

Fachbereich Erziehungswissenschaft und Psychologie

der Freien Universität Berlin

Ein diagnostisches Fallinventar: Entwicklung, Erprobung und Evaluation
einer fallbasierten Lerngelegenheit zur Förderung von diagnostischen
Kompetenzen im Lehramtsstudium

Dissertation

zur Erlangung des akademischen Grades

Doktorin der Philosophie (Dr. phil.)

vorgelegt von

Jacqueline Wißmann

Berlin, 30.06.2020

Erstgutachterin: Prof. Dr. Annette Kinder

Zweitgutachterin: Prof. Dr. Bettina Hannover

weitere Kommissionsmitglieder:

Prof. Dr. Felicitas Thiel, Prof. Dr. Sascha Hein, Dr. Bettina Röder

Datum der Disputation: 04.09.2020

Danksagung

Auf meinen Weg haben mich viele Menschen begleitet, ohne die diese Arbeit nicht entstanden wäre. Allen voran mein Ehemann und mein Sohn, die – genau wie ich – viele Entbehrungen auf sich genommen, meine Zweifel und Krisen ertragen, gewartet und gebetet haben, dass ich es bald schaffen möge. Auch meine „Doktormutter“ hat sicher das eine oder andere Stoßgebet gen Himmel geschickt, dass diese Arbeit fertig werde, und mit Geduld und Langmut versichert, dass ich es schaffe. Ich selbst habe in einigen Momenten nicht daran glauben können und bin jetzt erleichtert, dass ich die „Disco“ – wie mein Sohn immer sagte, wenn ich ins Schlafzimmer an den Heimarbeitsplatz entschwand – verlassen kann. Diese Arbeit ist fertig geworden. Den nötigen Zuspruch, hilfreiche Nachfragen, wie weit ich sei, und offene Ohren für verknotete Gedanken erhielt ich von meiner Klein- und Großfamilie, von Freund*innen (namentlich nennen möchte ich hier Pia, Julia, Heike und Nicole) und Kolleg*innen (insbesondere Judit und Maike). Nicht zuletzt hat mich meine Zweitgutachterin zum Endspurt motiviert, um diesen Termin zu halten. Ein herzliches Dankeschön gilt in diesem Zusammenhang auch der Schule meines Sohnes, die uns kurzerhand die Corona-Notbetreuung bis zu den Sommerferien ermöglichte. Besonderer Dank gilt schließlich den studentischen Mitarbeiter*innen, die mich und meine Arbeit bis zuletzt tatkräftig unterstützt haben (Ines, Antonia, Flora und Kevin).

Diese Dissertation ist mit Unterstützung des Projekts „K2teach – Know how to teach“ entstanden. „K2teach“ wird im Rahmen der gemeinsamen Qualitätsoffensive Lehrerbildung von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1802 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin.

» Siehst du den Mond dort stehen?
Er ist nur halb zu sehen und ist doch rund und schön!
So sind wohl manche Sachen, die wir getrost belachen,
weil unsere Augen sie nicht seh'n. «

Matthias Claudius, Volkslied

Zusammenfassung

Die KMK hat diagnostische Kompetenzen als eine der vier zentralen professionsbezogenen Kompetenzen von Lehrkräften herausgestellt, die es bereits im Studium vorzubereiten gilt. Bislang ist das Angebot an evaluierten Interventionen zur Förderung, vor allem in der universitären Ausbildungsphase, allerdings überschaubar. Dabei kann professionelles diagnostisches Vorgehen bereits im Lehramtsstudium vorbereitet werden, indem neben deklarativem Wissen auch prozedurales diagnostisches Wissen gefördert wird (diagnostische Expertise). Die Förderung prozeduralen Wissens ist jedoch mit Herausforderungen verbunden, weil dabei praktische Erfahrungen essentiell sind.

In der vorliegenden Arbeit wird die Entwicklung, Erprobung und Evaluation einer fallbasierten Lerngelegenheit vorgestellt, die den Aufbau prozeduralen Wissens von Lehramtsstudierenden unterstützt. Im Rahmen komplexitätsreduzierter, realistischer Fallvignetten wird damit auf einen Diagnoseanlass vorbereitet, der weitreichende Konsequenzen hat: die Feststellung von Lernbesonderheiten. Dieser Diagnoseanlass erfordert sogenannte formelle Diagnosen, die im Rahmen eines regelgeleiteten diagnostischen Prozesses unter Anwendung standardisierter diagnostischer Testverfahren erstellt werden. Zur Vorbereitung auf diese anspruchsvolle Aufgabe üben die Studierenden im Fallinventar die Vergabe von Diagnosevermutungen sowie die Ableitung von pädagogischen Anschlusshandlungen ein.

Bei der Bearbeitung des Fallinventars zeigte sich, dass den Studierenden die Ableitung von Anschlusshandlungen noch schwerer fällt als die Vergabe von Diagnosevermutungen. Diese Schwierigkeit konnte dadurch erklärt werden, dass die Studierenden formelle diagnostische Informationen noch nicht systematisch genug nutzten und dass sie ihre Empfehlungen nicht auf Grundlage der Diagnosen vergaben. Die Fähigkeit zur Vergabe von Empfehlungen konnte erfreulicherweise durch die Arbeit mit dem Fallinventar verbessert werden, indem prozedurales Wissen trainiert wurde. In Kombination mit einer aufgabenbasierten Übung konnte andererseits deklaratives Wissen gefördert werden, das die Trefferquoten bei Diagnosevermutungen erhöhte. Außerdem konnten erste affektiv-motivationale Voraussetzungen geschaffen werden, die die Wahrscheinlichkeit diagnostisch kompetenten Verhaltens in der späteren Berufspraxis erhöhen. Um diese Voraussetzungen zu stabilisieren und weiter auszubauen, bedarf es allerdings weiterer Lerngelegenheiten, die eine schrittweise Praxisannäherung im Lehramtsstudium ermöglichen.

Die Ergebnisse erweitern die Befundlage zur Förderung diagnostischer Kompetenzen im Lehramtsstudium und haben Implikationen für die Modellierung diagnostischer Kompetenzen, die bislang die Qualität der Anschlusshandlungen unzureichend berücksichtigt.

