described by a medium homogeneity score. The state transition graph of James' modeling process (Figure 4b) is characterized by medium complexity, as he includes a higher number of different modeling activities in his modeling process. The transitions between his modeling activities are rather equally distributed, reducing the dependence of the state transition graph from specific knots (i.e., modeling activities), leading to a high homogeneity score. Finally, the state transition graph of Raphael's modeling process (Figure 4c) is characterized by high complexity, as he includes every modeling activity but one in his modeling process. However, his homogeneity score is medium, as he transitions between the modeling activities 2.2



**FIGURE 4** Three examples of state transition graphs for Claudia (a), James (b), and Raphael (c). The modeling activities (Table 3) are displayed as knots connected by their transitions, labeled with their occurrence

(summarizing/describing observations) and 2.3 (input/output, pattern detection) quite often in comparison to the other modeling activities. A complete state transition graph combining the modeling processes of all 35 participants can be found in the supplementary material (Figure S2).

A Mann-Whitney  $U$  test confirmed that participants with modeling processes, which included activities of prediction, reached significantly higher complexity scores ( $M = 7.33$ ,  $SD = 1.72$ ) than participants who did not include modeling activities of prediction ( $M = 3.15$ ,  $SD = 1.73$ ;  $z = 4.70$ ,  $p < 0.001$ ;  $d = 2.62$ , large effect size measure). Moreover, a correlation analysis (Spearman) between the complexity of the modeling processes and the length of the participants' modeling processes revealed a significant correlation ( $r = 0.39$ ,  $p = 0.021$ ; medium effect size measure). Other sample characteristics, like the subject combination of each participant or

the enrollment status of the participants in either the bachelors' or masters' program, could not be shown to be systemically related to the modeling processes.

Summarizing, related to the modeling practices, the majority of our participants only engaged in modeling activities of exploration and model development, while activities of prediction were missing.

### 3.4 | Modeling product

Thirty-one participants produced a single modeling product at the end of their modeling processes. Two participants produced multiple models as possible solutions for the given task

TABLE 6 Integration of the three necessary concepts by the participants

	Concept 1 (water reservoirs filling up)	Concept 2 (parallel system of diverging paths)	Concept 3 (water reservoirs fully emptying at specific fill level)
Verbalized	32 (91.437%)	7 (20.00%)	11 (31.43%)
Drawn	32 (91.43%)	9 (25.71%)	3 (8.57%)
Verbalized and/or drawn	34 (97.142%)	9 (25.71%)	12 (34.29%)

TABLE 7 Overview of the number of concepts integrated by the participants verbally and/or drawn

Score (=number of concepts integrated)	Participants ( <i>n</i> ) integrating the concepts		Either verbally or drawn (=modeling product score)
	Verbally	Drawn	
0	2 (5.71%)	3 (8.57%)	1 (2.86%)
1	20 (57.14%)	21 (60.00%)	18 (51.43%)
2	9 (25.71%)	11 (31.43%)	12 (34.29%)
3	4 (11.43%)	0	4 (11.43%)

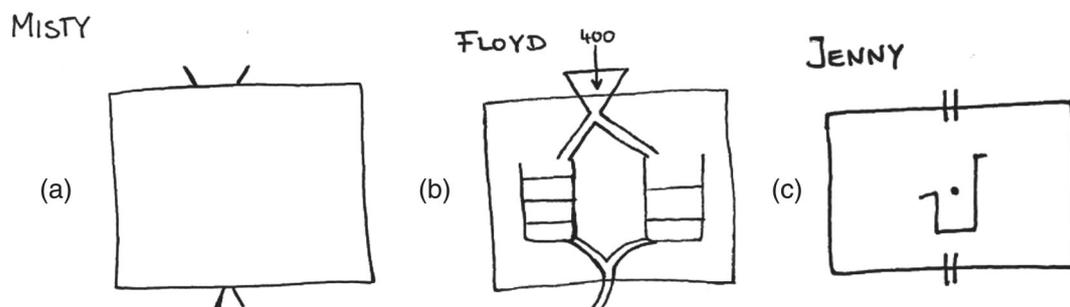


FIGURE 5 Examples of modeling products. Misty's modeling product was scored with 1 point, Floyd's and Jenny's modeling products with 2 points. Note that no participant drew a model that received full score (Table 6)

(*Claudia* presented five, *Kara* two). One participant (*Iris*) removed all her drawn modeling products before ending her modeling process, as she regarded none of the modeling products a suitable solution. Together with *Carlo*, she was one of two participants who had no drawn modeling product, although they verbalized their ideas.

The participants' modeling product scores ranged between zero and three with a mean score of 1.54 ( $SD = .73$ ), indicating most participants were able to develop a model that integrates one or two target concepts. Only four participants reached full score. Table 6 shows how often each specific concept was integrated by the participants either verbally or drawn. Notably, no participant was able to *draw* all three target concepts (Table 7).

The first concept (water reservoirs filling up with water) was included in all models except one: *Misty* (Figure 5a) removed a vessel that she had drawn before from her final model, claiming "it is physically not possible to work like that," concluding that she cannot explain her data with only this one concept. Instead, she proposed—but did not draw—a mechanism at the bottom of the black box consisting of "something like a valve, regulated by pressure." Alternative concepts like these were found with nine other participants. However, these concepts were rarely evaluated as sufficient for the given task by the participants themselves. Of the remaining 33 models, seven also included the second concept (water reservoirs are embedded in a parallel system of two diverging paths). For example, *Floyd's* model (Figure 5b) showed two vessels on the bottom, filling up with water simultaneously. Though he did not draw the third concept (water reservoirs being fully emptied at a specific fill level), as he was unsure how a mechanism working like this could be drawn. Like *Floyd*, three other participants (*Alice*, *Raphael*, and *Susi*) drew a model that included the first two concepts but were only able to verbalize the third concept. In contrast, four participants (like *Jenny*; Figure 5c) were able to draw a mechanism, which included the third concept of water reservoirs being fully emptied at a specific fill level. This behavior was commonly explained with a tilting or turning mechanism. None of the participants drew a siphon, like it is actually used in the presented black box (Figure 2). Moreover, the second concept is drawn more often than explicitly verbalized, while the third concept is drawn rarely but more often verbalized.

Additionally, a Mann–Whitney  $U$  test regarding the participants' subject combinations revealed that secondary preservice biology teachers studying a second scientific subject developed modeling products with significantly higher modeling product scores ( $M = 2.13$ ,  $SD = 0.84$ ) as opposed to secondary preservice biology teachers with nonscientific secondary subjects ( $M = 1.37$ ,  $SD = 0.63$ ,  $z = -2.34$ ,  $p < 0.019$ ;  $d = 0.86$ , large effect size measure). Neither the length of the modeling processes, nor the participants' enrollment status in either the bachelors' or the masters' program could be shown to be systemically related to the quality of the modeling products.

In summary, participant's modeling products were of medium quality, mostly incorporating two of the three target concepts.

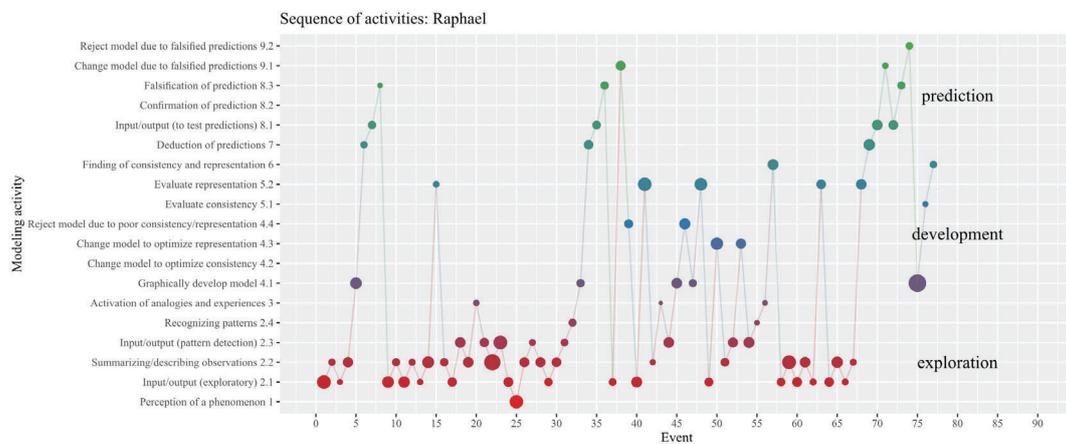
### 3.5 | Relationships between meta-modeling knowledge, modeling practice, and modeling product

The correlation analyses (Spearman) between the two measures of meta-modeling knowledge, the complexity and homogeneity of the modeling processes, and the modeling product scores revealed that there are no significant relationships between the five variables with one exception: the complexity of the modeling processes significantly correlates with the modeling product score (Table 8).

**TABLE 8** Correlation table for Spearman's rank correlations between the variables decontextualized and contextualized meta-modeling knowledge, homogeneity, and complexity of the modeling processes, and the modeling product score

Variable	2	3	4	5
1. Decontextualized meta-modeling knowledge	0.27 (n.s.)	0.17 (n.s.)	0.21 (n.s.)	0.05 (n.s.)
2. Contextualized meta-modeling knowledge		0.25 (n.s.)	0.10 (n.s.)	−0.08 (n.s.)
3. Complexity			−0.01 (n.s.)	<b>0.41 (<math>p &lt; 0.05</math>)</b>
4. Homogeneity				0.11 (n.s.)
5. Modeling product score				

Note:  $N = 29$  for correlation analyses including contextualized meta-modeling knowledge and  $N = 35$  for the other analyses.



**FIGURE 6** Sequence diagram showing the modeling process of Raphael in detail

Based on Mann–Whitney  $U$  tests, no significant differences between modeling processes which included modeling activities of prediction, and modeling processes which did not include modeling activities of prediction, were found for the variables decontextualized meta-modeling knowledge ( $p = 0.61$ ), contextualized meta-modeling knowledge ( $p = 0.19$ ), homogeneity ( $p = 0.59$ ), and modeling product score ( $p = 0.51$ ).

### 3.6 | Exemplary cases

In the following section, we present and discuss two exemplary cases to give more qualitative and holistic insights into the nature of the investigated relationships. With these two cases, we were especially interested in the only statistically significant relationship that was found between the modeling practices and the modeling products. Therefore, two cases with similar values along the complexity of the modeling processes and quality of the modeling products were chosen (case 1: *Raphael*, with high complexity score and high product score; case 2: *Angelina*, with medium complexity score and medium product score). These cases not only shed light on possible explanations for the found statistically significant relationship, but also generate insight into the other, statistically nonsignificant relationships as well.

### 3.6.1 | Example 1: Raphael

*Raphael's* modeling process (Figure 6) takes 1 h and 6 min and consists of 77 events. In his modeling process he shows a wide diversity of different modeling activities including exploration, model development, and prediction, and therefore received a high complexity score of 17 (Table S1).

He starts his modeling process by filling 400 ml of water into the black box twice, observing an output of roughly 400 ml after the second input. After documenting his observations on the board, he draws an initial modeling product (event 5) including the idea of an overflow vessel with a volume of 400 ml. From this modeling product, he deduces the hypothesis, that whatever his next input maybe, the output should be the same and tests his hypothesis by filling 100 ml of water into the black box. Contrary to his expectation, he cannot observe any output, which leads him to falsify and discard his modeling product as “obviously wrong.” He then enters a longer phase (events 11–25) of modeling activities of exploration, collecting data by varying his input volumes and speeds until he recognizes a repeating pattern in the collected data (event 26). From this, he tries to develop a second modeling product retracing his collected data step by step while also simultaneously predicting the next output for each data point and testing it on the black box. Although this directly fails in the first step, as he predicts an output of 100 ml and observes only 50 ml, this leads him to the important observation, that the volume is somehow halved inside the black box. This motivates *Raphael* to modify his modeling product, adding a second compartment in which water is equally distributed. He then constantly switches between modeling activities of exploration and model development, collecting data, retrospectively explaining his collected data and modifying existing or developing new modeling products, which mostly add more compartments (events 35–50). Still not being able to explain all his observations, he concludes, that he needs a mechanism that “if a critical volume is reached, empties all the water or even more.” Without integrating this idea into his drawn modeling product, he assumes that a very big input of 1300 ml would completely empty the black box, regardless of what is still left inside. Confirming this assumption, he verbalizes that the only thing unknown is the number and volume of compartments inside the black box (event 57). Therefore, with the black box now being empty again, he does a series of very small inputs comparing the outputs to his current modeling product (events 58–68). The now repeated observation of small inputs being halved, leads him to reason, that the black box contains two compartments. However, the small inputs did not help him in reasoning about the volume of each of the compartments, but he deduces the hypothesis “that the maximum volume of the black box should be around 1200 ml.” After testing and falsifying this hypothesis (events 70–74), he claims “I am all out of ideas” and draws his final modeling product (Figure 7), which reaches the highest modeling product score. It includes the concepts

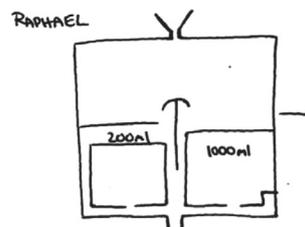


FIGURE 7 Modeling product of Raphael

1 (modeling product contains water reservoirs), 2 (water reservoirs are embedded in a parallel system), and 3 (water fully empties at a specific fill level), with concept 3 being verbalized, as he is unsure how to draw such a mechanism. After evaluating the appearance of his modeling product to an audience with “nobody will understand this, if I am not explaining it,” he deems his modeling product sound “I think, the inside looks something like this.” and ends his modeling process without further testing his modeling product to validate or falsify it.

Contrasting his modeling practice and modeling product, *Raphael's* answers to the questionnaire assessing decontextualized meta-modeling knowledge are all typical level 2 answers ( $M = 2.00$ ), in which he consistently highlights models as idealized representations or media for visualization and explanation. During his modeling practice, only two statements showing his contextualized meta-modeling knowledge ( $M = 2.50$ ) could be identified: In his first statement, he verbalizes a perspective on models more consistent with his answers to the questionnaire (“my model could be an explanation”; level 2). However, in his second statements he also makes clear, that he sees his model as a “hypothetical idea” (level 3), which better fits his modeling practice.

### 3.6.2 | Example 2: Angelina

*Angelina's* modeling process (Figure 8) takes 1 h and 1 min and consists of 62 events. In her modeling process, she shows a lower diversity of different modeling activities and only includes activities of exploration and model development. This leads to a medium complexity score (11).

*Angelina* starts her modeling process with one input of 400 ml. After observing no output, she inputs another 400 ml very slowly, resulting in the siphon inside the black box not emptying the vessel fully, but letting the 200 ml of water trickle out slowly. After documenting her observations, she follows up with a third, now faster, input, leading her to observe an output of 1000 ml as both vessels inside empty through their respective siphon. From this, *Angelina* verbalizes the idea of a vessel connected to a tilting mechanism (events 9–11). However, she does not attempt to draw a modeling product yet. Instead, she focuses on generating more data, following up with numerous, varying inputs, and observation of the outputs. She documents everything in a very thorough and systematic manner (events 12–26) until she claims to

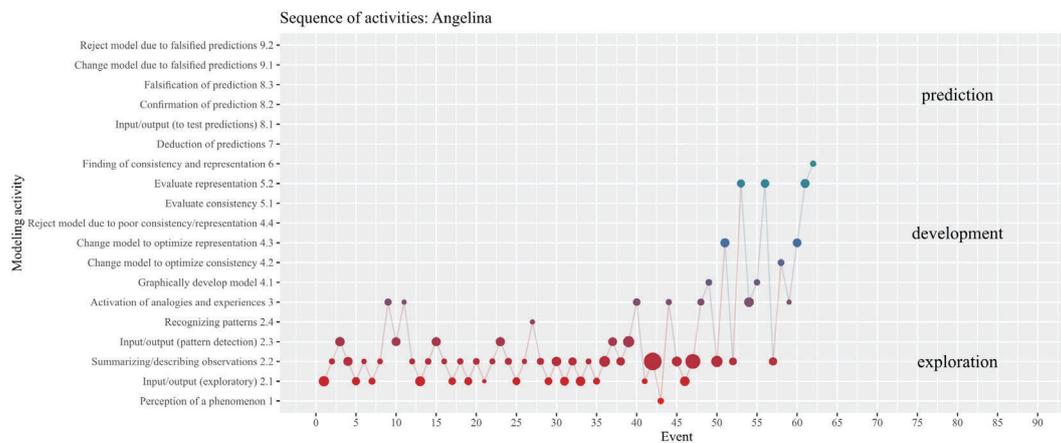


FIGURE 8 Sequence diagram showing the modeling process of Angelina in detail

recognize a reoccurring pattern. Still, she is unable to predict the next outputs from the previously collected data and continues to collect even more data, going into increasingly longer activities of summarizing her previous observations. After verbalizing the idea, that “to explain the varying outputs, multiple vessels with different volumes are necessary” she finally draws a modeling product, depicting three tilting vessels with different volumes. Reflecting on her “mathematically-oriented approach” and still not being able to explain her data with her modeling product, *Angelina* discards it. Activating the analogy of a wheel, she draws her final modeling product (Figure 9), consisting of a round, revolving vessel with different compartments. It reaches a medium modeling product score as it includes the concepts 1 (modeling product contains water reservoirs) and 3 (water fully empties at a specific fill level). With this modeling product she is now able to explain her inconsistent data. However, she expresses, that she is not convinced the modeling product is fully sound and can be seen as “provisional.” With this *Angelina* ends her modeling process.

Although, her scores regarding the modeling practice and modeling product are quite different than *Raphael's*, the scores of *Angelina* for decontextualized and contextualized meta-modeling knowledge perfectly match the scores of *Raphael*. Her answers to the questionnaire assessing decontextualized meta-modeling knowledge are also typical level 2 answers throughout ( $M = 2.00$ ), like *Raphael*, highlighting models as idealized representations or media for visualization and explanation. Even her two statements assigned to contextualized meta-modeling knowledge ( $M = 2.50$ ) are quite similar to *Raphael*: In her first statement, she is consistent with her answers in the questionnaire (“There are different containers, tilting and having different volumes, which explains the unregular volumes coming out”; level 2). In another statement, she reflects about her practice on a meta level, claiming that seeing her idea as a provisional model helps her thinking about it (level 3).

### 3.6.3 | Summary of the two cases

The cases of *Raphael* and *Angelina* illustrate two approaches to the modeling task at hand, which differ in their integration of the modeling product into the modeling practice: *Raphael* uses his modeling product as an epistemic tool by constantly switching between modeling phases of exploration, model development and predictions, in which he strongly relies on his modeling product to test his ideas. This test-driven modeling practice leads him to expose flaws in his modeling product, which he overcomes by then integrating new ideas (i.e., analogies) into his modeling product. Interestingly, while he shows this test-driven modeling practice with his initial, most simplistic modeling product of an overflow vessel (lowest modeling product score),



FIGURE 9 Modeling product of Angelina

he does not show this with his final modeling product (highest modeling product score). In contrast, *Angelina* approaches the modeling task using a more data-driven approach. After a long phase of exploration, collecting a lot of observations in a data table, she develops her first modeling product in the last third of her modeling practice attempting to explain the observed data retrospectively. While this also helps her in generating new ideas about the inner mechanism of the black box, she does not test her modeling product by drawing hypotheses and predicting outputs from the modeling product. Instead, she compares newly collected data to her previous observations and changes her modeling product to fit the data retroactively. Lastly, regarding their meta-modeling knowledge, both *Raphael* and *Angelina* achieve exactly the same scores for contextualized and decontextualized meta-modeling knowledge, although their modeling practice and product widely differ.

## 4 | DISCUSSION

In the present study, we set out to identify, characterize, and test the correlation between the three dimensions of modeling competence: meta-modeling knowledge, modeling practices, and modeling product. These three dimensions were operationalized as five variables, including decontextualized and contextualized meta-modeling knowledge for the meta-modeling knowledge dimension, complexity and homogeneity of the modeling processes for the modeling practices dimension, and a modeling product quality score. This approach aimed to uncover the relationships between the three dimensions, which literature proposes to be highly and positively related (Gobert & Pallant, 2004; Lee & Kim, 2014; Schwarz & White, 2005) but has not been investigated comprehensively yet (Chiu & Lin, 2019; Louca & Zacharia, 2012; Nicolaou & Constantinou, 2014).