Schlüsselwörter: diagnostische Kompetenz, Lehramtsstudierende, prozedurales Wissen, fallbasierte Lerngelegenheit, diagnostischer Prozess

Abstract

Developing diagnostic competence is a central issue in teacher as well as in pre-service teacher education (for Germany: KMK). To date, however, there are only few intervention programs to foster pre-service teachers' diagnostic competencies. When preparing professional diagnostic procedures during university teacher training, it is important to promote not only declarative but also procedural knowledge ("diagnostic expertise"). Supporting the acquisition of procedural knowledge is indeed challenging since gaining practical experience is essential.

In the present work, I introduce and evaluate a case-based learning opportunity which was designed to support the development of procedural knowledge for a particularly responsible diagnostic task, namely the diagnosis of learning exceptionalities. The so called "case inventory" presented here, contains realistic case descriptions of reduced complexity. Teacher students are asked to submit diagnostic judgments and to determine pedagogical recommendations for the virtual students described in the "case inventory". These two steps require the application of a rule-guided diagnostic process that relies on standardized diagnostic measures, such as academic performance tests.

Upon completing the case inventory, students struggled with selecting the correct diagnostic judgments and they were challenged even more with determining recommendations for the virtual students. First, these difficulties can be explained by an unsystematic use of formal diagnostic information. Second, pre-service teachers did not sufficiently relate their recommendations to their diagnostic judgments.

Fortunately, by increasing procedural knowledge, the case inventory supported their ability to derive recommendations. Combined with a conventional task-based exercise, the development of declarative knowledge was supported that increased the hit rates of diagnostic judgements. Furthermore, the results suggest that developing first motivational aspects of diagnostic competencies can be supported by combining these two learning opportunities. In this way, it is more likely that pre-service teachers will transfer their knowledge to later professional practice. However, further learning opportunities are necessary to stabilize, expand and approximate the diagnostic processing to practical requirements.

The results add to the empirical findings on fostering diagnostic competencies in pre-service teachers and have implications for modelling diagnostic competencies as well which did not focus on the quality of pedagogical recommendations until recently.

Keywords: diagnostic competencies, pre-service teachers, case-based learning opportunity, diagnostic process, procedural knowledge

Inhalt

Abbildungen	11
Tabellen	12
Einleitung und Zielsetzung	13
1 Diagnostische Tätigkeiten im Lehramt	19
1.1 Relevanz diagnostischer Kompetenzen	19
1.1.1 Diagnostische Kompetenzen tangieren Schülerinnen und Schüler.....	19
1.1.2 Diagnostische Kompetenzen als aktuelles Forschungsdesiderat.....	20
1.2 Diagnostische Aufgaben und Vorgehensweisen von Lehrkräften	21
1.2.1 Diagnostische Aufgaben von Lehrkräften.....	21
1.2.2 Pädagogische Diagnostik als besondere Form der Diagnostik	22
1.2.3 Verschiedene Arten pädagogischer Diagnosen	24
1.2.4 Der diagnostische Prozess	28
1.3 Zusammenfassung und Zwischenfazit	37
2 Indikatoren diagnostischer Kompetenz von Lehrkräften	38
2.1 Urteilsgenauigkeit als Kern diagnostischer Kompetenz?.....	38
2.1.1 Ausgewählte Befunde zur Urteilsgenauigkeit.....	39
2.1.2 Diagnostische Expertise als Erweiterung der Urteilsgenauigkeit.....	42
2.2 Prozessmerkmale als Indikator diagnostischer Kompetenz.....	43
2.2.1 Die Diagnose als Produkt und als Prozess.....	44
2.2.2 Modellierung diagnostischer Kompetenz anhand von Prozessmerkmalen	47
2.3 Zusammenfassung und Zwischenfazit	50
3 Vorüberlegungen zur Förderung diagnostischer Kompetenz	52
3.1 Überlegungen zum Begriff der Kompetenz.....	52
3.1.1 Kompetenz als Voraussetzung für Performanz	52
3.1.2 Kompetenz als Kontinuum	53
3.2 Kognitive Voraussetzungen diagnostischer Kompetenz	55
3.2.1 Deklaratives Wissen	56
3.2.2 Prozedurales Wissen	58

3.2.3	Erwerb deklarativen und prozeduralen Wissens	59
3.2.4	Inhaltsbereiche des diagnostischen Wissens	64
3.3	Affektiv-motivationale Voraussetzungen diagnostischer Kompetenz	66
3.3.1	Einstellungen als Prädiktor für professionelles Handeln	67
3.3.2	Selbstwirksamkeitserwartungen und kompetenzbezogene Selbsteinschätzungen	69
3.4	Zusammenfassung und Zwischenfazit	71
4	Ansätze zur Förderung diagnostischer Kompetenz	72
4.1	Verbesserung der Urteilsgenauigkeit	72
4.2	Verbesserung kompetenten Vorgehens im diagnostischen Prozess	73
4.2.1	Ansätze zur Förderung von Fähigkeiten zur Erstellung formeller Diagnosen	73
4.2.2	Ansätze zur Förderung unterrichtsbezogener diagnostischer Tätigkeiten	74
4.3	Förderung kognitiver Voraussetzungen diagnostischer Kompetenz	76
4.4	Zusammenfassung und Zwischenfazit	79
5	Hinführung zu den empirischen Studien	80
6	Die Entwicklung und Pilotierung einer fallbasierten Lerngelegenheit (Studie 1)	83
6.1	Hintergrund und Überblick	83
6.2	Diagnostik von Lernbesonderheiten als besondere diagnostische Aufgabe	84
6.2.1	Definition und Vorkommen von Lernbesonderheiten	84
6.2.2	Prinzipien der Diagnostik und Förderung bei Lernbesonderheiten	87
6.3	Fähigkeiten zur Diagnostik von Lernbesonderheiten fördern	89
6.3.1	Förderung von Grundlagenwissen zur Diagnostik von Lernbesonderheiten	90
6.3.2	Förderung prozeduralen Wissens zur Diagnostik von Lernbesonderheiten	91
6.4	Konzeption und Gestaltung eines diagnostischen Fallinventars	92
6.4.1	Komplexitätsreduzierte Fallvignetten	93
6.4.2	Darstellung prototypischer Fähigkeitsprofile	94
6.4.3	Relevante und irrelevante Informationen zu Fähigkeiten der Schüler*innen	97
6.4.4	Irrelevante und potentiell urteilsverzerrende Zusatzinformationen	99
6.4.5	Diagnosen und pädagogische Entscheidungen als diagnostische Produkte	99
6.5	Pilotierung des Fallinventars	102
6.5.1	Stichprobe	102