### 4.1 | Meta-modeling knowledge

In line with earlier studies, the 35 secondary preservice biology teachers' decontextualized meta-modeling knowledge seems to be rather limited (Table 4), as their views of models in their answers to the questionnaire suggest mainly medial perspectives on models (Krell & Krüger, 2016; Torres & Vasconcelos, 2015). More sophisticated views of models, in which models are seen as epistemic tools that are being tested by deducing and testing predictions, are rarely addressed (i.e., level III; Krell & Krüger, 2016). However, theoretically less informed views about scientific models, reflecting a naïve understanding of models as copies of reality, were also only rarely observed, contradicting previous reports of these being quite common (Torres & Vasconcelos, 2015). For contextualized meta-modeling knowledge, it is suggested that specific contexts, including the black box modeling task used here, provoke more elaborate views of models (Ke & Schwarz, 2020). While this could be confirmed for the aspect *nature of models* (Figure 3), the average level of contextualized meta-modeling knowledge is only slightly higher compared to the decontextualized assessment and not significantly different statistically (Table 5). This may suggest that the specific black box modeling task used in this study emphasizes the hypothetical character of models, as the secondary preservice biology teachers are constructing their own hypothetical model, being made aware that models do not have to be final solutions or ready-made explanations (e.g., *Raphael*). Generally, the participants'

decontextualized and contextualized meta-modeling knowledge varies between the different aspects of meta-modeling knowledge, suggesting unstable views about scientific models (cf., Krell et al., 2014).

## 4.2 | Modeling practice

Regarding the participants' modeling practices observed in this study, only some individual modeling processes can be seen as meaningful engagement, characterized by reflective insight, elaborate relational reasoning and monitoring the models' structure, and modeler's goals leading to the systematical testing of hypotheses and scoping of variables (Sins et al., 2005). While the correlation analysis between the complexity of the modeling processes and the length of the participants' modeling processes revealed a significant positive correlation, the nature of this relation does not become particularly clear, considering the qualitative data collected. However, it is obvious that with longer time on task, there is a higher probability of addressing more modeling activities. Still, only half of the modeling processes observed included activities of prediction. This is in line with previous studies suggesting that the predictive use of models is challenging for most sample groups, including teachers and students (Krell & Krüger, 2016; Passmore et al., 2014). Additionally, it is suggested in the literature that an evaluation of the modeling products by attempts of falsification or by systematically testing alternative hypotheses is also important for scientific modeling (Louca & Zacharia, 2015). Neither of these behaviors could be observed. A more in-depth discussion of our findings regarding the modeling practices including a comparison to other scientific practices like experimentation can be found in Göhner and Krell (2020a).

## 4.3 | Modeling product

In our sample group and for the given modeling task of the black box, the quality of the modeling products was rather low, with not a single participant drawing a modeling product incorporating all three necessary concepts (Table 6). This indicates a high difficulty of the given black box modeling task, which was initially chosen over an authentic and content-rich biological problem to reduce the influence of prior knowledge on the secondary preservice biology teachers' modeling processes. This advantage of the black box approach, on the other hand, is potentially limiting the generalizability of our findings and highlights the importance of context and content knowledge for modeling (Ruppert et al., 2017). This is further emphasized by some of our sample characteristics: secondary preservice biology teachers studying a second scientific subject developed modeling products with significantly higher modeling product scores than secondary preservice biology teachers with a nonscientific secondary subject. However, it is not evident from the qualitative data, if this correlation is grounded in the higher content knowledge, especially regarding physics knowledge, of participants studying two scientific subjects or if they are simply more familiar in working with scientific models. Regarding the latter, we rarely observed participants labeling and keying all the elements of their modeling products or developing explicit comparative modeling products, which was also observed by Bamberger and Davis's (2013) regarding student's modeling products prior to an instructional intervention.

#### 4.4 | Relationships between meta-modeling knowledge, modeling practice, and modeling product

Investigating the relationships between the three dimensions reveals that not all relationships proposed in the literature were observed in the present study. While the relationship between modeling practices and modeling products was found to be statistically significant, no statistical evidence was found for the most commonly proposed relationship between meta-modeling knowledge and modeling products.

##### 4.4.1 | The relationship between modeling practice and modeling product

In this study, a positive significant correlation between complexity of modeling processes and quality of modeling products was found (Table 8). This finding suggests a positive relationship between modeling practices and modeling products. This is reassuring, given that the analysis of modeling products is commonly used as a proxy for the more time-consuming analysis of modeling practice (Bamberger & Davis, 2013; Cheng & Lin, 2015; Ergazaki et al., 2007; Schwarz et al., 2009). However, the correlation coefficient indicates a medium effect size, suggesting a shared variance of about 17%. Most participants developed modeling products of rather low quality (Table 6 and Table S1). This matches their modeling processes of rather low complexity, which often did not include modeling activities of prediction. While in theory, this could be explained by low-quality modeling products limiting modelers in their modeling activities, making them unable to deduce suitable hypotheses and test them, no direct evidence for this causal relationship could be found in the qualitative data. Quite the contrary, the case of *Raphael* illustrates that even the most simplistic modeling product of a single overflow vessel can be used to deduce a hypothesis leading to the modeling product being tested and revised. Moreover, other observations suggest that the relationship between modeling practice and modeling product might be more complex. *Raphael*, ending his modeling process without testing his final highest-scoring modeling product in the same predictive manner, suggests that a high-quality modeling product does not automatically lead to complex modeling practices, although attempting a falsification of a developed model is considered to be an utterly important modeling practice (Giere et al., 2006).

Comparing *Raphael's* approach to the modeling task at hand with *Angelina's* approach, indicates that the *perceived soundness*—how we suggest to call it—of a modeling product may explain if and when a modeling product is used for predictions during the modeling practice. *Raphael* deems his modeling products as sound enough to reason with them and to test ideas, until his final modeling product is so sound it can be presented as a solution to the task. In contrast, *Angelina*, who shows a more data-driven approach and does not predict from her model at all, as she perceives all her modeling products, including the final one as not very sound (or “provisional”). She makes clear, that she is unsure in multiple instances, as she is not able to reproduce or to explain the black box phenomenon retrospectively until her final modeling product. This factor of perceived soundness could also explain, why *Angelina* does not draw a model at all during the first two thirds of her modeling practice, as she perceives her data as rather unsound until she recognizes a pattern, from which she is able to develop her first modeling product.

Furthermore, it is especially interesting that evidence for a relationship between modeling practices and modeling products was found by evaluating the modeling products based on the

concepts they integrated (i.e., their content) and not by using epistemic criteria, like for example parsimony or conceptual coherence (Pluta et al., 2011). This adds more evidence to a potential influence of context on performance (Krell et al., 2014; Schwarz, 2002; Sikorski, 2019). Content knowledge (and therefore context) can be seen as a limiting or moderating factor, whose important role for scientific modeling has already been emphasized by different authors (e.g., Ruppert et al., 2017). As most of the participants in this study were studying biology as their only scientific discipline ( $n = 30$ ), they might lack relevant physics knowledge to develop an appropriate modeling product for the inner mechanism of the black box. The task of discovering a black box is rather abstract, and hence, not representative for problem solving in authentic, content-rich scientific contexts (Leden et al., 2020). Therefore, it remains unclear how exactly domain-specific and situated content knowledge moderates the quality of the modeling products, as well as meta-modeling knowledge and modeling processes. Our findings emphasize a strong need for a systematic analysis of the role of content knowledge for modeling in research and the importance of context for interventions in educational settings.

#### 4.4.2 | The relationship between meta-modeling knowledge and modeling practice

In general, statistical evidence for a relationship between meta-modeling knowledge and modeling practices, as well as meta-modeling knowledge and the two other variables considered in this study (Table 8), were surprisingly absent. This contradicts common assumptions of meta-modeling knowledge guiding the practice (Louca & Zacharia, 2012; Schwarz et al., 2009) and indicates that meta-modeling knowledge might not be a valid predictor for the quality of engagement in modeling practices and the modeling product. Of course, the small sample size of this study and the limitations discussed below have to be considered in the generalizability of these findings, but they clearly show that more research regarding this relationship is needed.

Based on the average scores of meta-modeling knowledge, the present findings suggest that the secondary preservice biology teachers in our sample with rather low levels of meta-modeling knowledge are somehow able to engage in elaborate modeling practices. These modeling practices could be observed as modeling processes, which are characterized by high complexity and homogeneity scores, and include activities of prediction, using their models for reasoning about the black box (e.g., *Raphael*). On the contrary, some participants with high levels of meta-modeling knowledge, describing models consistently as hypothetical entities for knowledge generation, showed less elaborate modeling practices, including no activities of prediction in the herein observed modeling processes. These participants typically used their modeling products just to illustrate and communicate their solution (*Carlo*, see Table S1). Although Ke and Schwarz (2020) propose that meta-modeling knowledge in action might be stronger related with the practices (in contrast to abstract, decontextualized meta-modeling knowledge), the average contextualized meta-modeling knowledge was also not found to be significantly related to the complexity of the modeling processes. However, the response levels for both decontextualized and contextualized meta-modeling knowledge widely varied across the statements of single participants (e.g., *Claudia*, see Table S1), suggesting unstable views of models. In line with the observations regarding modeling products in this study, it seems likely that meta-modeling knowledge is activated by the participants based on the context of assessment and the aspect of meta-modeling knowledge they address. This is in concurrence to previous studies, discussing the context-dependency of meta-modeling knowledge and the construct

of modeling competence as a whole (Gobert & Pallant, 2004; Krell et al., 2014; Sikorski, 2019; Sins et al., 2009).

The qualitative analyses of the presented cases of *Raphael* and *Angelina* shed some light on the relationship between meta-modeling knowledge and modeling practices. On the one hand, their direct comparison provides more evidence to the absence of a relationship between meta-modeling knowledge and modeling practices, as both show the exact same scores regarding meta-modeling knowledge but engage in completely different modeling practices. On the other hand, the case of *Angelina* clearly shows someone with consistent scores over all assessed dimensions of modeling competence, implying there might be a relationship, which just could not be found to be of statistical significance in this study. Qualitative analyses of the other cases, not presented in this study, indicate, that single responses regarding the aspects *testing models* and *changing models* often are in line with the observed modeling processes. However, these statements leave unclear whether meta-modeling knowledge guides the practice or engagement in practices promotes understanding (Gobert & Pallant, 2004; Schwarz et al., 2009). Moreover, our findings suggest that helping secondary preservice biology teachers improve their meta-modeling knowledge, for example in university courses, may not necessarily improve their modeling practice (Shi et al., 2021).

#### 4.5 | Limitations of the study

This study has some limitations. First and foremost, the sample size is modest and only consists of secondary preservice biology teachers, which potentially limits the generalizability of our findings to this specific sample group. We hope our study can give an example to investigate other sample groups of science teachers. Further limitations must be considered regarding our methodology. As in most similar studies, meta-modeling knowledge (decontextualized and contextualized) was assessed as views on models (Krell & Krüger, 2016; Schwarz et al., 2009). Nicolaou and Constantinou (2014), though, distinguish between two dimensions of the overarching construct of modeling metaknowledge: the herein investigated meta-modeling knowledge (the epistemological awareness about the purpose and use of models) and metacognitive knowledge of the modeling process (the understanding of the actual modeling process).

Recently it has been suggested to also assess the latter (Lazenby et al., 2020). Yet to date, there is no established form of assessment in science education research, which addresses this dimension of metacognitive knowledge of the modeling process. The development of such an assessment tool based on a coherent and sound theoretical framework and independent of the traditional approaches to meta-modeling knowledge might yield further insights into the construct of modeling competence and its development (Nicolaou & Constantinou, 2014). Additionally, the participants in this study were not explicitly asked for verbalizing contextualized meta-modeling knowledge, leading to incomplete data sets, as not every participant verbalized thoughts about every aspect of meta-modeling knowledge throughout their modeling processes. Regarding the assessment of the modeling processes, the black box as a modeling task may have limited the participants in their engagement, as it is a rather abstract and complex task (Leden et al., 2020). Furthermore, we cannot be sure if all reasoning processes of the participants are fully revealed by them concurrently thinking-aloud. For the quality of the modeling products, the participants were limited to develop their modeling products by drawing. Although this approach is quite common in science education research (Chang et al., 2020), modeling products could also be developed via computational modeling or hands-on modeling. The given black box modeling task, which allowed only for drawing, may have limited some of the

participants in their modeling processes. Also, it should be acknowledged, that an evaluation of the modeling products based on epistemic criteria may yield further insights (Pluta et al., 2011). However, this was not applicable in the present study given the overall lack of complexity and variance of the modeling products.

Finally, this study solely focused on the modeling competence of secondary preservice biology teachers, without taking their ability to teach modeling competence to students into consideration, which is suggested to be even more challenging for teachers (Shi et al., 2021) and develops over vast time frames (Vo et al., 2019).

## 5 | CONCLUSION

The present study shows that the established assumptions, held by many science education researchers, about the dimensions that constitute modeling competence and their relationships, might not be empirically valid, at least for secondary preservice biology teachers. Firstly, our findings highlight that secondary preservice biology teachers' meta-modeling knowledge is not a reliable indicator for their engagement in modeling practices or for modeling competence in general. Secondly, the quality of modeling products seems to be a more reliable proxy assessment, as it corresponds to the observed modeling practice in this study. However, it becomes clear, that a valid assessment of secondary preservice biology teachers' modeling competence needs to consider meta-modeling knowledge, modeling practices, and modeling products holistically. Although more research is needed regarding the metacognitive knowledge of the modeling process or the influence of context on modeling, we hope that our study provides valuable insights for researchers, who work on an assessment of modeling competence. Untangling the construct of modeling competence further, could lead to further development of educational interventions, aiming to foster modeling competence in teachers and, consequently, improve the integration of models and modeling in classrooms worldwide.

## ACKNOWLEDGMENTS

We thank the German Research Foundation (DFG) for funding the present project (award nr. 327507949). Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the DFG. Open Access funding enabled and organized by Projekt DEAL.

## ORCID

Maximilian Felix Göhner  <https://orcid.org/0000-0003-4938-372X>

Tom Bielik  <https://orcid.org/0000-0002-8227-5870>

Moritz Krell  <https://orcid.org/0000-0003-2226-0383>

## REFERENCES

- Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education*, 22(7), 665–701.
- ACARA (2015). *The Australian curriculum: science*. Retrieved from <http://www.australiancurriculum.edu.au/>
- Andrienko, N., & Andrienko, G. (2018). State transition graphs for semantic analysis of movement behaviours. *Information Visualization*, 17(1), 41–65.
- Bamberger, Y. M., & Davis, E. A. (2013). Middle-school science students' scientific modelling performances across content areas and within a learning progression. *International Journal of Science Education*, 35(2), 213–238.

- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–48). Springer US.
- BCMOE (2019). *Curriculum redesign*. Retrieved from <https://curriculum.gov.bc.ca/rethinking-curriculum>
- Bielik, T., Opitz, S., & Novak, A. (2018). Supporting students in building and using models: Development on the quality and complexity dimensions. *Education Sciences*, 8(3), 149.
- Campbell, T., Oh, P. S., Maughn, M., Kiriazis, N., & Zuwallack, R. (2015). A review of modeling pedagogies: Pedagogical functions, discursive acts, and technology in modeling instruction. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(1), 159–176.
- Carlson, J., & Daehler, K. (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.
- Chang, H.-Y., Lin, T.-J., Lee, M.-H., Lee, S. W.-Y., Lin, T.-C., Tan, A.-L., & Tsai, C.-C. (2020). A systematic review of trends and findings in research employing drawing assessment in science education. *Studies in Science Education*, 56(1), 77–110.
- Cheng, M.-F., & Lin, J.-L. (2015). Investigating the relationship between students' views of scientific models and their development of models. *International Journal of Science Education*, 37(15), 2453–2475.
- Cheng, M.-F., Wu, T.-Y., & Lin, S.-F. (2021). Investigating the relationship between views of scientific models and modeling practice. *Research in Science Education*, 51, 307–323. <https://doi.org/10.1007/s11165-019-09880-2>
- Chiu, M.-H., & Lin, J.-W. (2019). Modelling competence in science education. *Disciplinary and Interdisciplinary Science Education Research*, 1, 1–11. <https://doi.org/10.1186/s43031-019-0012-y>
- Constantinou, C. P. (1999). The Cocoa Microworld as an environment for developing modelling skills in physical science. *International Journal of Continuing Engineering Education and Life Long Learning*, 9(3-4), 201–209.
- Crawford, B. A., & Cullin, M. J. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 309–323). Springer.
- Dounas-Frazer, D. R., & Lewandowski, H. J. (2018). The modelling framework for experimental physics: Description, development, and applications. *European Journal of Physics*, 39(6), 64005.
- Ergazaki, M., Zogza, V., & Komis, V. (2007). Analysing students' shared activity while modeling a biological process in a computer-supported educational environment. *Journal of Computer Assisted Learning*, 23(2), 158–168.
- Fretz, E. B., Wu, H.-K., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2002). An investigation of software scaffolds supporting modeling practices. *Research in Science Education*, 32(4), 567–589.
- Giere, R. N. (1999). *Science without laws*. The University of Chicago Press.
- Giere, R. N., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning*. Thomson Wadsworth.
- Gilbert, J. K., & Justi, R. (2016). *Modelling-based teaching in science education*. Springer.
- Gobert, J. D., & Pallant, A. (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology*, 13(1), 7–22.
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, 21(5), 725–740.
- Göhner, M., & Krell, M. (2020a). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*. <https://doi.org/10.1007/s11165-020-09945-7>
- Göhner, M., & Krell, M. (2020b). Analyzing the relationships between pre-service biology teachers' modeling processes, scientific reasoning competencies and general cognitive abilities. In M. Ergazaki & K. Kampourakis (Eds.), *Contributions from biology education research*. Springer Nature.
- Günther, S. L., Fleige, J., Upmeier zu Belzen, A., & Krüger, D. (2019). Using the case method to foster preservice biology teachers' content knowledge and pedagogical content knowledge related to models and modeling. *Journal of Science Teacher Education*, 30(4), 321–343.
- Harlow, D. B., Bianchini, J. A., Swanson, L. H., & Dwyer, H. A. (2013). Potential teachers' appropriate and inappropriate application of pedagogical resources in a model-based physics course: A “knowledge in pieces” perspective on teacher learning. *Journal of Research in Science Teaching*, 50(9), 1098–1126.
- Harré, R. (1970). *The principles of scientific thinking*. University of Chicago Press.

- Jonassen, D. H., & Ionas, I. G. (2008). Designing effective supports for causal reasoning. *Educational Technology Research and Development*, 56(3), 287–308.
- Justi, R., & Van Driel, J. H. (2006). The use of the interconnected model of teacher professional growth for understanding the development of science teachers' knowledge on models and modelling. *Teaching and Teacher Education*, 22, 437–450.
- Ke, L., & Schwarz, C. V. (2020). Supporting students' meaningful engagement in scientific modeling through epistemological messages: A case study of contrasting teaching approaches. *Journal of Research in Science Teaching*, 58(3), 335–365.
- Khan, S. (2011). What's missing in model-based teaching. *Journal of Science Teacher Education*, 22(6), 535–560.
- KMK (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung [Common federal requirements for science and science education in teacher education]*. Retrieved from [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_16-Fachprofile-Lehrerbildung.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf)
- KMK (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife [Educational standards in biology for the higher education entrance qualification]*. Retrieved from [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2020/2020\\_06\\_18-BildungsstandardsAHR\\_Biologie.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf)
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, 42(2), 262–271.
- Krell, M., & Krüger, D. (2016). Testing models: A key aspect to promote teaching activities related to models and modelling in biology lessons? *Journal of Biological Education*, 50(2), 160–173.
- Krell, M., Upmeyer zu Belzen, A., & Krüger, D. (2014). Context-specificities in students' understanding of models and modelling. In C. P. Constantinou, N. Papadouris, & A. Hadjigeorgiou (Eds.), *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning*, Nicosia, Cyprus. pp. 1024–1035.
- Krell, M., Walzer, C., Hergert, S., & Krüger, D. (2019). Development and application of a category system to describe pre-service science teachers' activities in the process of scientific modelling. *Research in Science Education*, 49(5), 1319–1345.
- Krüger, D., Hartmann, S., Nordmeier, V., & Upmeyer zu Belzen, A. (2020). Measuring scientific reasoning competencies: Multiple aspects of validity. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student Learning in German Higher Education* (pp. 261–280). Springer.
- Lazenby, K., Stricker, A., Brandriet, A., Rupp, C., Mauger-Sonnek, K., & Becker, N. (2020). Mapping undergraduate chemistry students' epistemic ideas about models and modeling. *Journal of Research in Science Teaching*, 57, 794–824. <https://doi.org/10.1002/tea.21614>
- Leden, L., Hansson, L., & Ideland, M. (2020). The mangle of school science practice: Teachers' negotiations of two nature of science activities at different levels of contextualization. *Science Education*, 104(1), 5–26.
- Lederman, N. G., & Abd-El-Khalick, F. (2002). Avoiding de-natured science: Activities that promote understandings of the nature of science. In W. McComas (Ed.), *The Nature of Science in Science Education* (pp. 83–126). Springer.
- Lee, S., & Kim, H.-B. (2014). Exploring secondary students' epistemological features depending on the evaluation levels of the group model on blood circulation. *Science & Education*, 23(5), 1075–1099.
- Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. *Handbook of Child Psychology*, 6, 153–196.
- Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Lenhard, W., & Lenhard, A. (2016). *Computation of effect sizes*. Retrieved from [https://www.psychometrica.de/effect\\_size.html](https://www.psychometrica.de/effect_size.html)
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R: IST 2000 R*. Hogrefe.
- Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471–492.
- Louca, L. T., & Zacharia, Z. C. (2015). Examining learning through modeling in k-6 science education. *Journal of Science Education and Technology*, 24(2), 192–215.