6.5.2 Instrument	102
6.5.3 Durchführung	103
6.5.4 Auswertung.....	103
6.5.5 Ergebnisse.....	104
6.5.6 Diskussion und Schlussfolgerungen	105
6.6 Entwicklung der finalen Version des Fallinventars	107
6.6.1 Überarbeitung der Aufgabenstellungen	108
6.6.2 Überarbeitung der Fallvignetten.....	108
6.6.3 Zwischenfazit zu Studie 1	111
7 Diagnostische Produkt- und Prozessvariablen: Untersuchung der Bearbeitung des Fallinventars durch Studierende (Studie 2).....	113
7.1 Hintergrund und Überblick	113
7.2 Der diagnostische Prozess im Forschungsinteresse	113
7.3 Das diagnostische Fallinventar als Szenario basierte Umgebung.....	118
7.3.1 Fragestellungen und Hypothesen.....	119
7.3.2 Methoden.....	121
7.3.3 Ergebnisse.....	129
7.3.4 Diskussion der Ergebnisse	138
8 Evaluation des diagnostischen Fallinventars als fallbasierte Lerngelegenheit (Studie 3)	149
8.1 Hintergrund und Überblick	149
8.2 Aufbau diagnostischen Wissens bei angehenden Lehrkräften.....	149
8.3 Kognitive Veränderungen und weitere Zielvariablen in der Lehrkräftebildung	151
8.3.1 Vier Ebenen der Evaluation für Interventionen in der Lehrkräftebildung.....	151
8.3.2 Evaluation von Interventionen für Lehramtsstudierende: Kognitive und affektiv-emotionale Voraussetzungen im Zentrum.....	153
8.4 Evaluation des diagnostischen Fallinventars.....	154
8.4.1 Fragestellungen und Hypothesen.....	155
8.4.2 Methoden.....	158
8.4.3 Ergebnisse und Interpretation.....	170
8.5 Diskussion der Ergebnisse.....	179
9 Abschließende Diskussion	190

9.1 Zusammenfassung der empirischen Untersuchungen	190
9.2 Einordnung der Ergebnisse in den Forschungskontext	192
9.2.1 Hinweise zur Förderung diagnostischer Kompetenzen	192
9.2.2 Implikationen für die Modellierung diagnostischer Kompetenz	196
9.3 Ausblick und Vorschlag eines erweiterten Rahmenmodells für die Erforschung diagnostischer Kompetenzen	199
9.4 Fazit	201
Literatur	202
Anhänge	218
Anhang 2A. Zusammenstellung der Fallhefte für Studie 2	218
Anhang 2B. Antworten für einzelne Fallvignetten in Studie 2	219
2B.1 Fall 1: Julia / Maximilian	219
2B.2 Fall 2: Lena / Christian	220
2B.3 Fall 3: Katrin / Felix	221
2B.4 Fall 4: Sophie / Lukas	222
2B.5 Fall 5: Hanna / Jonas	223
2B.6 Fall 6: Jana / Luca	224
2B.7 Fall 7: Leonie / Jan	225
2B.8 Fall 8: Laura / Thomas	226
Anhang 2C. Scatterplots zu Korrelationsanalysen in Studie 2	227
2C.1 Zusammenhang der Lösungsraten von Diagnosen und Empfehlungen	227
2C.3 Zusammenhang des Nutzungsanteils formeller Diagnoseinformationen und Lösungsraten ...	228
Anhang 3.A Konzeptualisierung des Wissenstests in Studie 3	229
Anhang 3.B Itembeschreibungen und –statistiken für Studie 3	230
3.B.1 Evaluationsebene 1: Rezeption	230
3.B.2 Evaluationsebene 2: Lernen	230
Anhang 3.C Voranalysen zu potentiellen Störvariablen für Studie 3	236
Anhang 3.D Faktorenanalyse des zweidimensionalen Wissenstests in Studie 3	237
Anhang 3.E Einschätzungen der Lerngelegenheiten durch die Dozierenden	238
Anhang 3.F Deskriptive Ergebnisse der Studierenden ohne Übung	239

Eigenständigkeitserklärung 240

Abbildungen

ABBILDUNG 1.1 DER DIAGNOSTISCHE PROZESS ALS MODELL MIT FLEXIBLEN ABLÄUFEN. DARSTELLUNG IN ANLEHNUNG AN HERPPICH ET AL. (2017, S. 82)	30
ABBILDUNG 1.2 DER DIAGNOSTISCHE PROZESS ALS MODELL MIT REGELGELEITETEM VORGEHEN.....	31
ABBILDUNG 1.3 DER ERWEITERTE DIAGNOSTISCHE PROZESS IM PÄDAGOGISCH-PSYCHOLOGISCHEN SETTING.	36
ABBILDUNG 2.1 VIER-KOMPONENTEN-MODELL DER DIAGNOSEQUALITÄT. DARSTELLUNG IN ANLEHNUNG AN BEHRMANN & VAN OPHUYSEN (2017, S. 39).	45
ABBILDUNG 2.2 SCHEMATISCHE DARSTELLUNG DES NeDiKo–ARBEITSMODELLS DIAGNOSTISCHER KOMPETENZ. (DARSTELLUNG IN ANLEHNUNG AN HERPPICH ET AL., 2017, 2018)	49
ABBILDUNG 3.1 MODELL DER KOMPETENZ ALS KONTINUUM. (DARSTELLUNG IN ANLEHNUNG AN BLÖMEKE ET AL. (2015, S. 7))	54
ABBILDUNG 3.2 WISSENSARTEN IM RAHMEN DES ERWERBS PROFESSIONELLER KOMPETENZ.....	61
ABBILDUNG 6.1 RAHMENMODELL SCHRITTWEISER PRAXISANNÄHERUNG IN DER LEHRKRÄFTEBILDUNG (GROSSMAN ET AL., 2009) ...	92
ABBILDUNG 6.2 SCHEMATISCHER AUFBAU EINES FALLBLATTES IM DIAGNOSTISCHEN FALLINVENTAR	100
ABBILDUNG 7.1 HÄUFIGKEITEN DER ANTWORTOPTIONEN ZUM SUBJEKTIV EMPFUNDENEN NUTZEN DES FALLINVENTARS ALS LERNGELEGENHEIT (SECHSSTUFIGE ANTWORTSKALA)	130
ABBILDUNG 7.2 MITTLERE LÖSUNGSRATEN FÜR DIAGNOSEN UND EMPFEHLUNGEN IN ABHÄNGIGKEIT.....	135
ABBILDUNG 7.3 NUTZUNGSANTEILE DER DIAGNOSTISCHEN INFORMATIONEN ENTSPRECHEND DER REIHENFOLGE VON DIAGNOSE- UND EMPFEHLUNGSENTSCHEIDUNGEN.	137
ABBILDUNG 8.1 GRUPPENUNTERSCHIEDE BEIM DIAGNOSTISCHEN WISSEN	177
ABBILDUNG 8.2 GRUPPENUNTERSCHIEDE BEI DIAGNOSTIKBEZOGENEN SELBSTEINSCHÄTZUNGEN.....	178
ABBILDUNG 8.3 GRUPPENUNTERSCHIEDE BEI DIAGNOSTIKBEZOGENEN EINSTELLUNGEN	179
ABBILDUNG 9.1 VORSCHLAG EINES ERWEITERTEN RAHMENMODELLS DIAGNOSTISCHER KOMPETENZ.....	200

Tabellen

TABELLE 1.1 UNTERSCHIEDUNG VON DIAGNOSEARTEN NACH DEM FORMALISIERUNGSGRAD	27
TABELLE 3.1 LERNZIELEBENEN IM ÜBERBLICK	63
TABELLE 6.1 DEFINITIONEN UND PRÄVALENZEN VON EINZELNEN LERNSTÖRUNGEN	85
TABELLE 6.2 DEFINITION UND PRÄVALENZEN VON LERNBESONDERHEITEN IM KOGNITIVEN BEREICH.....	86
TABELLE 6.3 KATEGORISIERUNG VON LERNSCHWIERIGKEITEN HINSICHTLICH ZEITLICHER UND BEREICHSBEZOGENER STABILITÄT.....	88
TABELLE 6.4 PROTOTYPISCHE FÄHIGKEITSPROFILE DER LERNBESONDERHEITEN	95
TABELLE 6.5 ERGEBNISSE DER PILOTIERUNG DES FALLINVENTARS FÜR DREI SEMINARE	105
TABELLE 6.6 KURZBESCHREIBUNG DER FALLVIGNETTEN IN DER FINALEN VERSION DES FALLINVENTAR	109
TABELLE 6.7 KENNWERTE FÜR DIE AUSWAHL UND ZUSAMMENSTELLUNG DER WEIBLICHEN UND MÄNNLICHEN FALLNAMEN	111
TABELLE 7.1 VALIDIERTE LÖSUNGSVORSCHLÄGE FÜR DIAGNOSEN UND EMPFEHLUNGEN	127
TABELLE 7.2 FALLBEZOGENE AUFGABENSCHWIERIGKEITEN UND SICHERHEITS RATINGS	132
TABELLE 7.3 NUTZUNGSANTEILE FÜR DIE ARTEN DIAGNOSTISCHER INFORMATIONEN IM VERGLEICH	136
TABELLE 8.1 KURZBESCHREIBUNG DER INHALTE DER AUFGABENBASIERTEN ÜBUNG	159
TABELLE 8.2 SCHEMA ZUR ENTSCHEIDUNG ÜBER LERNZIELEBENEN	161
TABELLE 8.3 INDIKATOREN FÜR DEN VERGLEICH DER LERNGELEGENHEITEN (FRAGESTELLUNG 2A).....	163
TABELLE 8.4 INSTRUMENTE ZUR ANALYSE DER INTERVENTIONSGRUPPEN (FRAGESTELLUNG 2B)	166
TABELLE 8.5 EVALUATIONSDESIGN ZUR ANALYSE DER INTERVENTIONSGRUPPEN.....	167
TABELLE 8.6 ERGEBNISSE DES EXPERT*INNENRATINGS ZU DEN ÜBUNGEN.....	171
TABELLE 8.7 ERGEBNISSE DES EXPERT*INNENRATINGS ZUM WISSENSTEST	172
TABELLE 8.8 LÖSUNGSRATEN FÜR DIAGNOSE- UND EMPFEHLUNGSENTSCHEIDUNGEN IM SZENARIOTEST.....	173
TABELLE 8.9 ANTEIL BEARBEITETER ARBEITSAUFRÄGE IN DEN INTERVENTIONSGRUPPEN	174
TABELLE 8.10 ERGEBNISSE FÜR DEN VERGLEICH DER LERNGELEGENHEITEN (FRAGESTELLUNG 2A).....	175

Einleitung und Zielsetzung

Zwei Fallbeispiele von Schüler*innen mit Lernbesonderheiten

Sarah¹ ist endlich in der siebten Klasse, an einer weiterführenden Schule², die sie sich gewünscht hat. Bis hierher ist es ein harter Weg gewesen. Seit der ersten Klasse hat Sarah gravierende Schwierigkeiten im Rechnen, ihr zwei Jahre jüngerer Bruder kann bald besser rechnen als sie. In der Schule müssen sie und ihre Mutter immer wieder dafür kämpfen, dass ihre Rechenschwäche anerkannt wird, und dass Sarah Unterstützung bekommt. Ohne Förderunterricht und Nachteilsausgleich kommt Sarah im Unterricht nicht mit, und schlimmer noch, ihr Selbstwertgefühl leidet massiv und das wirkt sich auch auf andere Fächer aus. Sarah hat keine Kraft mehr, jetzt auch noch ihre neuen Lehrer*innen davon zu überzeugen, dass sie eigentlich nichts versteht in Mathe. In der Grundschule konnte sie durch intensive Unterstützung seitens der Mutter und außerschulische Förderung, die die Eltern privat zahlten, immer korrekte Hausaufgaben vorweisen. In den Tests bekam sie meist die Note Drei oder Vier – kein wirklicher Grund zur Sorge, befanden die Lehrkräfte, zumal es in der Klasse noch andere Schülerinnen gäbe, die noch viel schlechter seien als Sarah. Wieviel Kraft und Zeit Sarah allerdings investiert hat, um diese Noten zu erreichen, mit denen sie nicht zufrieden ist, und wie sehr sie sich mit ihren Schwächen statt mit ihren Stärken beschäftigt, das ist ihr von außen nicht anzusehen.