- Mahr, B. (2011). On the epistemology of models. In G. Abel & J. Conant (Eds.), *Berlin studies in knowledge research: Rethinking epistemology* (Vol. 1, 1st ed., pp. 301–352). De Gruyter.
- Meister, S., Krell, M., Göhner, M., & Upmeyer zu Belzen, A. (2021). Pre-service biology teachers' responses to first-hand anomalous data during modelling processes. *Research in Science Education*, *51*, 1459–1479. <https://doi.org/10.1007/s11165-020-09929-7>
- Nehring, A., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2015). Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic variables. *International Journal of Science Education*, *37*(9), 1343–1363.
- 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*. Information Age.
- Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.
- NGSS Lead States (2013). Next generation science standards: For states, by states: Appendix F - science and engineering practices in the NGSS. In NGSS Lead States (Ed.), *Next generation science standards: For states, by states*. The National Academies Press.
- Nicolaou, C. T., & Constantinou, C. P. (2014). Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, *13*, 52–73.
- Nielsen, S. S., & Nielsen, J. A. (2021a). A competence-oriented approach to models and modelling in lower secondary science education: Practices and rationales among Danish teachers. *Research in Science Education*, *51*, 565–593. <https://doi.org/10.1007/s11165-019-09900-1>
- Nielsen, S. S., & Nielsen, J. A. (2021b). Models and modelling: Science teachers' perceived practice and rationales in lower secondary school in the context of a revised competence-oriented curriculum. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(4) 1–18. <https://doi.org/10.29333/ejmste/10790>
- NRC. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
- Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education*, *33*(8), 1109–1130.
- Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, *25*(2), 177–196.
- Passmore, C. M., Gouvea, J. S., & Giere, R. (2014). Models in science and in learning science: Focusing scientific practice on sense-making. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 1171–1202). Springer.
- Passmore, C. M., & Svoboda, J. (2012). Exploring opportunities for argumentation in modeling classrooms. *International Journal of Science Education*, *34*(10), 1535–1554.
- Patton, M. (1990). Purposeful sampling. *Qualitative Evaluation and Research Methods*, *2*, 169–186.
- Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching*, *48*, 486–511.
- Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). Communities in networks. *Notices of the American Mathematical Society*, *56*(9), 1082–1097.
- Ronqui, J. R. F., & Travieso, G. (2015). Analyzing complex networks through correlations in centrality measurements. *Journal of Statistical Mechanics: Theory and Experiment*, *2015*(5), P05030.
- Ruppert, J., Duncan, R. G., & Chinn, C. A. (2017). Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education*, *49*, 921–948. <https://doi.org/10.1007/s11165-017-9656-9>
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE Publications.
- Schwarz, C. V. (2002). Is there a connection? The role of meta-modeling knowledge in learning with models. Retrieved from [http://schwarz.wiki.educ.msu.edu/file/view/Schwarz\\_ICLS\\_metapaper.pdf](http://schwarz.wiki.educ.msu.edu/file/view/Schwarz_ICLS_metapaper.pdf)
- Schwarz, C. V., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654.
- Schwarz, C. V., & White, B. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, *23*(2), 165–205.
- Shi, F., Wang, L., Liu, X., & Chiu, M.-H. (2021). Development and validation of an observation protocol for measuring science teachers' modeling-based teaching performance. *Journal of Research in Science Teaching*, *58* (9), 1359–1388. <https://doi.org/10.1002/tea.21712>

- Sikorski, T.-R. (2019). Context-dependent “Upper Anchors” for learning progressions. *Science & Education*, 28(8), 957–981.
- Sins, P. H., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The difficult process of scientific modelling: An analysis of novices’ reasoning during computer-based modelling. *International Journal of Science Education*, 27(14), 1695–1721.
- Sins, P. H., Savelsbergh, E. R., van Joolingen, W. R., & van Hout-Wolters, B. H. (2009). The relation between students’ epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education*, 31(9), 1205–1229.
- Torres, J., & Vasconcelos, C. (2015). Nature of science and models: Comparing Portuguese prospective teachers’ views. *EURASIA Journal of Mathematics, Science and Technology Education*, 11(6), 1473–1494.
- Upmeier zu Belzen, A., Krüger, D., & van Driel, J. H. (Eds.) (2019). Models and modeling in science education. In *Towards a competence-based view on models and modeling in science education*. Springer International Publishing.
- Upmeier zu Belzen, A., van Driel, J. H., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger, & J. H. van Driel (Eds.), *Towards a competence-based view on models and modeling in science education* (pp. 3–19). Springer International Publishing.
- VCAA (2016). Victorian curriculum: F-10. Retrieved from <https://victoriancurriculum.vcaa.vic.edu.au/science/curriculum/f-10>
- Vo, T., Forbes, C. T., Zangori, L., & Schwarz, C. (2015). Fostering 3rd-grade students’ use of scientific models with the water cycle: Elementary teachers’ conceptions and practices. *International Journal of Science Education*, 37(15), 2411–2432.
- Vo, T., Forbes, C. T., Zangori, L., & Schwarz, C. (2019). A longitudinal investigation of primary inservice teachers’ modelling the hydrological phenomena. *International Journal of Science Education*, 41(18), 2788–2807.
- Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941–967.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.

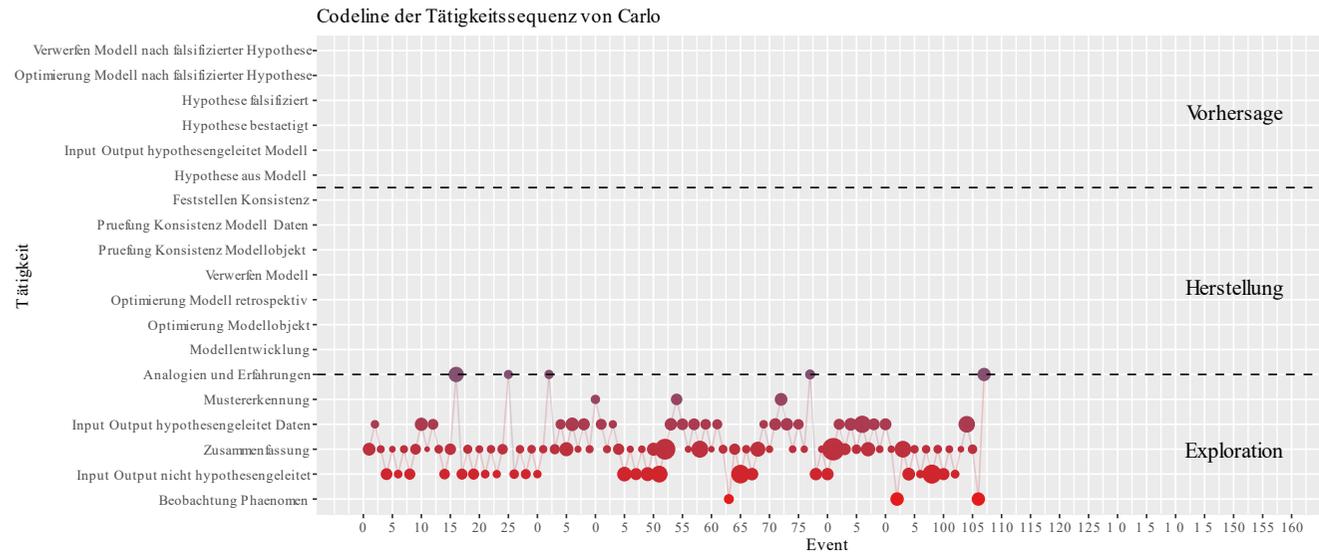
**How to cite this article:** Göhner, M. F., Bielik, T., & Krell, M. (2022). Investigating the dimensions of modeling competence among preservice science teachers: Meta-modeling knowledge, modeling practice, and modeling product. *Journal of Research in Science Teaching*, 59(8), 1354–1387. <https://doi.org/10.1002/tea.21759>

## Anhang

### 9.3 Material & Daten

#### 9.3.1 Codelines & Zeichnungen der Modellierprodukte (geordnet nach Typ)

##### Typ 1

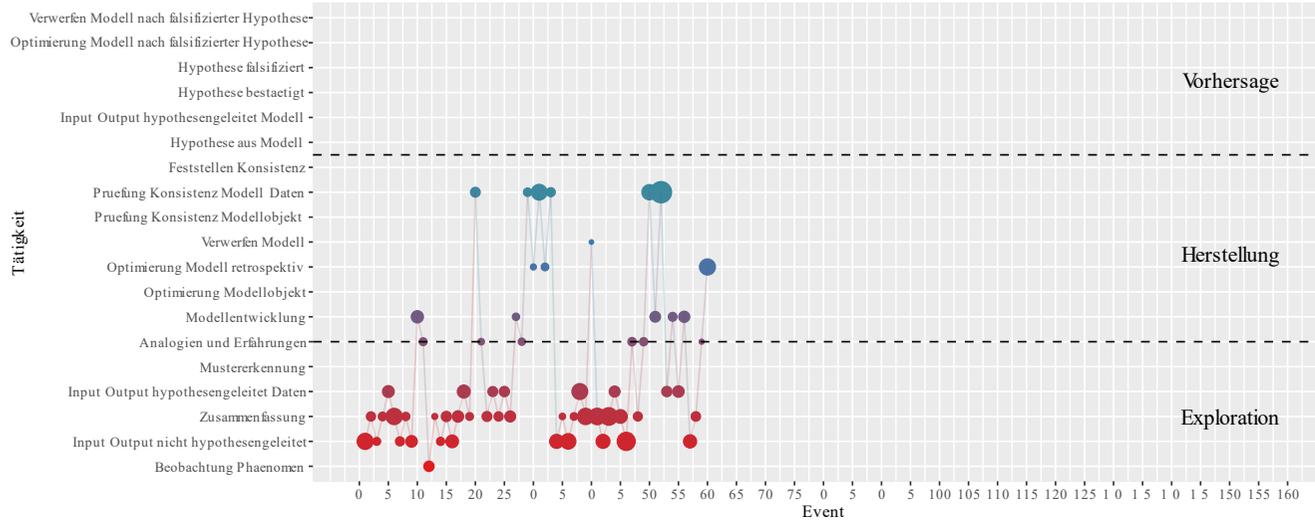


Carlo stellt kein Modellierprodukt her

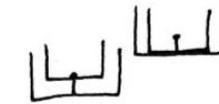
# Anhang

## Typ 2a

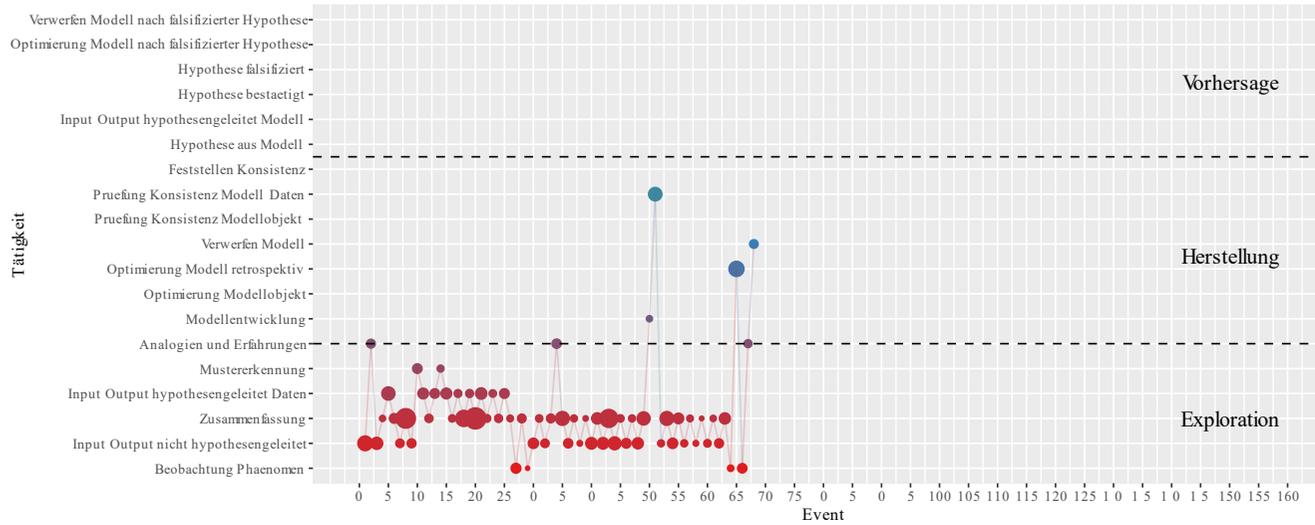
Codeline der Tätigkeitssequenz von Alexandra



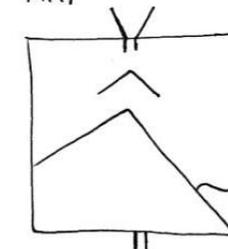
ALEXANDRA



Codeline der Tätigkeitssequenz von Amy



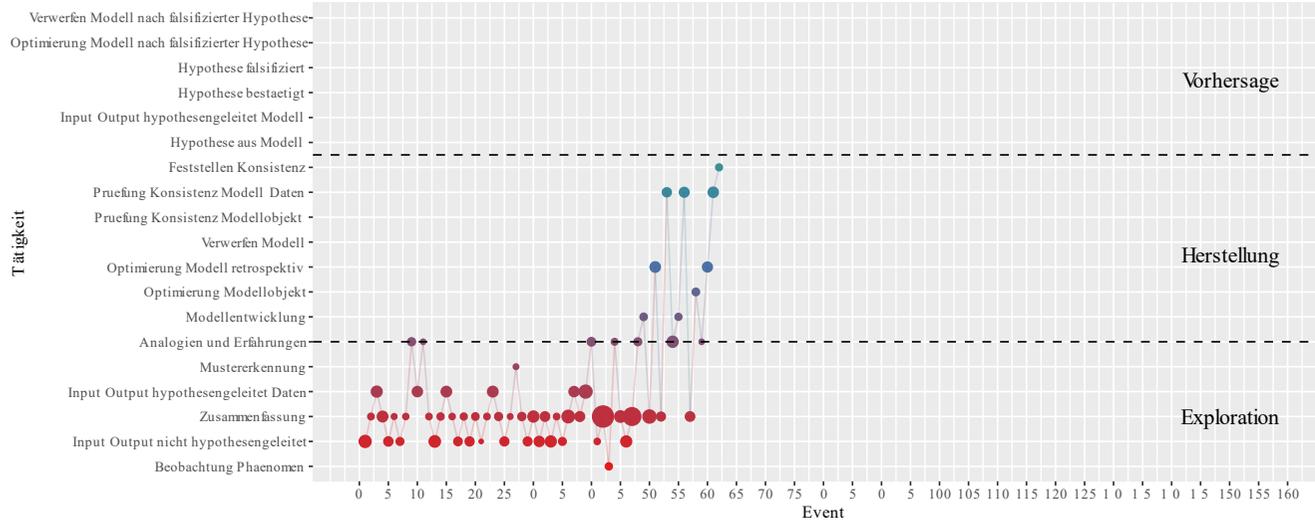
Amy



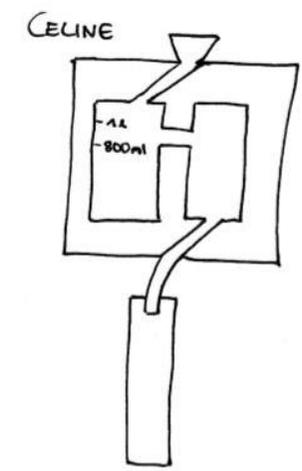
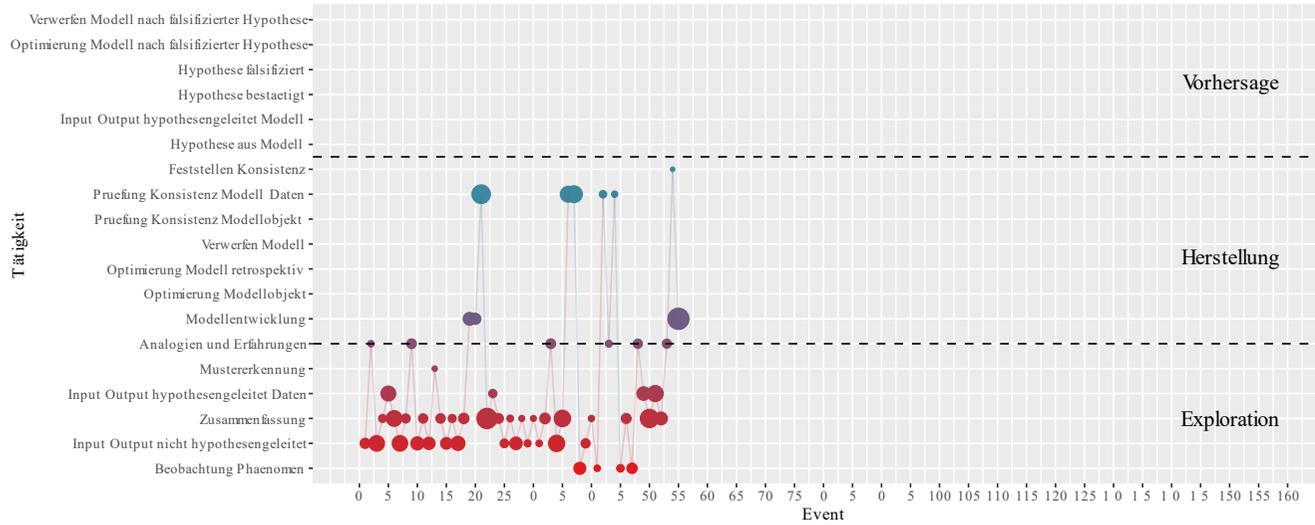
(erased it last second)

# Anhang

Codeline der Tätigkeitssequenz von Angelina

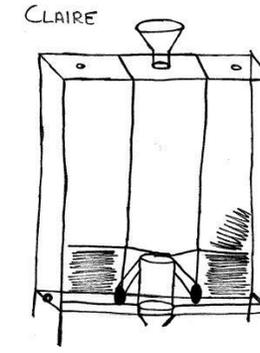
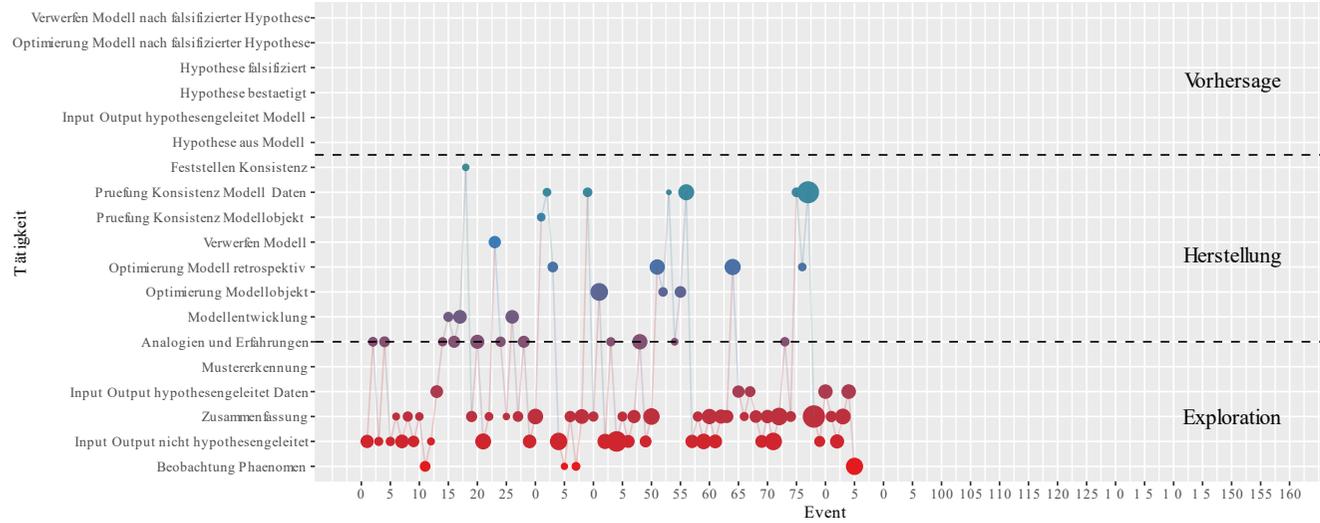


Codeline der Tätigkeitssequenz von Celine

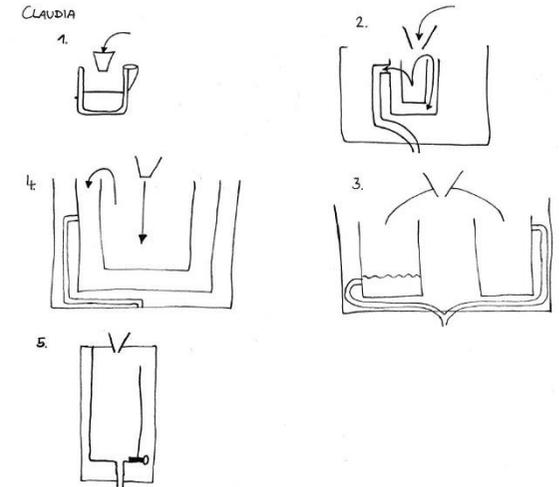
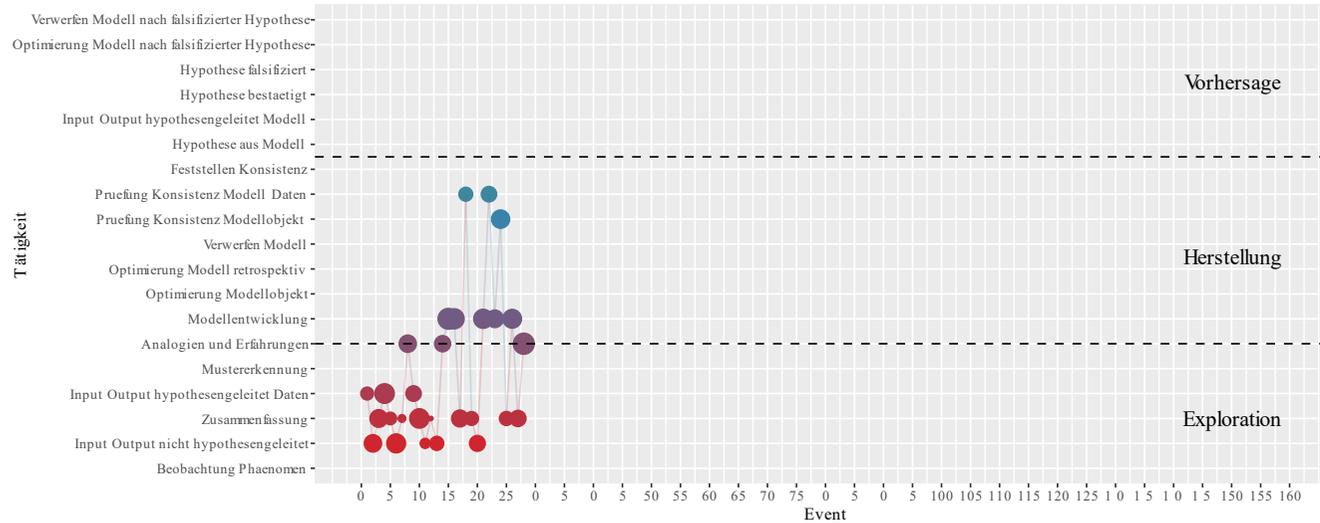


# Anhang

Codeline der Tätigkeitssequenz von Claire

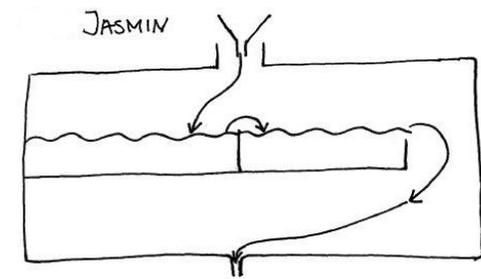
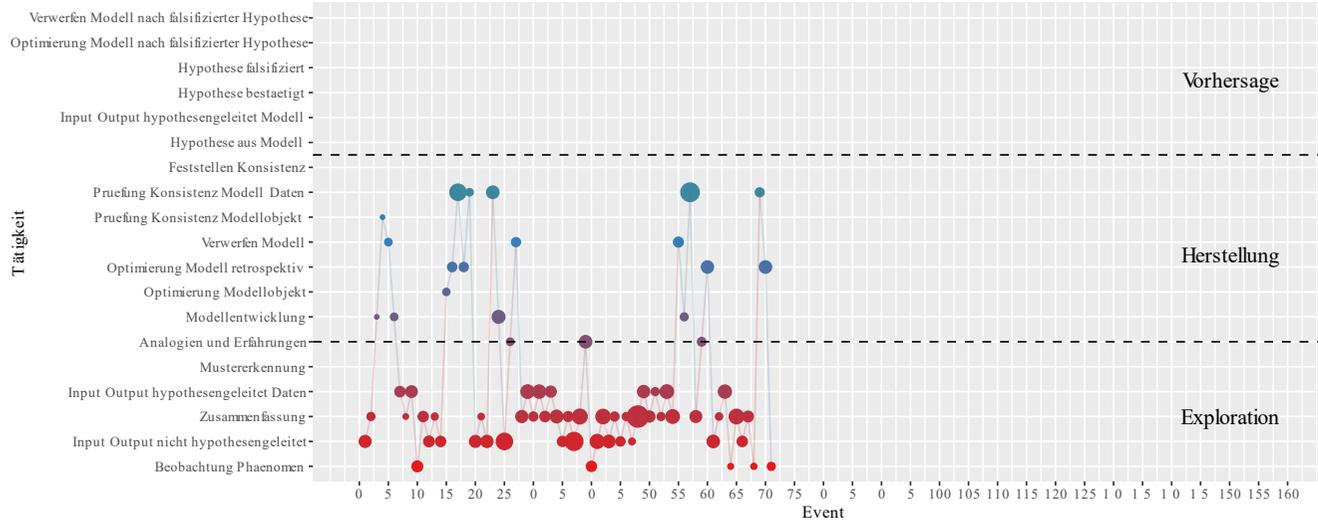


Codeline der Tätigkeitssequenz von Claudia

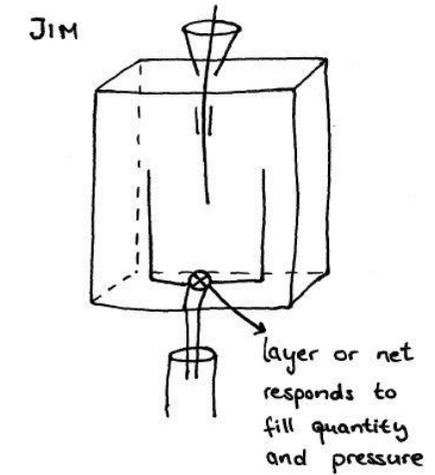
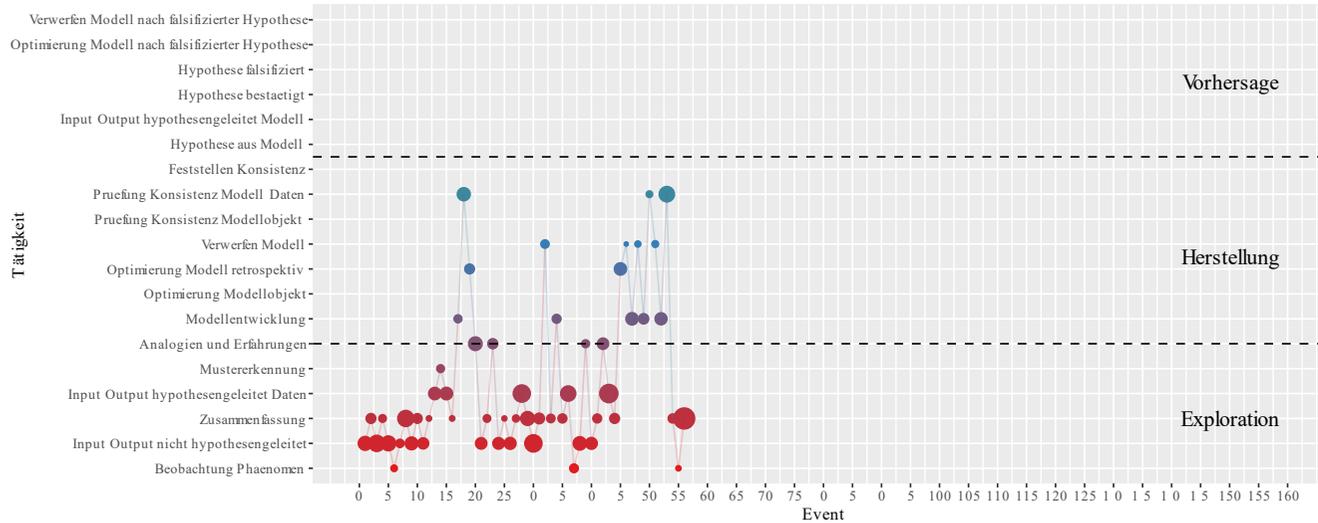


# Anhang

Codeline der Tätigkeitssequenz von Jasmin

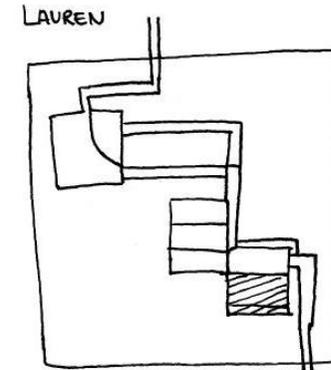
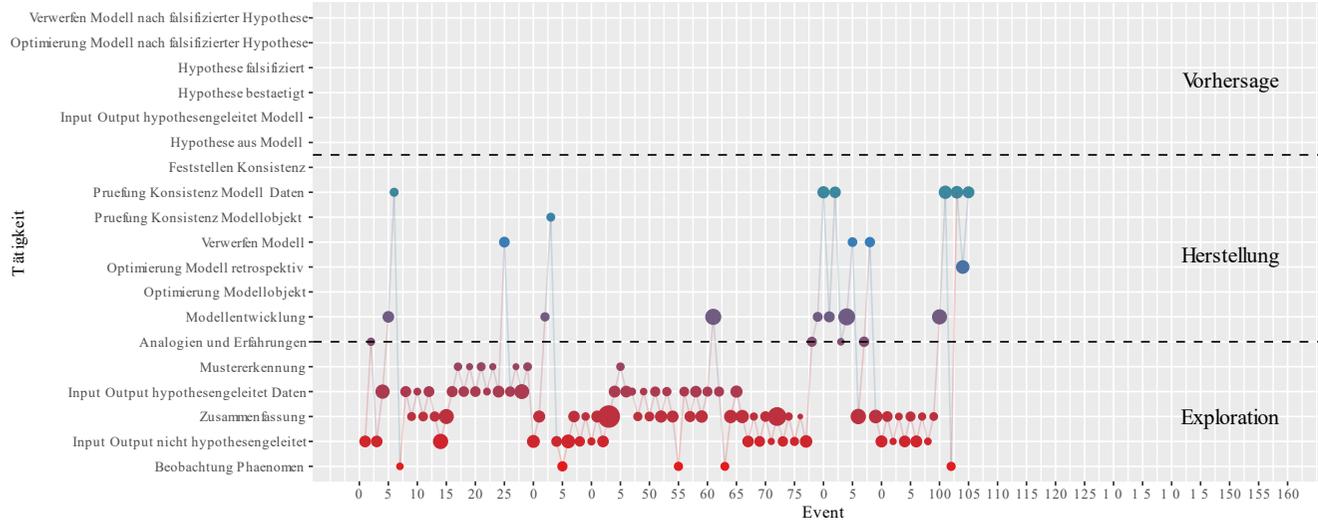


Codeline der Tätigkeitssequenz von Jim

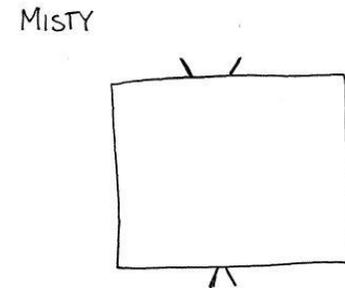
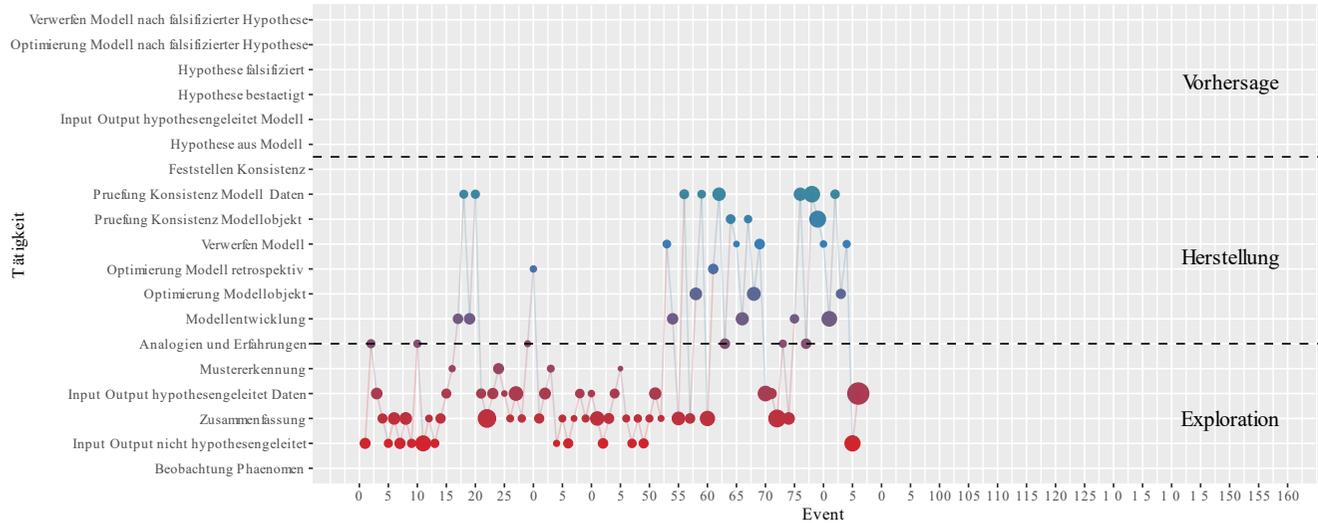


# Anhang

Codeline der Tätigkeitssequenz von Lauren

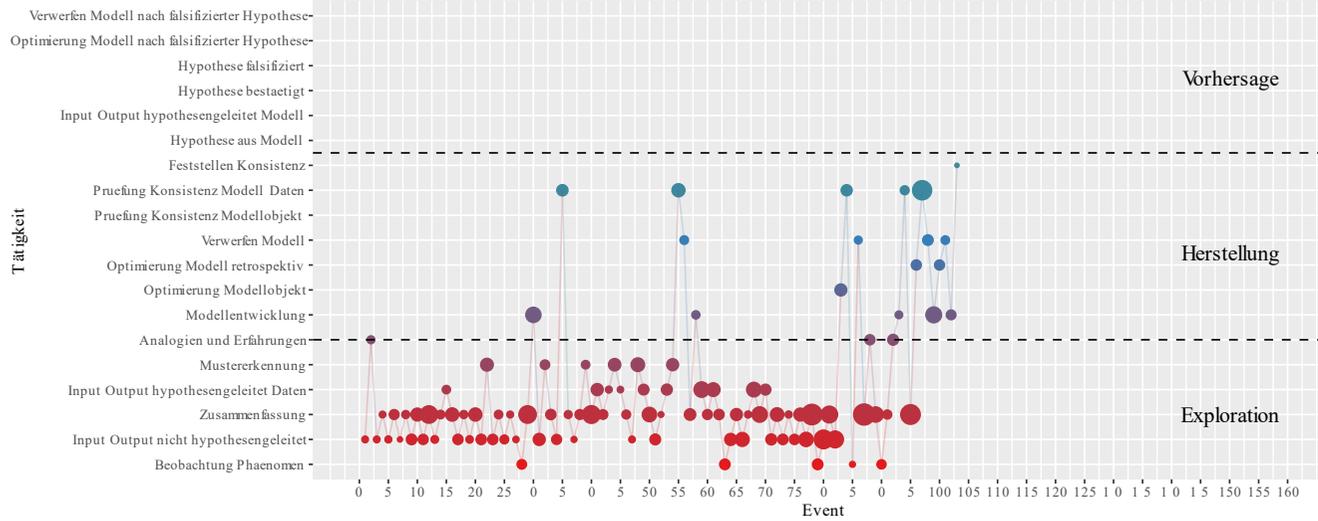


Codeline der Tätigkeitssequenz von Misty

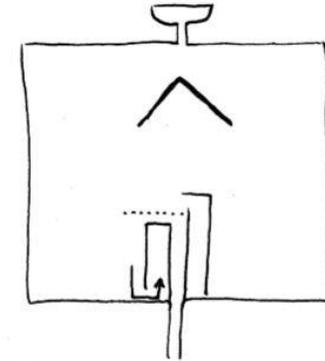


# Anhang

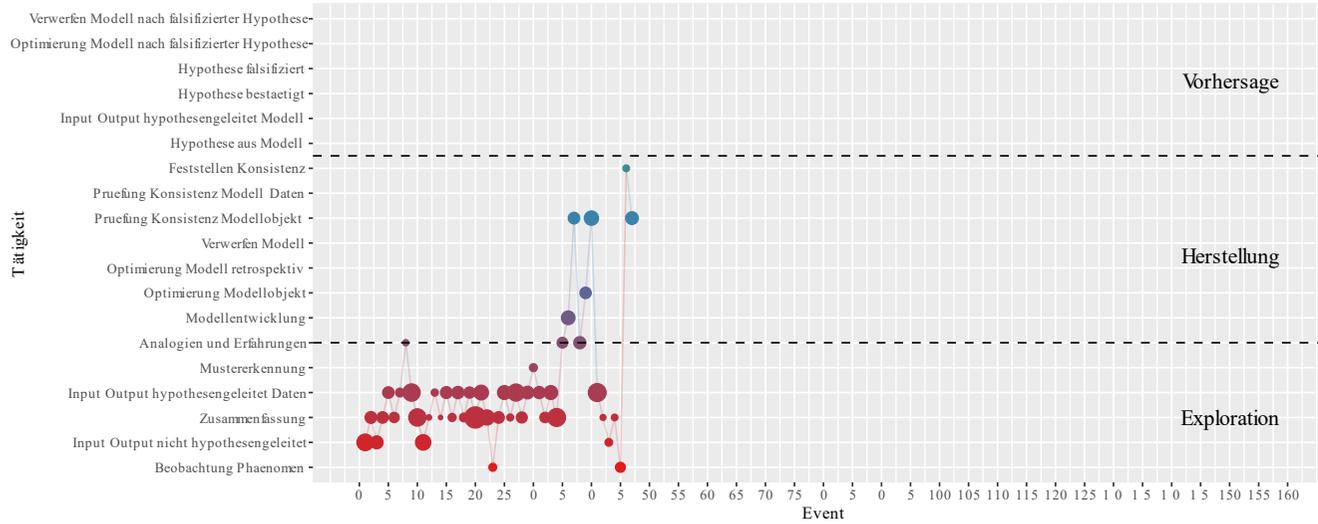
Codeline der Tätigkeitssequenz von Natalie