Ricardo¹ ist aktuell in der zehnten Klasse eines Gymnasiums. Er schreibt Theaterstücke, führt Regie, hat recht gute Noten in der Schule. Mit seinen Schulleistungen fällt er in der Masse nicht negativ auf, aber mit seinen Forderungen. Er möchte mehr Zeit bei schriftlichen Arbeiten, und er möchte, dass seine Rechtschreibung nicht benotet wird. Allerdings hat er aus Sicht der Lehrkräfte nur leichte Probleme bzgl. Orthografie und Grammatik, und dafür Eltern, die übertreiben. Tatsächlich unterstützt die Mutter ihren Sohn intensiv, und auch Ricardo übt beharrlich und ehrgeizig. Seine Lese- und Rechtschreibleistungen sind aktuell – hart erarbeitet – im unteren Durchschnittsbereich anzusiedeln. Aber er leidet darunter, weil er das, was er in seiner Vorstellung hat, nicht auf das Papier übertragen kann. Er schämt sich für jeden Rechtschreibfehler, vor allem vor den Mitschüler*innen. Vor jeder Präsentation überfordern ihn seine Gefühle, denn seine Mitschüler*innen haben wenig Nachsicht. In schriftlichen Tests kommentieren sie die mühsam erkämpfte Zeitverlängerung argwöhnisch. Auch die Lehrkräfte sind skeptisch, gewähren den Nachteilsausgleich jedoch – wenn Ricardo regelmäßig ein neues Attest vorlegt. Das ist bei schriftsprachlichen Leistungen im Durchschnittsbereich allerdings nicht leicht zu bekommen.

In den Fallbeschreibungen werden Schüler*innen vorgestellt, die den leistungsbezogenen bzw. sozialen Anforderungen in ihrer Schule nicht gewachsen sind. Beide Schüler*innen haben Lernschwierigkeiten, die sich ganz unterschiedlich äußern, und die von ihren jeweiligen Lehrkräften nicht in ausreichendem Maße

¹ Die beschriebenen Fälle stammen aus der lerntherapeutischen Arbeit der Autorin. Es handelt sich um reale Fallbeschreibungen, bei denen aus Datenschutzgründen Namen, Geschlecht und einige Details geändert wurden.

² In Berlin erstreckt sich die Grundschulzeit bis zum Ende der sechsten Klassenstufe.

erkannt werden. Dabei hat Sarah inzwischen diverse Befunde testpsychologischer Untersuchungen, die ihr eindeutig eine umschriebene Rechenschwäche bzw. –störung attestieren. Noch immer ist sie dem zählenden Rechnen verhaftet, eine Vorstellung vom dekadischen Stellenwertsystem hat sie nicht entwickelt. Bei Ricardo ist es weniger eindeutig. Auch er hat sich mehrfach von verschiedenen Kinder- und Jugendlichenpsychiater*innen untersuchen lassen, hat zunächst eine Lese-Rechtschreibstörung diagnostiziert und vom Jugendamt aufgrund der psychischen Belastung eine lerntherapeutische Förderung finanziert bekommen. Inzwischen erfüllt er die Kriterien einer Lese-Rechtschreibstörung nicht mehr, jedoch sind die Lernschwierigkeiten nach wie vor präsent. Das liegt daran, dass seine allgemeinen kognitiven Fähigkeiten – im Gegensatz zu den durchschnittlichen schriftsprachlichen Fähigkeiten – im weit überdurchschnittlichen Bereich anzusiedeln sind. Diese Diskrepanz belastet Ricardo als (bislang nicht offiziell diagnostizierter) hochbegabter Minderleister deutlich. Für die Lehrkräfte bleibt diese Diskrepanz unsichtbar, da sie ihren eigenen Beobachtungen und Einschätzungen vertrauen, die sie im Rahmen ihres Unterrichts gewonnen haben. Auf diese Weise begehen sie Urteilsfehler, die für Sarah und Ricardo spürbar sind und die sich auf ihre schulische sowie ihre psychische Entwicklung auswirken.

Lehrkräfte für eine datenbasierte Diagnostik befähigen und motivieren

Diese zwei Fallbeispiele verdeutlichen, wie wichtig es ist, dass Lehrkräfte ihre „beiläufig gewonnenen Eindrücke“ systematisch reflektieren, anhand objektiver Kriterien analysieren und gegebenenfalls korrigieren (vgl. Hascher, 2008; Helmke, 2009; Schrader, 2010). In Berlin können Lehrkräfte insbesondere auf die Feststellung individueller Lernausgangslagen (z. B. ILeA 2-6: LISUM, 2018), die Ergebnisse aus Vergleichsarbeiten (VERA-3 und VERA-8; IQB, 2018) sowie auf Ergebnisse standardisierter Schulleistungstests (z. B. Hamburger Schreibprobe: May, 2012) zurückgreifen, um ihre *eigenen Einschätzungen mit objektiven Daten* zu vergleichen. Wenn Lehrkräfte die Daten aus diesen objektiven Verfahren nutzen, vermindert das nicht nur die Gefahr von Urteilsfehlern. Diese Daten unterstützen sie auch bei der Feststellung von Förderschwerpunkten für einzelne Schüler*innen sowie bei der Unterrichtsentwicklung allgemein (z. B. Helmke & Lenske, 2013; Mandinach & Gummer, 2016). Die unlängst aktualisierten schulrechtlichen Regelungen in Berlin sehen die jeweiligen Fachlehrkräfte sogar in der Verantwortung, die schriftsprachlichen sowie numerisch-rechnerischen Fähigkeiten ihrer Schüler*innen dahingehend zu untersuchen, ob umschriebene Schwierigkeiten im Lesen, Schreiben oder Rechnen vorliegen – und hierfür sollen gezielt objektive Verfahren eingesetzt werden, die die differenzierte Feststellung solcher Schwierigkeiten erlauben (vgl. SenBJW, 2012).