NATALIE



Codeline der Tätigkeitssequenz von Selena



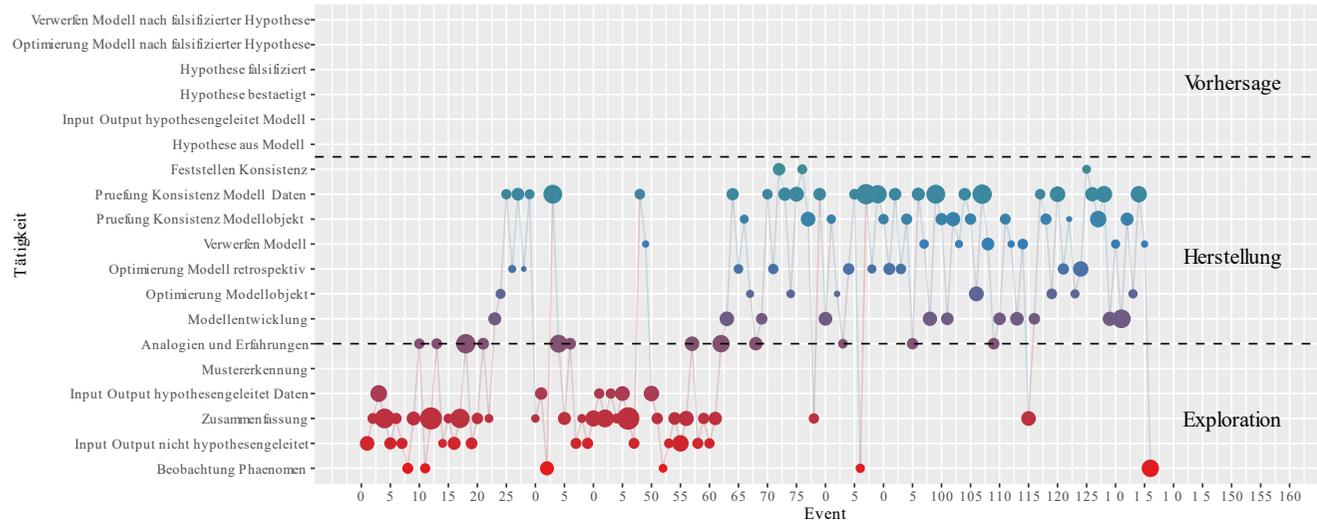
SELENA



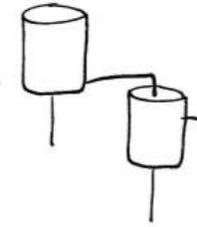
# Anhang

## Typ 2b

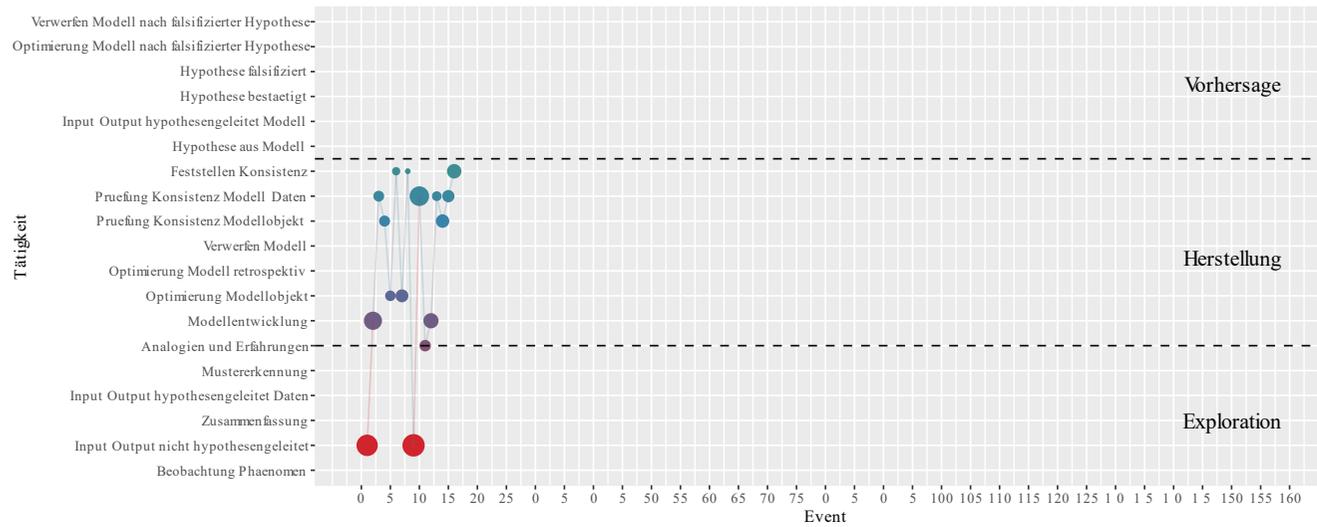
Codeline der Tätigkeitssequenz von Donna



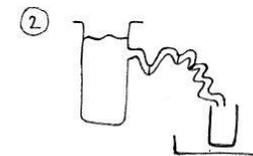
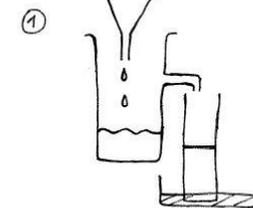
DONNA



Codeline der Tätigkeitssequenz von Kara



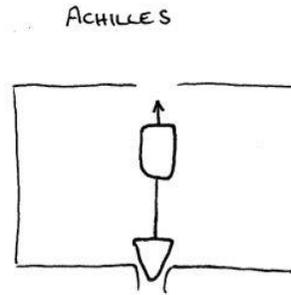
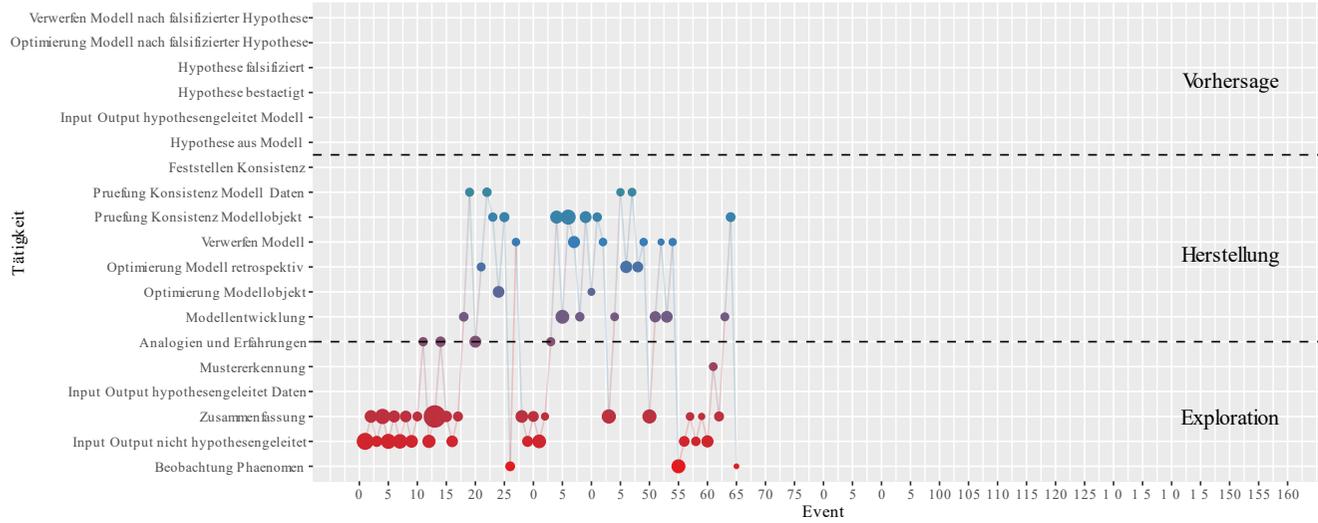
KARA



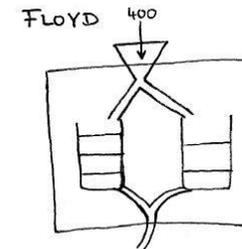
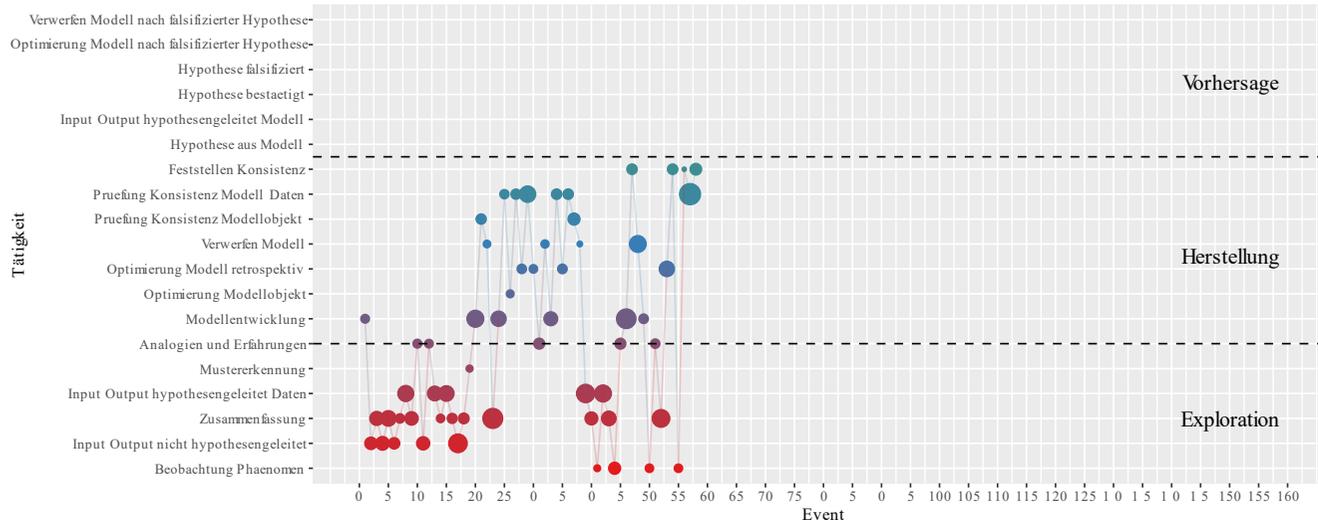
# Anhang

## Typ 2c

Codeline der Tätigkeitssequenz von Achilles

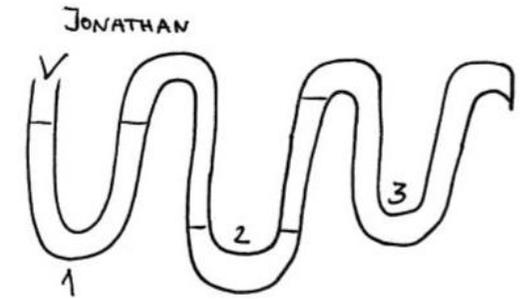
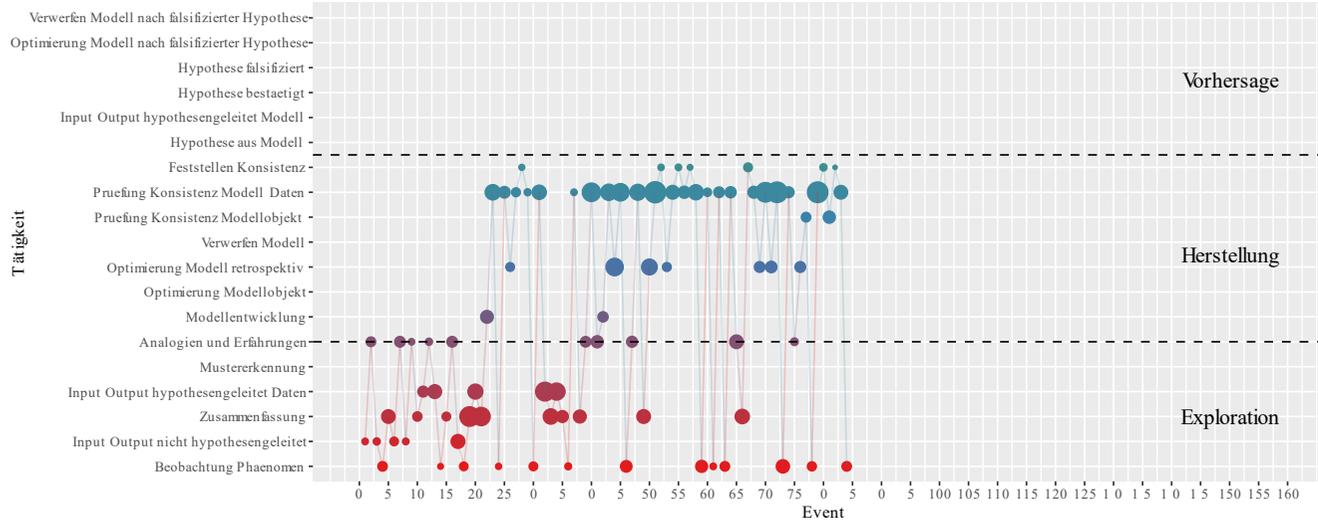


Codeline der Tätigkeitssequenz von Floyd

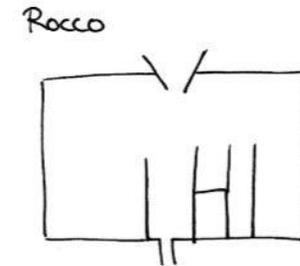
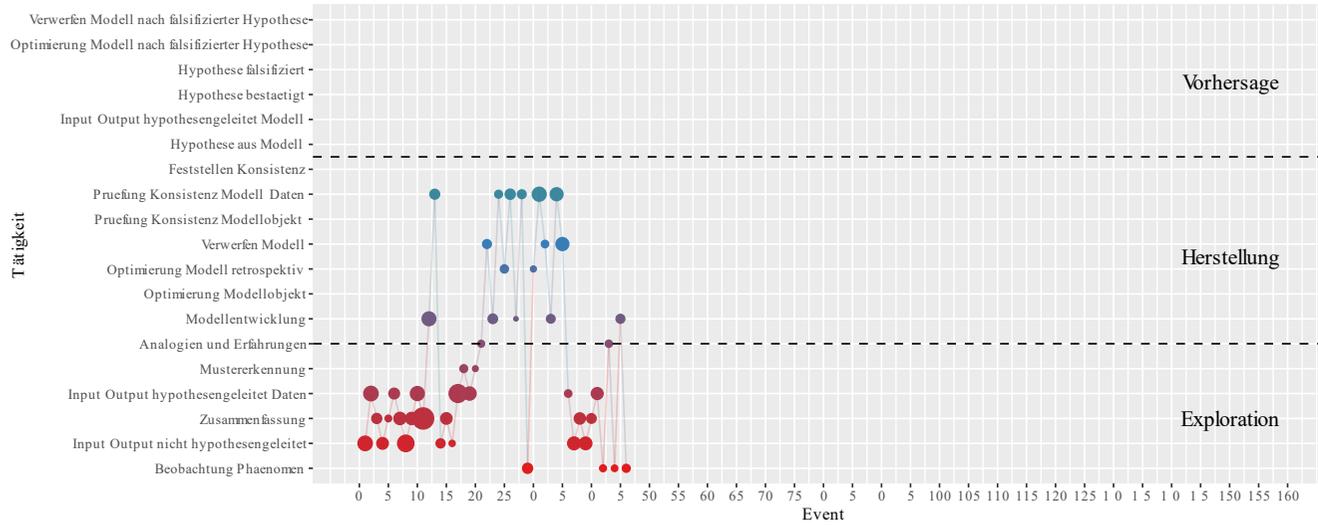


# Anhang

Codeline der Tätigkeitssequenz von Jonathan

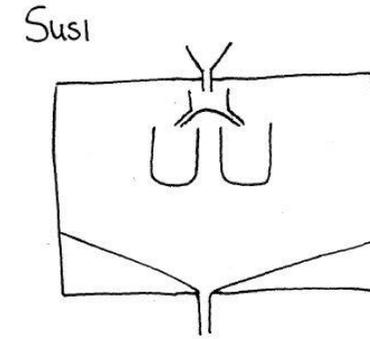
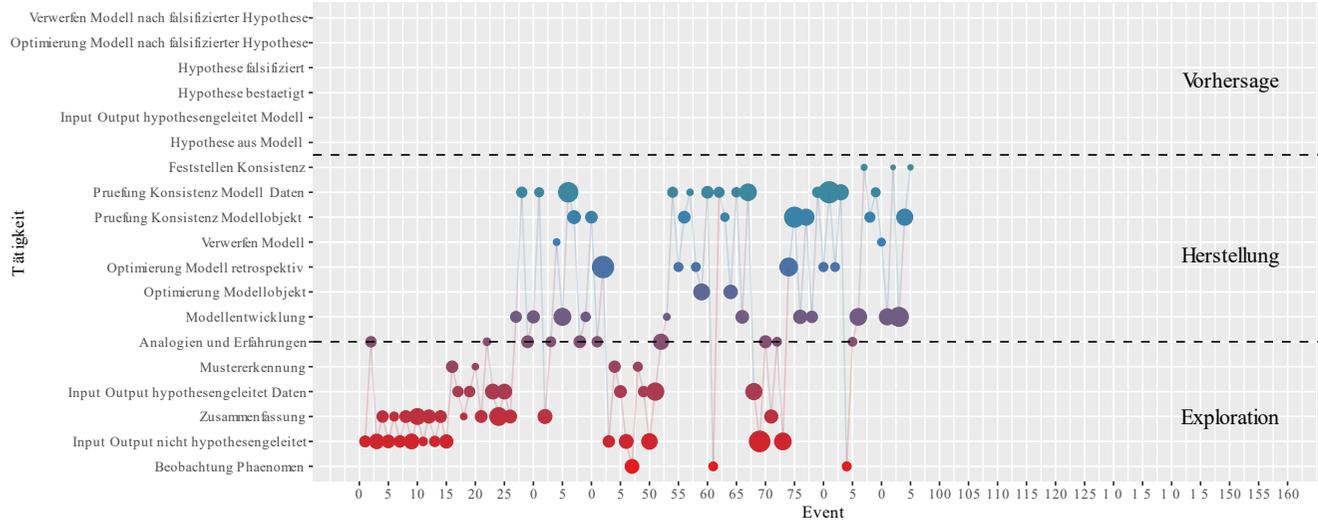


Codeline der Tätigkeitssequenz von Rocco

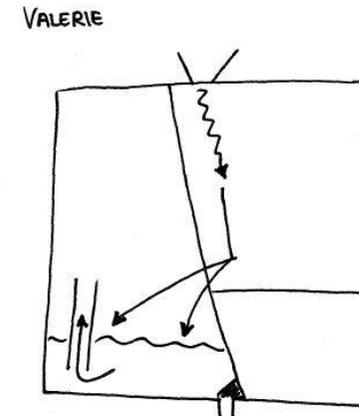
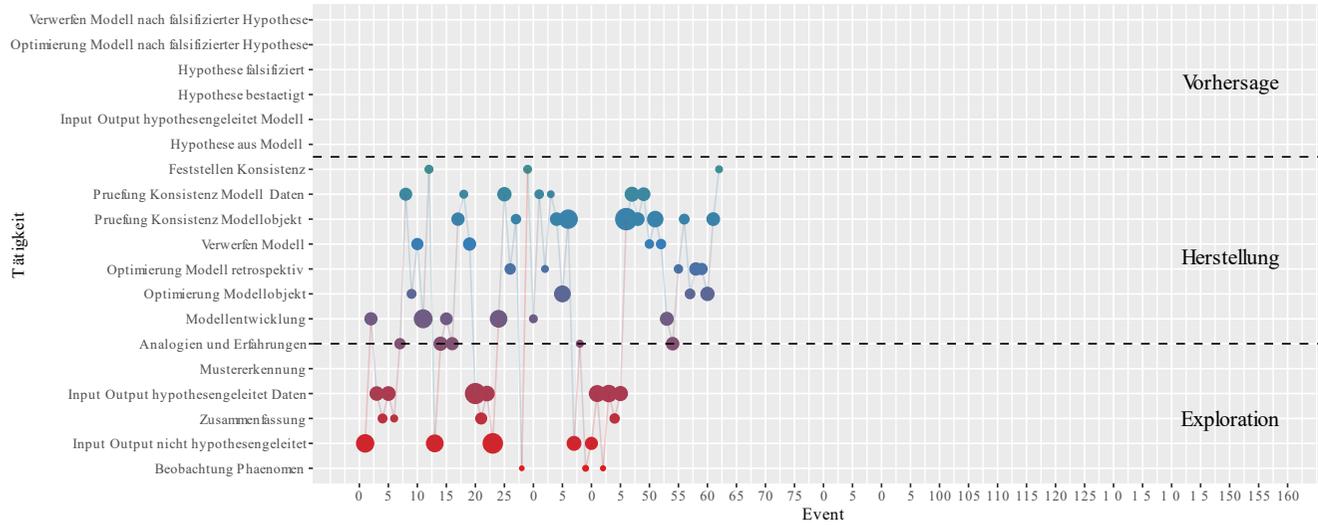