Dass Lehrkräfte solche Daten erheben oder nutzen können, führt jedoch nicht zwangsläufig dazu, dass sie dies auch tun (Hosenfeld & Groß Ophoff, 2007; Wurster et al., 2017). Befragungen zur Nutzung von Vergleichsarbeiten etwa zeigen eindrücklich, dass Lehrkräfte diese Daten „selektionsdiagnostisch“ verwenden, d. h. häufig nur dann, wenn sie ihren eigenen Eindrücken entsprechen (Maier, 2008). Dabei ist es gerade ein Ziel von Vergleichsarbeiten, dass Lehrkräfte regelmäßig zur Reflektion ihrer Urteile sowie widersprüchlicher Befunde angeregt werden und so ihre diagnostischen Kompetenzen verbessern

(Helmke & Hosenfeld, 2005). Die *Nutzung diagnostischer Daten* kann also dazu beitragen, dass Lehrkräfte ihre Schüler*innen und deren Fähigkeiten korrekter einschätzen, und damit einen entscheidenden Beitrag dazu leisten, diagnostisch professionell zu handeln (Behrmann & van Ophuysen, 2017; Helmke, 2009; Hesse & Latzko, 2017; Jäger, 2006). Im Zentrum dieser professionellen Diagnostik steht ein regelgeleiteter diagnostischer Prozess (Hesse & Latzko, 2017). Dieser sieht die Erhebung und Nutzung diagnostischer Daten explizit vor, da diese es den Lehrkräften ermöglichen, zu einem begründeten diagnostischen Urteil zu kommen, auf dieser Basis Förderstrategien und –maßnahmen für ihre Schüler*innen abzuleiten und diese bestenfalls auch zu evaluieren (Gummer & Mandinach, 2015; Mandinach & Gummer, 2016). Damit Lehrkräfte in der Praxis ein derart professionelles Handeln zeigen können, sind sie u. a. auf eine gute Aus- und Weiterbildung angewiesen, in der sie entsprechende Fähigkeiten zur Datennutzung sowie die Bereitschaft, diese auch in der Schulpraxis umzusetzen, erwerben. Die „Qualitätsoffensive Lehrerbildung“ (BMBF, 2016) strebt eine solch *umfassende Bildung für Lehrkräfte* an, die durch eine Verzahnung des Kompetenzerwerbs in Lehramtsstudium und Berufspraxis erreicht werden soll, d. h. durch eine Verbindung wissenschaftlich-theoretischer Inhalte mit relevanten Praxisaspekten bereits im Lehramtsstudium. Das Projekt „K2teach: Know how to teach“ (K2teach, 2019) an der Freien Universität Berlin, als ein Teil der Qualitätsoffensive, verfolgt seit Ende 2015 das Ziel, Lerngelegenheiten zu entwickeln, die angehende Lehrkräfte konkret auf ihre beruflichen Anforderungen vorbereiten. Im Teilprojekt „Forschungskompetenzen für eine evidenzbasierte adaptierte Unterrichtsentwicklung“ steht die Förderung von Datennutzungskompetenzen im Vordergrund, durch die die Lehramtsstudierenden u. a. in die Lage versetzt werden sollen, Daten aus objektiven diagnostischen Verfahren oder Verfahren der Unterrichtsevaluation verstehen und anwenden zu können und eine Bereitschaft zur Datennutzung zu entwickeln (Thoren et al., 2020).

In diesem Projektrahmen wurde eine *fallbasierte Lerngelegenheit* entwickelt, die angehende Lehrkräfte speziell auf die Nutzung individualdiagnostischer Daten vorbereitet. Vor dem Hintergrund meiner Tätigkeit als Lerntherapeutin nimmt diese Lerngelegenheit speziell die Feststellung von Lernschwierigkeiten und -besonderheiten in den Blick. Dadurch, dass ich Kinder und Jugendliche verschiedener Altersgruppen in der Diagnostik und Förderung ihrer Lernschwierigkeiten und –schwächen begleitet habe, spiegeln die im „Fallinventar“ arrangierten Fälle reale diagnostische Fragestellungen wider, die Studierende mithilfe realistischer diagnostischer Informationen beantworten können. Durch die Arbeit mit diesen realistischen Fallvignetten sollen angehende Lehrkräfte dafür sensibilisiert werden, dass zur Feststellung von Lernbesonderheiten eine bloße Eindrucksbildung nicht ausreichend ist, sondern evidenzbasierte Urteile im Sinne von Diagnosen erforderlich sind (vgl. Helmke, 2009). Diese gezielte Diagnostik von Lernstörungen (umschriebene Entwicklungsstörungen schulischer Fertigkeiten) sowie Besonderheiten kognitiver Lernvoraussetzungen (allgemeine Lernschwäche / Lernbehinderung, Hochbegabung), die ich unter dem Begriff der Lernbesonderheiten zusammenfasse, setzt voraus, dass objektive diagnostische Daten herangezogen oder erhoben werden. Nur diese formelle Diagnostik ermöglicht es den Lehrkräften, ihre subjektiven Eindrücke zu prüfen und gegebenenfalls zu revidieren (Schrader, 2009, 2010). Eine sorgfältige Diagnostik dient der Planung von Fördermaßnahmen, die den betroffenen Schüler*innen ein

Lernen nach ihren Fähigkeiten und Potentialen ermöglichen, damit sie diese im Rahmen des Unterrichts ausschöpfen und weiterentwickeln können. In der Arbeit mit dem „diagnostischen Fallinventar“ (Wißmann & Kinder, 2018) können Lehramtsstudierende bereits im Studium die Diagnostik und Ableitung von Fördermöglichkeiten kennenlernen. Auch wenn sich tatsächliche diagnostische Expertise erst mit wachsender Berufspraxis entwickelt, weil sie „die fortlaufende Konfrontation mit den typischen Aufgabenstellungen eines Arbeitsfeldes“ erfordert (Schrader, 2017, S. 253), kann diese Lerngelegenheit dazu beitragen, dass die Voraussetzungen hierfür bereits vor der eigentlichen Berufspraxis geschaffen werden.

Zur vorliegenden Arbeit

Die vorliegende Arbeit fasst zunächst theoretische und empirische Befunde zusammen, die im Zusammenhang der Entwicklung diagnostischer Expertise relevant sind. Der *theoretische Teil der Arbeit* wird durch eine Beschreibung der umfangreichen und herausfordernden diagnostischen Aufgaben von Lehrkräften in der Berufspraxis eröffnet (Kap. 1). Im zweiten Kapitel wird das Konstrukt der diagnostischen Kompetenz erörtert, indem auf die verschiedenen diagnostischen Tätigkeiten Bezug genommen wird. Das folgende Kapitel 3 umfasst Vorüberlegungen dahingehend, auf welche Weise diagnostische Kompetenz gefördert werden kann. Dabei werden insbesondere die Voraussetzungen diagnostischer Kompetenz beleuchtet, da diese im Rahmen des Studiums besonders gut implementiert werden können. Neben kognitiven Voraussetzungen, d. h. Wissen als einer der essentiellen Voraussetzungen von Expertise (z. B. Hascher, 2008), werden hier auch affektiv-motivationale Aspekte beachtet, die als Grundlage für die Entwicklung diagnostischer Kompetenz angesehen werden (Blömeke et al., 2015). Im letzten Kapitel des Theorieteils (Kap. 4) werden einige Ansätze zur Förderung diagnostischer Kompetenz zusammengestellt, wobei deutlich wird, dass ein Großteil dieser Ansätze für praktisch tätige Lehrkräfte, jedoch weniger für Lehramtsstudierende geeignet ist. Diese Ansätze werden um Interventionen ergänzt, die für die erste Phase der Lehrkräftebildung entwickelt wurden. Diese beziehen sich i. d. R. auf spezielle diagnostische Settings und können nur bedingt für die Diagnostik von Lernbesonderheiten bezogen werden.

Vor diesem Hintergrund ergeben sich drei Forschungsdesiderata, die im *empirischen Teil der Arbeit* bearbeitet werden. Im Zentrum der ersten Studie (Kap. 6) wird die Entwicklung und Pilotierung der fallbasierten Lerngelegenheit beschrieben. Hierbei wird auf die spezifische diagnostische Aufgabe der Feststellung von Lernbesonderheiten sowie Modelle und Befunde zu Urteilsprozessen eingegangen. Im Rahmen der zweiten Studie (Kap. 7) wird das erprobte diagnostische Fallinventar eingesetzt, um die Performanz der Studierenden beim Bearbeiten der zwei beschriebenen pädagogisch-diagnostischen Aufgaben (Diagnose und Anschlusshandlung) zu untersuchen. Dabei werden auch Merkmale des diagnostischen Prozesses analysiert, die zu mehr oder weniger treffsicheren Diagnosen und Anschlusshandlungen führen. Diese Analysen erlauben Aussagen darüber, wie diagnostische Prozesse gestaltet sein sollten, um zufriedenstellende Ergebnissen zu erbringen. Im letzten empirischen Abschnitt

(Kap. 8) wird das Fallinventar im Vergleich zu einer konventionellen aufgabenbasierten Übung auf verschiedenen Ebenen evaluiert. Neben der Erfassung kognitiver Voraussetzungen (deklaratives und prozedurales Wissen) werden hier auch affektiv-motivationale Aspekte (diagnostikbezogene Einstellungen, selbsteingeschätzte diagnostische Kompetenzen) erfasst, die diagnostisch kompetentes Handeln in der Praxis voraussetzt. Zum Abschluss der Arbeit (Kap. 9) werden die Befunde hinsichtlich des Erwerbs sowie der Konzeptualisierung diagnostischer Kompetenzen diskutiert.

Die vorliegende Arbeit liefert damit einerseits empirische Hinweise zur Ausprägung und Entwicklung diagnostischer Kompetenz in der ersten Phase der Lehrkräftebildung. Andererseits leistet sie einen Beitrag zur aktuellen Diskussion innerhalb der deutschen Forschungslandschaft zur Frage, was diagnostische Kompetenz im Kern beschreibe und welche Faktoren sie beeinflussen (Karst & Förster, 2017). Diese konzeptuellen Entwicklungen sind eine Voraussetzung dafür, dass theoretisch fundierte Ansätze zur Förderung diagnostischer Kompetenz entwickelt werden können. Auch hier besteht derzeit weiterer Bedarf, da die bisher vorliegenden Ansätze nur einen Anfang hinsichtlich der Wirksamkeitsforschung im Bereich diagnostischer Kompetenz darstellen (Karing & Seidel, 2017).

Neben diesen Beiträgen zur wissenschaftlichen Debatte bezüglich des Konstrukts der diagnostischen Kompetenz soll die vorliegende Arbeit auch in der *schulischen Praxis von Nutzen* sein. Ich hoffe, mit der vorliegenden Arbeit dazu beizutragen, dass angehende Lehrkräfte, die mit dieser Übung arbeiten, ihre diagnostischen Eindrücke prüfen und sorgfältig diagnostizieren, insbesondere wenn der Verdacht von Lernbesonderheiten im Raum steht. Indem sie das tun, erhalten Kinder und Jugendliche mit ihren unterschiedlichsten Lernvoraussetzungen die Chance, dass ihre Besonderheiten im Unterricht berücksichtigt werden können. Inklusiver Unterricht sollte nämlich nicht nur Schüler*innen mit offensichtlichen körperlichen oder geistigen Behinderungen berücksichtigen, sondern diverse Besonderheiten des Lernens und Leistens, wie es etwa bei Lernstörungen der Fall ist, mitdenken (s. a. Werning, 2017). Ein professionell gestalteter pädagogisch-diagnostischer Prozess kann Schüler*innen wie Ricardo oder Sarah damit nicht nur viel Leid ersparen, sondern ihnen v. a. Möglichkeiten der gesellschaftlichen Teilhabe ermöglichen.

I Theoretischer Rahmen

1 Diagnostische Tätigkeiten im Lehramt

Dieses Kapitel gibt einen Überblick über diagnostische Tätigkeiten, die Lehrkräfte im schulischen Kontext ausführen. Es wird gezeigt, wie wichtig diese diagnostischen Tätigkeiten im Rahmen pädagogischen Handelns sowie als Forschungsgegenstand sind. Die Besonderheiten pädagogischer Diagnostik werden herausgearbeitet und es wird verdeutlicht, dass im Rahmen pädagogischer Diagnostik verschiedene Arten von Diagnosen erstellt werden können. Die Verarbeitung diagnostischer Informationen in bestimmten Situationen wird über verschiedenartige diagnostische Prozesse nachgezeichnet, die zu Diagnosen unterschiedlicher Güte führen können.