# Anhang

Codeline der Tätigkeitssequenz von Susi



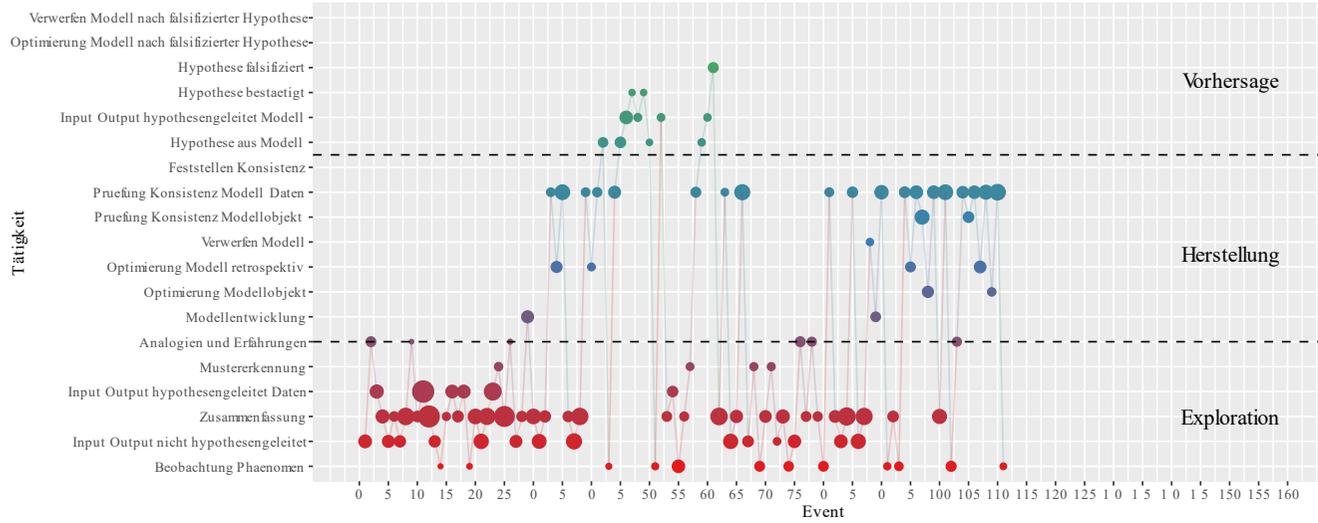
Codeline der Tätigkeitssequenz von Valerie



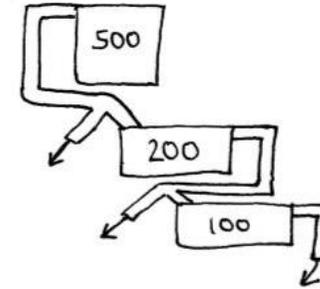
# Anhang

## Typ 3a

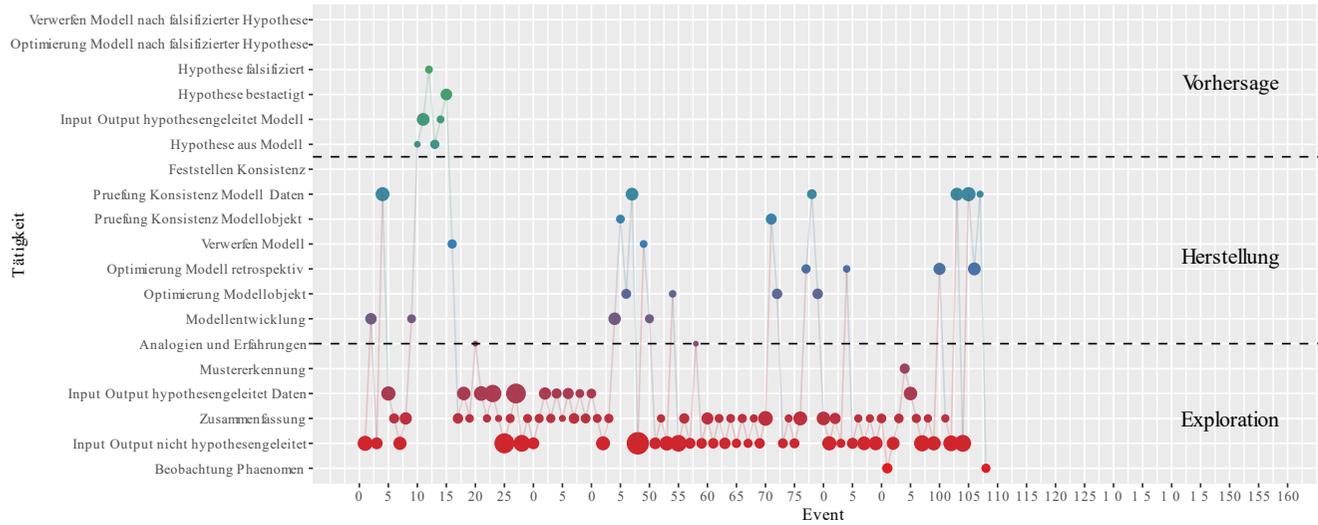
Codeline der Tätigkeitssequenz von Alice



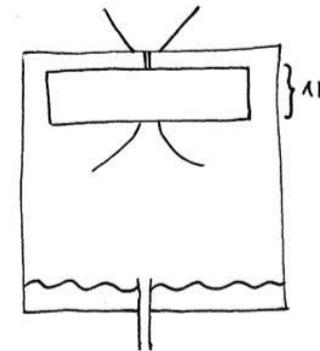
ALICE



Codeline der Tätigkeitssequenz von Ben

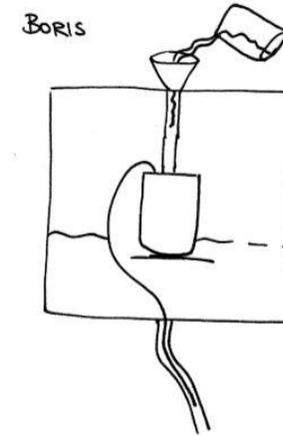
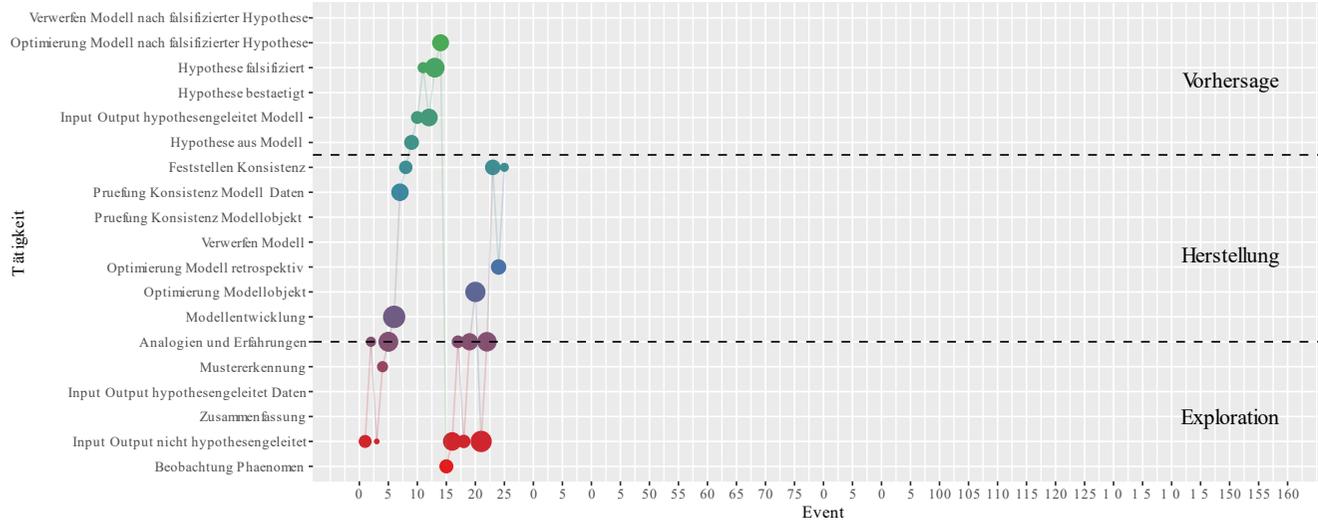


BEN

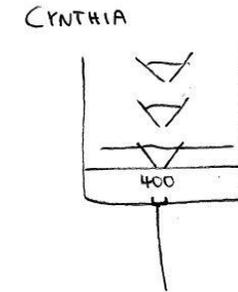
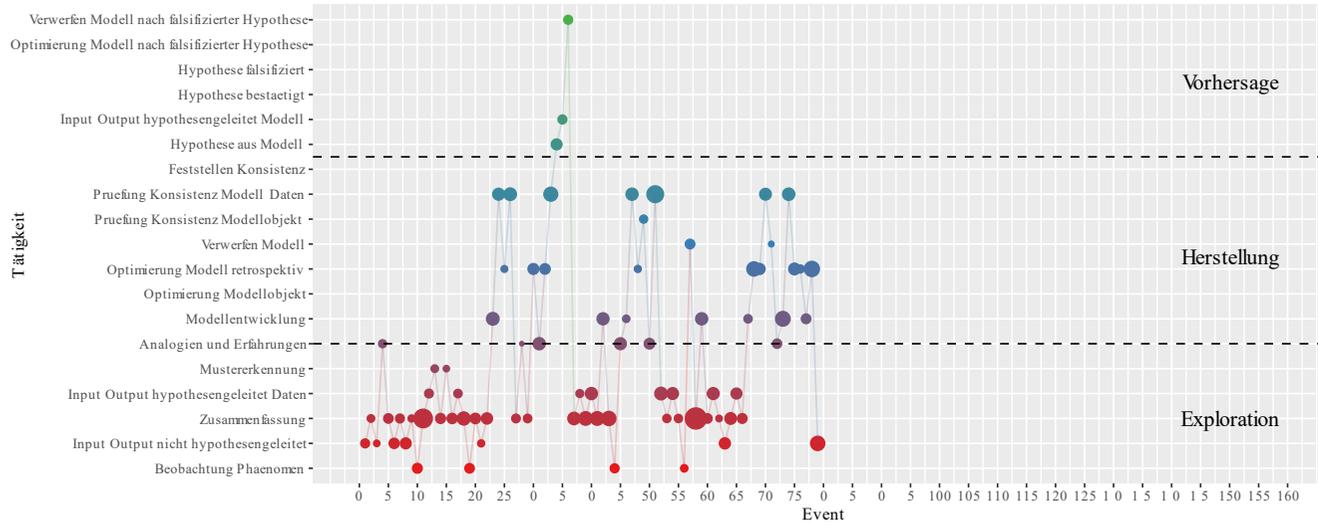


# Anhang

Codeline der Tätigkeitssequenz von Boris

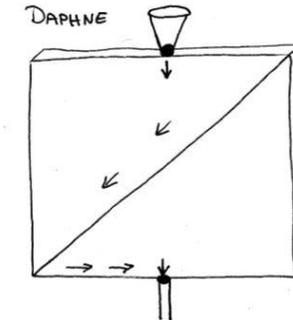
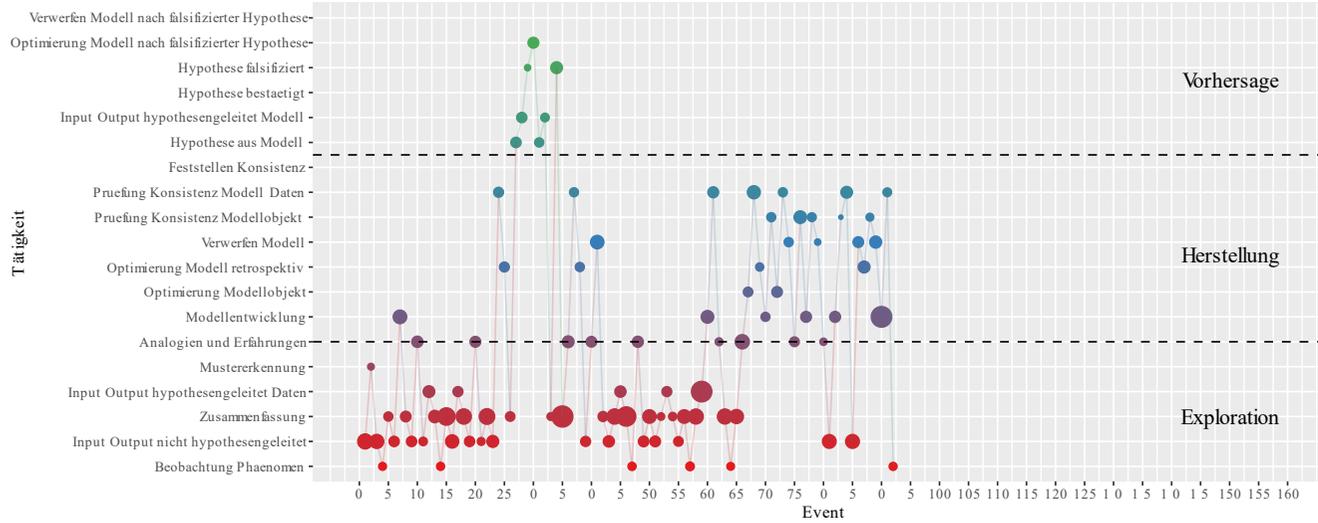


Codeline der Tätigkeitssequenz von Cynthia

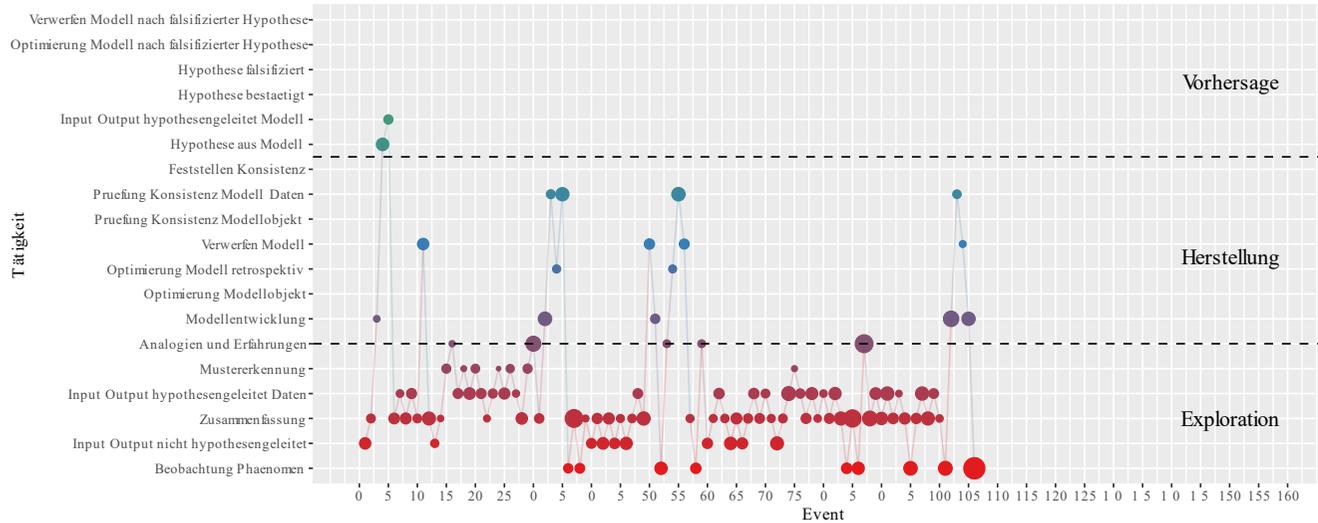


# Anhang

Codeline der Tätigkeitssequenz von Daphne



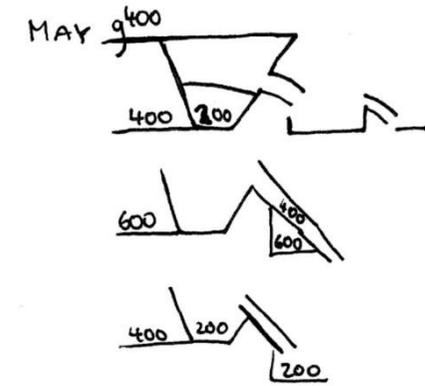
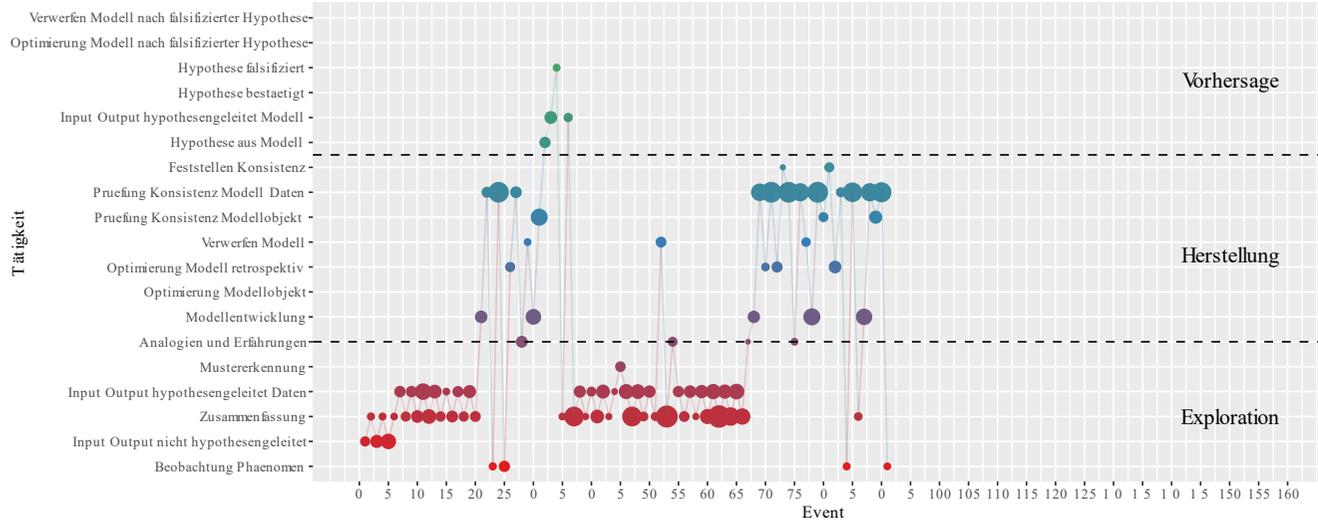
Codeline der Tätigkeitssequenz von Iris



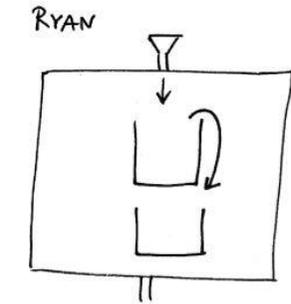
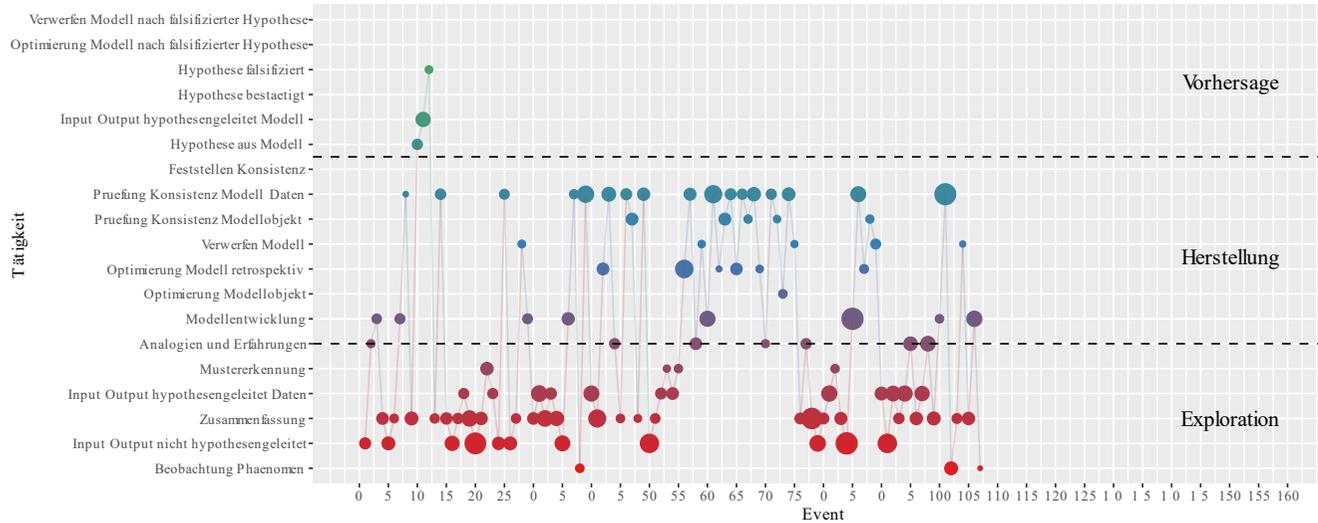
Iris verwirft ihr finales Modellierprodukt

# Anhang

Codeline der Tätigkeitssequenz von May

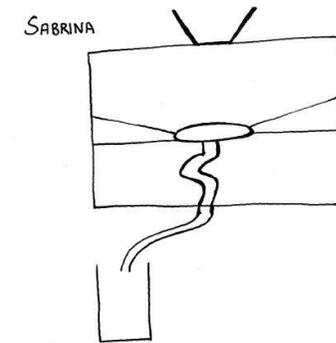
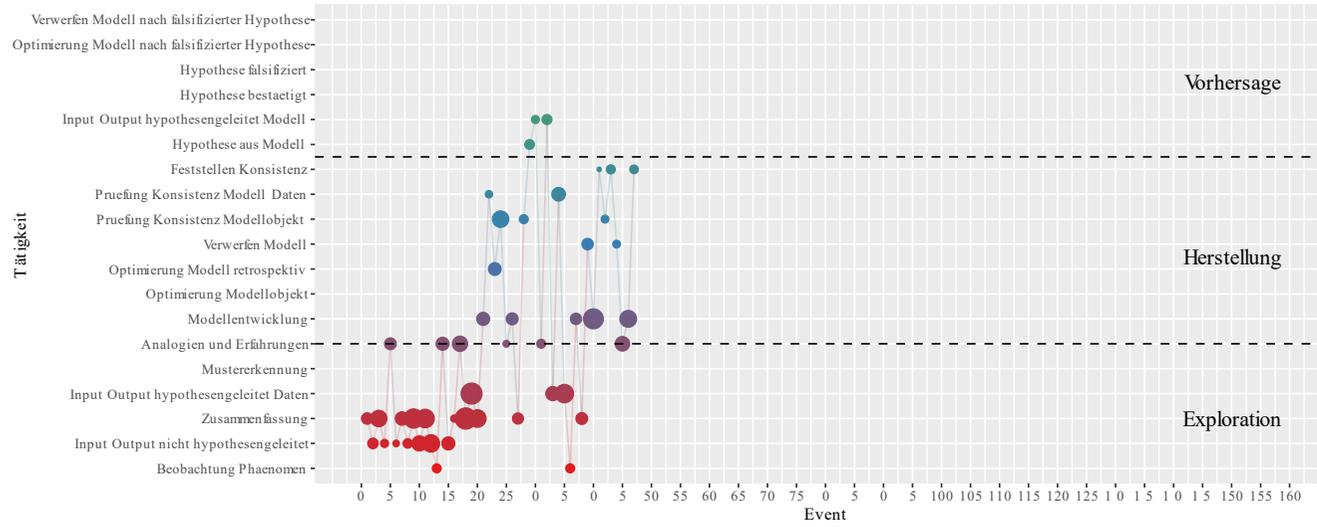


Codeline der Tätigkeitssequenz von Ryan



# Anhang

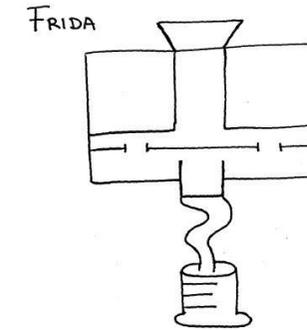
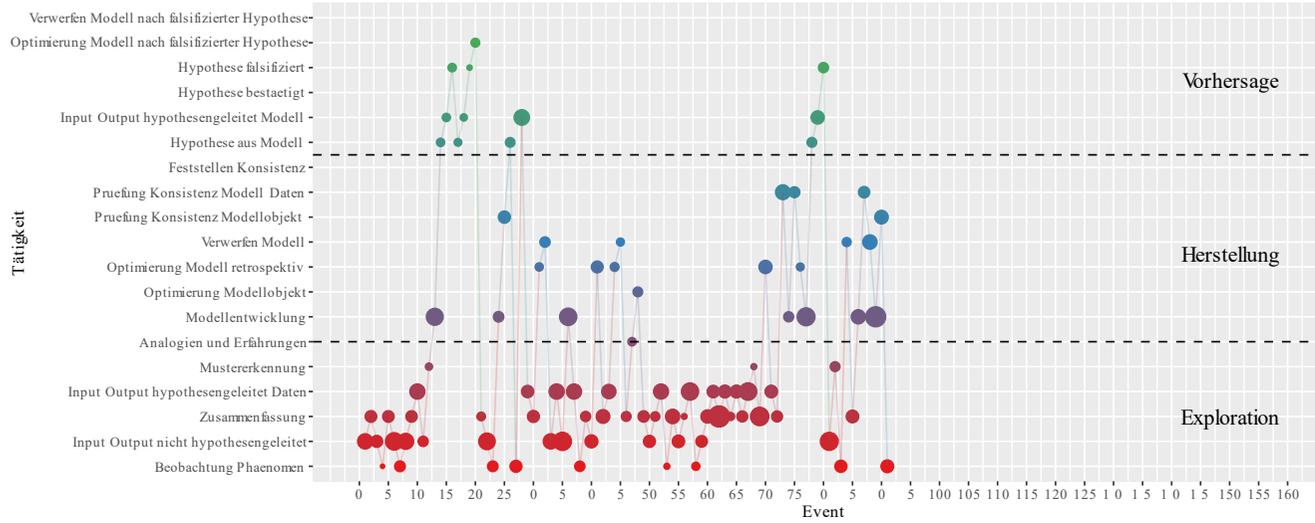
Codeline der Tätigkeitssequenz von Sabrina



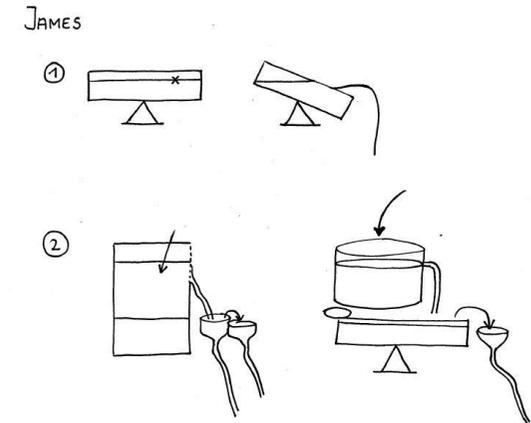
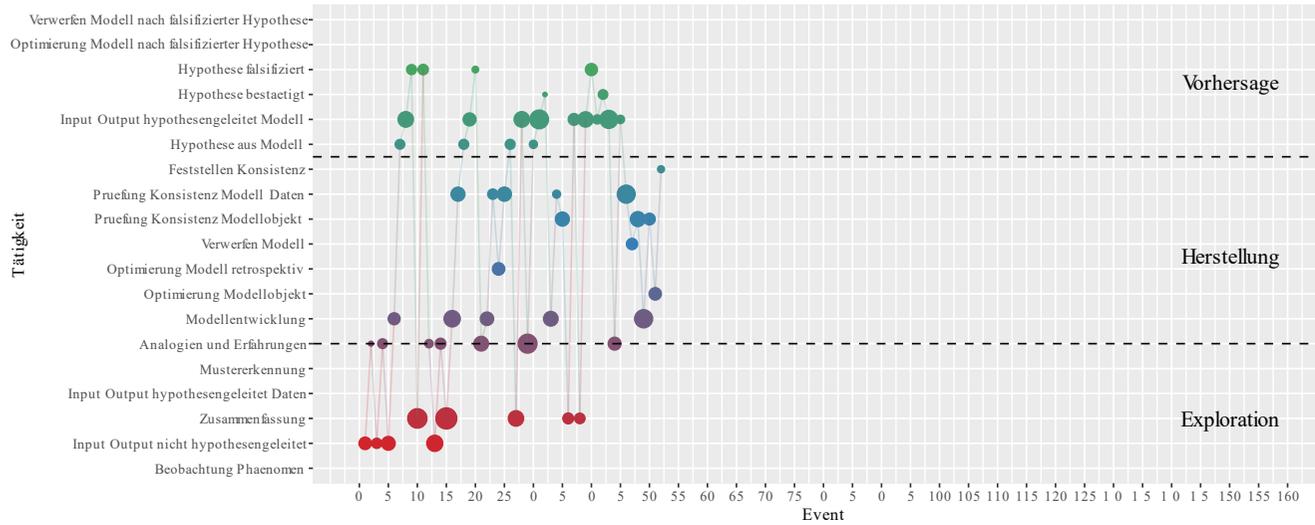
# Anhang

## Typ 3b

Codeline der Tätigkeitssequenz von Frida

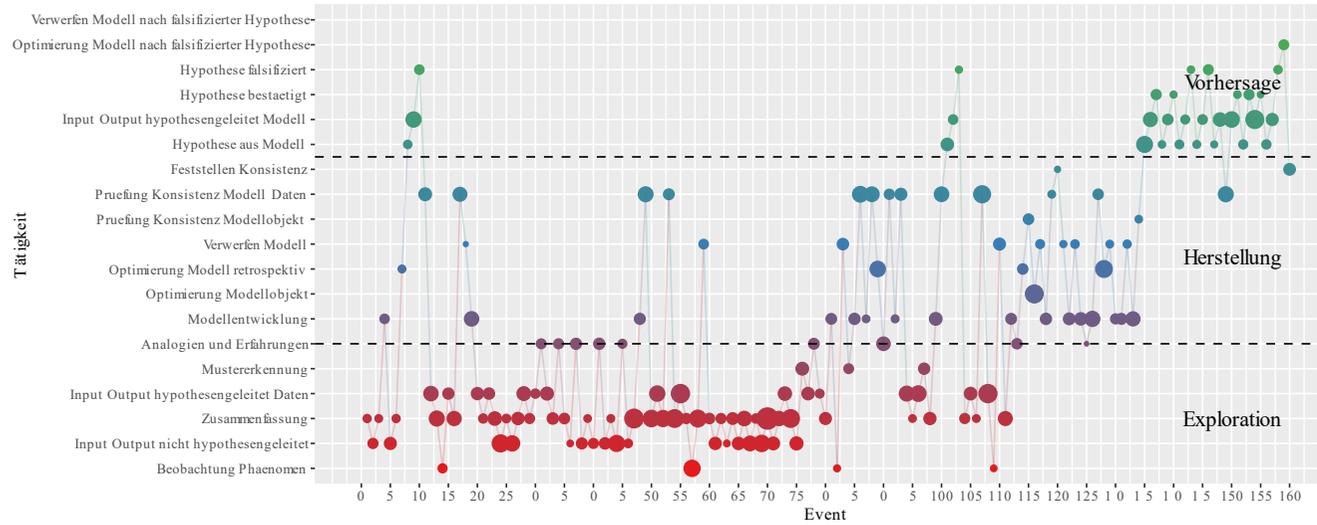


Codeline der Tätigkeitssequenz von James

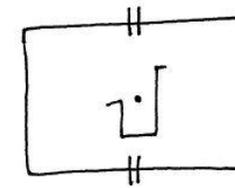


# Anhang

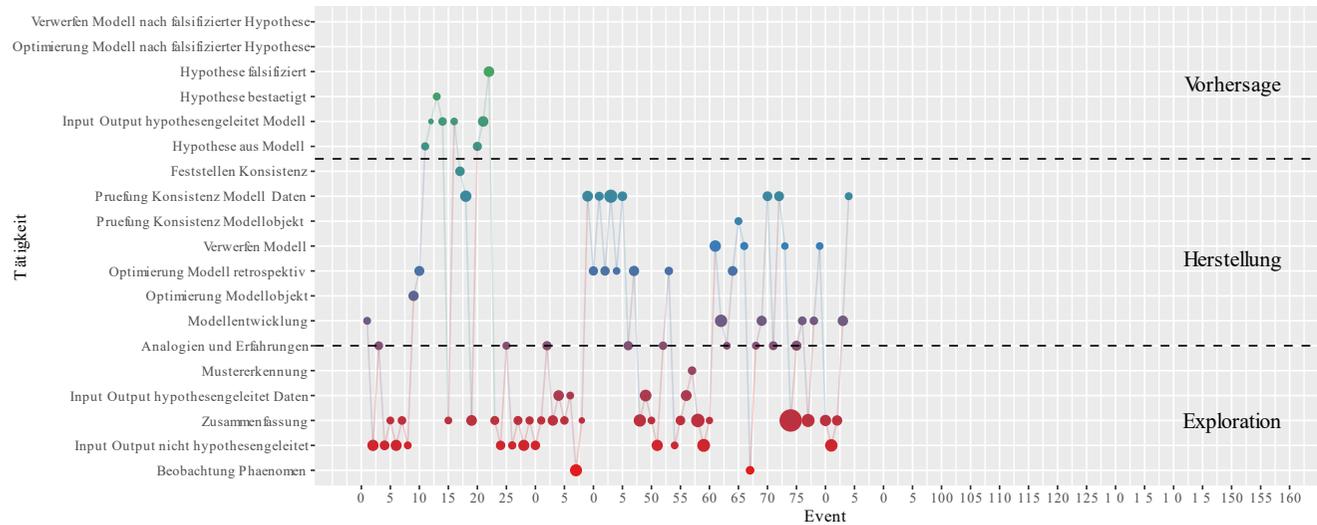
Codeline der Tätigkeitssequenz von Jenny



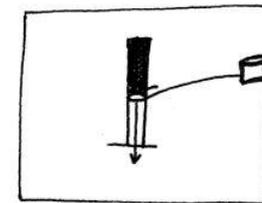
JENNY



Codeline der Tätigkeitssequenz von Lana

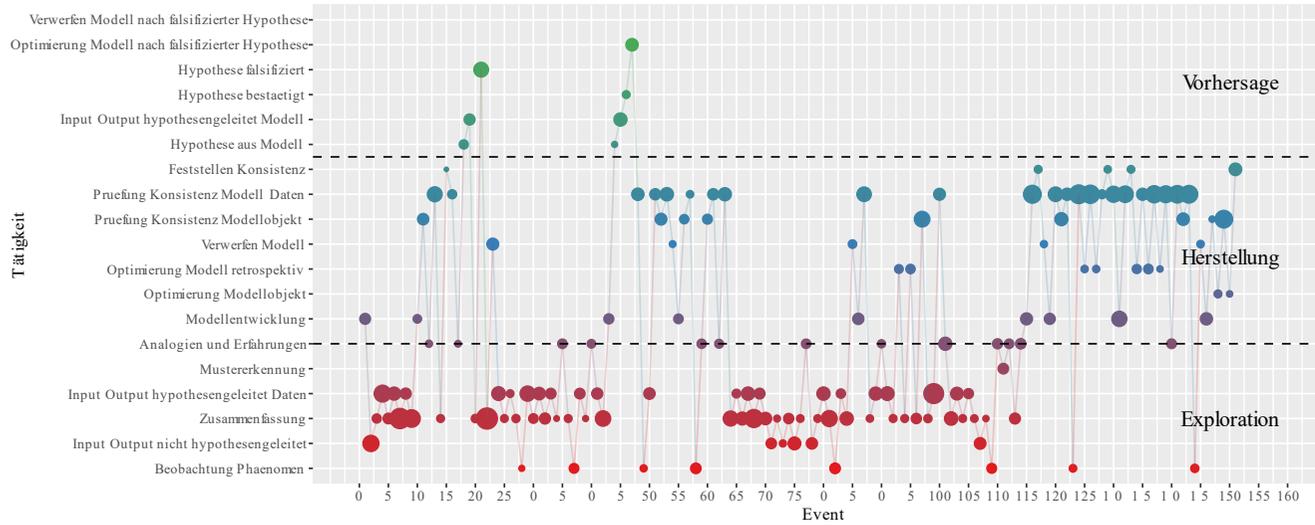


LANA

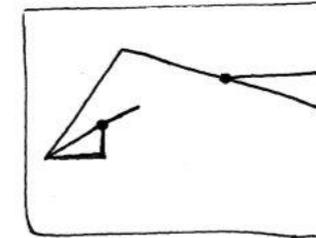


# Anhang

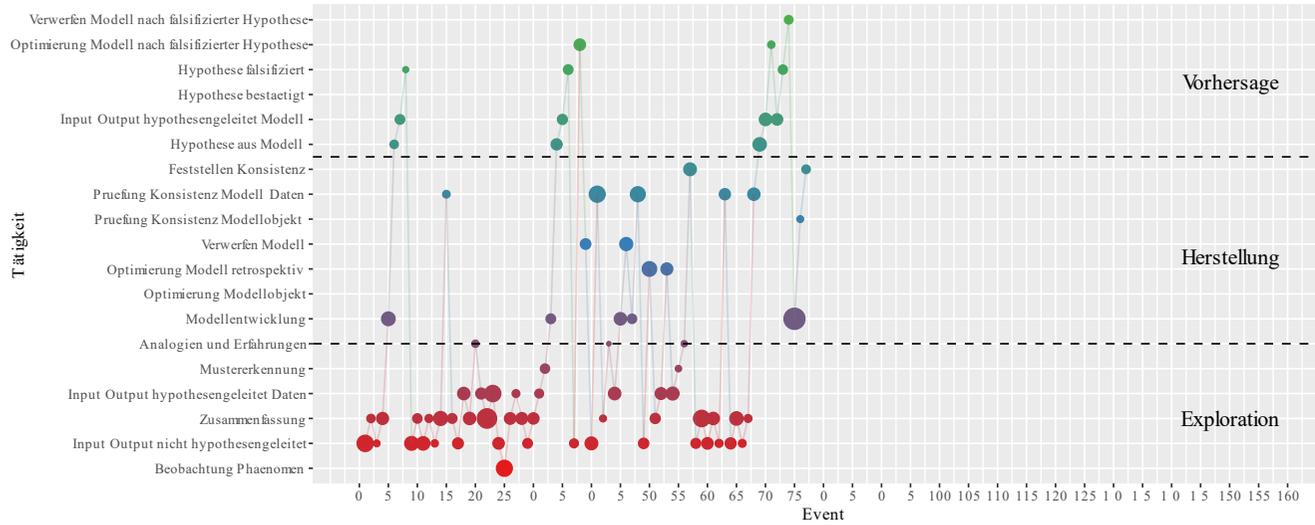
Codeline der Tätigkeitssequenz von Martin



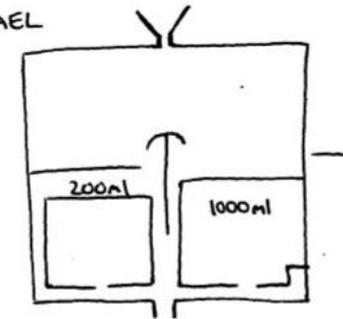
MARTIN



Codeline der Tätigkeitssequenz von Raphael



RAPHAEL



## Anhang

### 9.3.2 Forschungseinwilligung

Liebe Studentin, lieber Student,

das Projekt *TypMoL* wird von der Deutschen Forschungsgemeinschaft gefördert und von der Arbeitsgruppe Didaktik der Biologie an der Freien Universität Berlin durchgeführt. Ziel des Projektes ist die Beschreibung von Modellierungsstrategien Biologie-Lehramtsstudierender. Langfristig sollen hierauf aufbauend spezifische Reflexionsangebote für die Lehramtsausbildung entwickelt werden.

Was wird gefilmt?

Wir möchten Sie bei der Untersuchung einer Blackbox und der zeichnerischen Entwicklung eines Modells des Inneren der Blackbox filmen. Es werden sowohl Bild- als auch Tonaufnahmen angefertigt. Es ist nicht auszuschließen, dass Ihr Name in den Aufnahmen hörbar ist. Die Aufzeichnung erfolgt in Räumen der Freien Universität Berlin.

Was wird mit den Videoaufnahmen gemacht?

Die Videos werden auf einem passwortgeschützten Laufwerk dauerhaft gespeichert und an der Freien Universität Berlin aufbewahrt. Die Videos werden im Rahmen des Projektes *TypMoL* ausgewertet. Zum Zweck der Präsentation des Projektes werden die Videos und/ oder Standbilder auf öffentlich beschränkten Veranstaltungen für Fachpublikum (z. B. Fachkongresse) gezeigt und zur Anfertigung schriftlicher Veröffentlichungen (z. B. Publikationen in Fachzeitschriften) verwendet.

Wie wird der Datenschutz gewährleistet?

Die Teilnahme an der Aufnahme ist freiwillig. Es entstehen Ihnen keinerlei Nachteile, wenn Sie Ihr Einverständnis zur Teilnahme an den Videoaufnahmen nicht erteilen. Sie können eine erteilte Einverständniserklärung innerhalb von sechs Wochen nach den durchgeführten Videoaufnahmen schriftlich widerrufen. Gesichter, die auf Standbildern zu sehen sind, werden vor einer Veröffentlichung unkenntlich gemacht (z. B. verpixelt). Die Auswertung und ggf. Veröffentlichung von Videos und/ oder Standbildern erfolgt pseudonymisiert (d. h. ohne Nennung Ihres Namens).

Wir würden uns sehr über Ihre Unterstützung freuen! Für Rückfragen zum Projekt stehen wir Ihnen gerne zur Verfügung.

Mit freundlichen Grüßen

(Maximilian Göhner)

Ihre Einverständniserklärung:

(Name, Datum)

(Unterschrift)

## Anhang

### 9.3.3 Hinweise zur Durchführung der Studie

#### Studienbeginn

- ca. eine Stunde vor Beginn vor Ort sein und den Raum vorbereiten:
- drei Kameras aufbauen und ausrichten
- Blackbox, Gefäße und Wasser bereitstellen [ein Gefäß mit ca. 405ml Wasser füllen]
- Tafel säubern und Impuls anbringen
- Kreide & Schwamm
- Stuhl für Studienleiter
- kurz vor Eintreffen der Probanden Kameras anstellen

#### Während der Studie

- Sachen der Probanden (Jacken etc.) im Nebenraum (R 3.02) deponieren
- in R 3.02 kurze Einführung geben:
- „Ich danke dir für die Bereitschaft, mich heute bei meiner Studie zu unterstützen. Ich werde dir gleich eine sogenannte Blackbox zeigen, deren Verhalten nicht direkt verständlich ist, vielleicht sogar etwas paradox wirkt. Deine Aufgabe wird sein, eine mögliche Erklärung dafür zu finden. Du hast dafür so lange Zeit, wie du haben möchtest. Zunächst bitte ich dich aber darum, diesen Fragebogen zu beantworten.“

[Fragebogen zu Modellen aushändigen.]

- „Vielen Dank! Wir gehen gleich zusammen in den Nebenraum. Dort stehen drei Kameras, damit ich mir später evtl. erneut ansehen kann, wie das Ganze abgelaufen ist. Ergänzend bitte ich dich, dieses Aufnahmegerät zu tragen. [Gerät anlegen.] Damit ich später besser verstehe, warum du bestimmte Handlungen ausgeführt hast, bitte ich dich, alle Überlegungen, die dir während der Arbeit mit der Blackbox durch den Kopf gehen, laut auszusprechen. Dabei ist es wichtig, dass du nicht versuchst, zu erklären oder zu strukturieren, was du tust. Stell dir einfach vor, du wärst ganz allein im Raum und sprichst mir dir selbst. Wichtig ist, dass du möglichst immer redest. Wir üben dieses „laute Denken“ gleich noch an drei kleinen Beispielen.“
- „Ich benutze die Aufnahmen nur für meine Arbeit und werte alles anonymisiert aus. Ich werde im Raum anwesend sein um zu schauen, dass die Technik läuft und dich ggf. an die Aufgabe zum „lauten Denken“ erinnern. Ich stehe aber nicht für sonstige Nachfragen zur Verfügung. Deine Aufgabe ist es, mit Hilfe der bereitgestellten Materialien das Verhalten der Blackbox zu untersuchen und ein Modell des Inneren der Blackbox zu entwickeln und fortlaufend auf einer Tafel zu dokumentieren. Die einzige Einschränkung ist: Du darfst die Blackbox nicht öffnen.“
- „Hast du noch Fragen?“
- Übungsaufgaben zum lauten Denken
  - Aufgabe 1 Zählen Sie, wie viele Fenster in Ihrer Wohnung sind. ● ● ●
  - Aufgabe 2 Beschreiben Sie den Weg von der Eingangstür dieses Institutsgebäudes bis zur Tür des Raumes, in dem wir uns momentan befinden. ● ● ●
  - Aufgabe 3 Verbinden Sie die neun Punkte rechts mit vier geraden Linien, ohne dabei den Stift abzusetzen. [Blatt & Stift vorlegen.] ● ● ●

## Anhang

- gemeinsam den Raum betreten
- Einführung der Blackbox:
- „Hier ist die Blackbox. Du kannst in die Blackbox Wasser schütten und beobachten, was passiert. Ich empfehle, mit dem mehrmaligen eingießen von 400ml zu beginnen, um einen ersten Eindruck der Blackbox zu bekommen.“
- Studienleiter nimmt auf dem vorgesehenen Stuhl Platz
- bei Sprechpausen von mehr als 10 Sekunden an Aufgabe zum lauten Denken erinnern: „Bitte sprich weiterhin!“ „Erzähle weiterhin, was du denkst!“ „Bitte vergiss nicht, deine Gedanken zu äußern!“

### Studienende

- bei den Probanden für die Mitarbeit bedanken
- Raum aufräumen
- Daten auf einem PC sichern und benennen: JahrMonatTag\_KAMERA (z. B. 13. Dez. 2013: 20131213\_Übersicht, 20131213\_Tafel, 20131213\_Blackbox) bzw. JahrMonatTag\_Audio

## Anhang

### 9.3.4 Kodierleitfaden Modellverstehen

#### Grundsätzliche Kodierhinweise

- Wenn die Lehrer/-innen unterschiedliche Niveaus zu einer Teilkompetenz zum Ausdruck bringen, nur das höhere Niveau notieren.
- Bei Antworten, die sich nicht in eines der drei Niveaus der jeweiligen Teilkompetenz einordnen lassen, eine 888 notieren.
- Bei Antworten, die eine Abhängigkeit vom einzelnen Modell betonen und sonst keine Beschreibung liefern (z. B. „Kommt ganz auf das Modell an.“), ebenfalls eine 888 notieren.
- Bei Antworten, die sich nur mit großem Interpretationsaufwand von (implizit) enthaltenen Ideen einordnen lassen, auch eine 888 notieren.
- Die alleinige Erwähnung einzelner Schlüsselwörter (z. B. Fragestellung, Erkenntnisgewinnung) reicht nicht aus, wenn nicht klar ist, ob diese Begriffe didaktisch oder wissenschaftlich gemeint sind.
- Bei gar keiner Antwort, eine 999 notieren.

Eigenschaften von Modellen		
Frage im Testheft: Inwieweit entspricht ein Modell Ihrer Meinung nach seinem biologischen Ausgangsobjekt?		
Niveau I Modelle sind Kopien von etwas	Niveau II Modelle sind idealisierte Repräsentationen von etwas	Niveau III Modelle sind theoretische Rekonstruktionen von etwas
<p><b>KATEGORIEN</b>                      Modell als Kopie                      gleicht dem Original (vergrößert/verkleinert) Kopie des Originals als maßstabsgetreues Duplikat des Originals akzeptiert, weil ein großes Vertrauen gegenüber Wissenschaft, wissenschaftlichen Methoden oder Wissenschaftlern besteht                      Modell mit sehr großer Ähnlichkeit                      (nahezu) originalgetreues Duplikat des Originals; Defizite in der Modellerstellung führen zu Abweichungen; Modell als Kopie/Duplikat wäre optimal                      Modell entspricht (nicht) subjektiver Vorstellung vom Original                      vergleichen und beurteilen das Modell nach eigenem Vorwissen, persönlichen Erfahrungen oder subjektiven Vorstellungen über das Original</p> <p><b>SCHLÜSSELIDEEN</b>                      Ein Modell ist (nahezu) identisch, sehr ähnlich, besitzt zum Original nur Unterschiede im Maßstab, ...</p>	<p><b>KATEGORIEN</b>                      Modell ist in Teilen eine Kopie                      gleicht nur in bestimmten Merkmalen dem Original, andere Merkmale können bei geringer Information/ Kenntnis über das Original nicht beurteilt werden                      Modell als eine mögliche Variante                      gleicht möglicherweise (nicht) dem Original, abstrakte Aussage zu übereinstimmenden Merkmalen                      eine denkbare Version unter vielen; Gründe für Möglichkeit verschiedener Versionen werden nicht/wenig ausgeführt                      Modell als fokussierte Darstellung                      fokussiert auf einen Ausschnitt des Originals, hebt bestimmte Merkmale/Eigenschaften hervor (z. B. Struktur, Funktion, 2D D, ...)</p> <p><b>SCHLÜSSELIDEEN</b>                      Ein Modell ist idealisiert, vereinfacht, zeigt eingeschränkte, teilweise Gleichheit, zeigt Gleichheit nur in Bezug auf wenige/wesentliche Aspekte/Schwerpunkte, ...</p>	<p><b>KATEGORIEN</b>                      Modell als hypothetische Darstellung                      stellt eine Hypothese über das Original dar                      es wird ausgeführt/ diskutiert, dass Modelle stets auf Annahmen/ Theorien beruhen</p> <p><b>SCHLÜSSELIDEEN</b>                      Ein Modell ist: Gedankenmodell, Hypothese, Theorie, Rekonstruktion, ...</p>

## Anhang

Alternative Modelle		
Frage im Testheft: Aus welchen Gründen gibt es Ihrer Meinung nach zu einem biologischen Ausgangsobjekt verschiedene Modelle? (Bitte nennen Sie wenigstens zwei Gründe.)		
Bei didaktischen Perspektiven, in denen die Existenz verschiedener Modelle eines Originals lediglich mit unterrichtsmethodischen Überlegungen begründet wird (z. B. Binnendifferenzierung), eine 888 notieren.		
Niveau I Unterschiede zwischen den Modellobjekten	Niveau II Ausgangsobjekt ermöglicht Herstellung verschiedener Modelle von etwas	Niveau III Modelle für verschiedene Hypothesen
<p><b>KATEGORIEN</b>            Modelle sind gleich            Modelle sind alle gleich/ keine Beschreibung von Unterschieden zwischen den Modellen            Verschiedene Modelle zu unterschiedlichen Originalen            jedes Modell repräsentiert ein anderes Original            Nur ein endgültiges und richtiges Modell            nur eines der alternativen Modelle ist korrekt, die anderen sind falsch            nur eines der alternativen Modell ist das endgültige Modell; sind nicht zeitgleich gültig weil sich Modelle im Laufe der Geschichte/ aufgrund neuer Erkenntnisse ändern            Unterschiedliche Modellobjekteigenschaften            verschiedene Darstellungsweisen (u. a. 2D/3D, verschiedene Farben)            verschiedene Modellmerkmale (u. a. beweglich/ unbeweglich, weich/ hart)            unterschiedliche Konstruktionsmöglichkeiten (u. a. dünne/ dicke Materialien, getrennte/ nicht getrennte Bausteine)</p> <p><b>SCHLÜSSELIDEEN</b>            Es gibt verschiedene Modelle von verschiedenen Originalen, wegen verschiedener Materialien, unterschiedlicher Hersteller, unterschiedlich teurer Herstellung, historischer Weiterentwicklung in der Wissenschaft... aber <i>eigentlich</i> gibt es nur ein Modell zu einem Original.</p>	<p><b>KATEGORIEN</b>            Unterschiedliche inhaltliche Schwerpunkte            Komplexität des Originals ermöglicht unterschiedliche Perspektiven bzw. Foci auf das Original (u. a. innen/ außen, Längs- / Querschnitt, Struktur/ Funktion, verschiedene Ausschnitte bzw. Zustände des Originals)</p> <p><b>SCHLÜSSELIDEEN</b>            Es gibt verschiedene Modelle wegen der Komplexität des Originals, Schwerpunktsetzungen (z. B. auf Bau bzw. Funktion, Struktur bzw. Funktion), Perspektiven Blickwinkel auf das Original, ...</p>	<p><b>KATEGORIEN</b>            Unterschiedliche Annahmen            verschiedene Annahmen/ Ideen über das Original/ unterschiedliche Modelle sind alle zeitgleich gültig            unterschiedliche Interpretationen der Daten            Unterschiedliche Annahmen mit wissenschaftlicher Anwendungsperspektive            verschiedene Annahmen über das Original; wissenschaftlicher Nutzen wird genannt (z. B. Diskussionsgrundlage, Vergleich verschiedener Annahmen, Überprüfung der Annahmen mit den Modellen)</p> <p><b>SCHLÜSSELIDEEN</b>            Es gibt verschiedene Modelle wegen verschiedener Vorstellungen, Hypothesen, Meinungen, Wege zu Problemlösungen, Vorhersagen, konkurrierender Positionen... die dazu führen, dass es <i>zeitgleich</i> verschiedene Modelle zu einem Original gibt.</p>

## Anhang

Zweck von Modellen		
Frage im Testheft: Welchen Zweck erfüllen Ihrer Meinung nach Modelle in der Biologie? (Bitte nennen Sie wenigstens zwei unterschiedliche Zwecke.)		
Bei Aussagen, die – ohne weitere Ausführungen – auf „lässt sich nicht pauschal beantworten“ reduzierbar sind, eine            notieren. So lange es darum geht, Strukturen Zusammenhänge Funktionen Vorgänge „nur“ zu zeigen, dominiert das Verb und es ist Niveau I. Werden hingegen Strukturen/ Zusammenhänge/ Funktionen/ Vorgänge erklärt oder nachvollzogen, ist es Niveau II.		
Niveau I Modellobjekt zur Beschreibung von etwas einsetzen	Niveau II Bekannte Zusammenhänge und Korrelationen von Variablen im Ausgangsobjekt erklären	Niveau III Zusammenhänge von Variablen für zukünftige neue Erkenntnisse voraussagen
<b>KATEGORIEN</b> Modell zum Darstellen eines Sachverhaltes stellt Sachverhalte Merkmale des Originals dar, die zu klein zu groß ... sind zeigen/ veranschaulichen von Teilen/ Lage einer Struktur oder Funktion  <b>SCHLÜSSELIDEEN</b> Der Zweck von Modellen ist Strukturen/ Funktionen/ Vorgänge zu zeigen, veranschaulichen, darstellen, vereinfachen, visualisieren, abbilden, hervorheben...	<b>KATEGORIEN</b> Modell zum Erkennen von Zusammenhängen beschreibt Zusammenhänge/ Abläufe zwischen verschiedenen Aspekten im Original und dient dazu, bekannte Tatsachen nachzuvollziehen Modell zum Erklären von Zusammenhängen beschreibt und erklärt Zusammenhänge/ Abläufe zwischen verschiedenen Aspekten im Original und dient dazu, bekannte Tatsachen nachzuvollziehen  <b>SCHLÜSSELIDEEN</b> Der Zweck von Modellen ist Zusammenhänge/ Abläufe an Strukturen oder Funktionen zu erklären nachzuvollziehen zu verstehen, ...	<b>KATEGORIEN</b> Modell zum Überprüfen von Ideen dient zur Überprüfung von Hypothesen über das Original dazu, Schlüsse über das Original zu ziehen  <b>SCHLÜSSELIDEEN</b> Der Zweck von Modellen ist vorhersagen, testen/ überprüfen von Vermutungen, durchführen von Experimenten, Ableiten von Annahmen, ...

## Anhang

Testen von Modellen		
Frage im Testheft: Wie lässt sich Ihrer Meinung nach überprüfen, ob ein biologisches Modell seinen Zweck erfüllt? (Bitte nennen Sie wenigstens zwei Vorgehensweisen.)		
Bei didaktischen Perspektiven, in denen das Überprüfen/Testen von Schüler/-innen mit Hilfe eines Modells beschrieben wird, eine 888 notieren.		
Niveau I Modellobjekt überprüfen	Niveau II Parallelsieren mit dem Ausgangsobjekt, Modell von etwas testen	Niveau III Überprüfen von Hypothesen bei der Anwendung, Modell für etwas testen
<b>KATEGORIEN</b> Keine Testung des Modells Ablehnung einer Testung des Modells an sich bzw. mit dem Modell Überprüfung des Materials Überprüfung der Widerstandsfähigkeit des Materials (u. a. Beweglichkeit, Stabilität, Elastizität, Gewicht) Überprüfung der Grundvoraussetzungen Nennung grundlegender Anforderungen für den Einsatz an das jeweilige Modell (z. B. Haltbarkeit, Robustheit, allg. Funktionsfähigkeit)	<b>KATEGORIEN</b> Vergleich zwischen Original und Modell Vergleich der Eigenschaften (Struktur und/oder Funktion) des Originals mit denen des Modells Vergleich und Passung zwischen Original und Modell zusätzlich zum Vergleich der Eigenschaften eine Beschreibung notwendiger Übereinstimmungen (Passung) zwischen Modell und Original/Nennung von Kriterien für ein gutes Modell	<b>KATEGORIEN</b> Bildung und/ oder Überprüfung von Hypothesen Bildung und/ oder Überprüfung von Hypothesen über das Original mit dem Modell und Nennung allgemeiner Ideen für Untersuchungen Überprüfung von Hypothesen mit Forschungsdesign Beschreibung einer konkrete Anwendung des Modells (Forschungsdesign) zur Überprüfung einer Hypothese über das Original
<b>SCHLÜSSELIDEEN</b> Bezüge nur zum Modellobjekt (Funktionalität, Material, ...)	<b>SCHLÜSSELIDEEN</b> Bezug explizit zum Original, Modellkritik (d.h. Vergleich zwischen Modell und Original)	<b>SCHLÜSSELIDEEN</b> Annahmen/ Hypothesen zulassen/ bilden/ prüfen

## Anhang

Ändern von Modellen		
Frage im Testheft: Nennen Sie Gründe, warum ein gegebenes biologisches Modell Ihrer Meinung nach verändert wird. (Bitte nennen Sie wenigstens zwei unterschiedliche Gründe.)		
Bei didaktischen Perspektiven, die eine Veränderung/Anpassung des Modells an z. B. neue Lerngruppen beschreiben, eine 888 notieren.		
Niveau I Mängel am Modellobjekt beheben	Niveau II Modell als Modell von etwas durch neue Erkenntnisse oder zusätzliche Perspektiven revidieren	Niveau III Modell für etwas aufgrund falsifizierter Hypothesen revidieren
<p><b>KATEGORIEN</b> Kein Anlass für eine Änderung Ablehnung einer Änderung des Modells Ändern zur Darstellung unterschiedlicher Originale Herstellung verschiedener Modelle für unterschiedliche Originale; jedes Original wird durch ein Modell repräsentiert Ändern zur Verbesserung des Modellobjekts Optimierung der Funktionsfähigkeit/ der Ästhetik/ der Zeitgemäßheit des Modellobjekts Optimierung der Technik zur Modellherstellung Ändern bei Fehlern im Modellobjekt grundsätzliche Überlegungen zur Behebung von Fehlern am Modell Verweis auf konkrete fehlerhafte Eigenschaften des Modells (z. B. Materialmängel) Ändern bei Nichterfüllung der Grundvoraussetzungen Überprüfung der grundlegenden Anforderungen an das jeweilige Modell und ggf. Behebung des Mangels</p> <p><b>SCHLÜSSELIDEEN</b> Bezüge nur zum Modellobjekt (Funktionalität, Material, neue Darstellungsweise, Größe, Farbe, ...), Optimierung der Technik Modernisierung</p>	<p><b>KATEGORIEN</b> Ändern bei mangelnder Passung mit dem Original Optimierung der mangelnden Passung des Modells (Struktur und/oder Funktion) mit dem Original mit Blick auf notwendige Übereinstimmungen zwischen Original und Modell Ändern bei neuen Erkenntnissen über das Original Integration neuer (wissenschaftlicher) Erkenntnisse über das Original ins Modell; verbesserte Technik führt zu neuen Erkenntnissen über das Original Ändern bei Veränderung des Originals Berücksichtigung von Veränderungen (z. B. Individualentwicklung) bzw. Weiterentwicklungen (z. B. evolutive Anpassung) des Originals als neue Informationen im Modell</p> <p><b>SCHLÜSSELIDEEN</b> Bezug explizit zum Original, neue Forschungsergebnisse/ Erkenntnisse der Wissenschaft, Widersprüche/ mangelnde Passung zum Original</p>	<p><b>KATEGORIEN</b> Ändern bei Erkenntnissen aus Modellexperiment Anpassung des Modells an Erkenntnisse über das Original auf der Grundlage eines Modellexperiments/ Falsifizieren der Hypothese, die dem Modell zugrunde liegt Anpassen des Modells zur Untersuchung/ Berücksichtigung weiterer/ neuer Variablen in Modellexperiment</p> <p><b>SCHLÜSSELIDEEN</b> Annahmen/ Hypothesen aus Modellexperiment/ mit Hilfe des Modells abgeleitet, sind widerlegt/ falsifiziert</p>