1.1 Relevanz diagnostischer Kompetenzen

1.1.1 Diagnostische Kompetenzen tangieren Schülerinnen und Schüler

Eine fundierte Ausbildung im Bereich der Diagnostik ist ein zentrales Anliegen der Lehrkräftebildung, für das inzwischen verschiedene umfangreiche Lehrbücher vorliegen (z. B. Hesse & Latzko, 2017; Jürgens & Lissmann, 2015). Tatsächlich sind diagnostische Kompetenzen von Lehrkräften *entscheidend für das Leistungsniveau* ganzer Schulklassen (Anders et al., 2010) sowie für die Lernentwicklung einzelner Schülerinnen und Schüler (Förster & Souvignier, 2017). Diagnostische Kompetenzen stehen in einem engen Zusammenhang mit didaktischen Kompetenzen und ihr Zusammenspiel kann den Lernerfolg von Schüler*innen entscheidend beeinflussen. Beispielsweise lernen Schüler*innen besonders dann erfolgreich, wenn eine hohe Diagnosekompetenz mit einer starken Unterrichtsstrukturierung einhergeht – und umgekehrt profitieren Schüler*innen kaum von einem hoch strukturierten Unterricht, wenn ihre Lehrkraft über eine gering ausgeprägte Diagnosekompetenz verfügt (Schrader & Helmke, 1987).

Diagnostische Kompetenzen sind daher entscheidend für eine adaptive Unterrichtsgestaltung (Bruder et al., 2010): Lehrkräfte, die eine hohe diagnostische Kompetenz zeigen, haben die „Fähigkeit, das didaktische und fachdidaktische Repertoire flexibel zu handhaben und adaptiv an die Problemlagen der einzelnen Schüler anpassen zu können“ (Hesse & Latzko, 2017, S. 94). Sie können also die Bedürfnisse und Potenziale ihrer Schüler*innen besser erkennen und dementsprechend ihre individuellen Lernprozesse optimal unterstützen. Eine bedarfs- und entwicklungsorientierte, adaptive Unterrichtsgestaltung ist insbesondere im Rahmen inklusiver Beschulungsprozesse nötig, in denen die Schüler*innen individualisierte Lernwege und Förderangebote bekommen (Lenske et al., 2015). Im Rahmen der inklusiven Beschulung von Schüler*innen sind diagnostische Tätigkeiten als ein ganz zentraler Bestandteil der Qualität unterrichtlicher Tätigkeiten anzusehen: Hier werden nicht nur individualisierte Lehrpläne, sondern auch Förder- und Entwicklungspläne erstellt, die in besonderem Maße diagnostische Kompetenzen erfordern (Schäfer & Rittmeyer, 2015). Der Ansatz der *Aptitude-Treatment-Interaction* (Cronbach & Snow, 1977) aus dem Bereich der Sonderpädagogik basiert auf dieser Wechselwirkung zwischen den Fähigkeiten der Lernenden einerseits sowie individuell ausgewählten Methoden andererseits. Diagnostische Tätigkeiten von Lehrkräften tragen dazu bei, dass die Fähigkeiten

der Lernenden berücksichtigt werden können, und begünstigen damit positive Lernentwicklungen. So konnten verschiedene Studien an sonderpädagogisch geförderten Schüler*innen zeigen, dass das Erstellen von Lernverlaufsdiagnostiken positive Effekte auf die Lernverläufe dieser Schüler*innen hat, und dass diese höhere Lernzuwächse erzielen, wenn die Lehrinhalte und –methoden an ihre Lernvoraussetzungen angepasst werden (Stecker et al., 2005). Auch für die Förderung der Lesekompetenzen im Regelunterricht konnte unlängst dargestellt werden, dass sich diagnostische Tätigkeiten der Lehrkraft positiv auf die Kompetenzentwicklung ihrer Schüler*innen auswirken (vgl. Förster & Souvignier, 2017). Hier zeigte sich, dass sowohl die Durchführung einer Lernverlaufsdiagnostik an sich (Souvignier & Förster, 2011) als auch die auf Basis der diagnostischen Informationen implementierte Leseförderung (Förster & Souvignier, 2016) zu einer Kompetenzsteigerung auf Seiten der Schüler*innen führen können. Für Schüler*innen mit Entwicklungsverzögerungen im Bereich des Lesens oder Rechnens ist eine frühzeitige Diagnostik sogar essentiell, um das Auswachsen dieser Verzögerung zu einer manifesten Störung zu verhindern (Ise et al., 2012): Möglichst frühzeitig eingeleitete Interventionen können schwerwiegende psychosoziale Folgen bei den Betroffenen abmildern, die umso wahrscheinlicher werden, je später eine Lese-Rechtschreib- oder Rechenschwäche erkannt wird. Erfreulicherweise stehen aktuell eine Reihe von Verfahren zur Verlaufsdiagnostik zur Verfügung, die die Lernentwicklung von Schüler*innen mit unterschiedlichen Bedarfen sichtbar machen und so eine optimale Unterstützung der Lernprozesse ermöglichen (s. Hasselhorn et al., 2014).

1.1.2 Diagnostische Kompetenzen als aktuelles Forschungsdesiderat

Obwohl die diagnostischen Fähigkeiten von Lehrkräften einen entscheidenden Einfluss auf die Entwicklung ihrer Schülerinnen und Schüler haben, wurden sie in der gesellschaftlichen und wissenschaftlichen Wahrnehmung eine lange Zeit vernachlässigt (Artelt & Gräsel, 2009; Hesse & Latzko, 2017) und in den letzten Jahren wieder zunehmend „ins Bewusstsein gerückt“, weil deutsche Schüler*innen in internationalen Schulleistungsvergleichen negativ aufgefallen seien (Hesse & Latzko, 2017, S. 9). Als Reaktion auf die schlechten Ergebnisse bei internationalen wie nationalen Leistungsmessungen an deutschen Schüler*innen hat die Kultusministerkonferenz (KMK, 2004) Standards für Lehrkräftebildung veröffentlicht, die die Beurteilungskompetenzen von Lehrkräften als eine von vier zentralen professionellen Kompetenzen definieren. Die nötigen Voraussetzungen, um die einzelnen diagnostischen Aufgaben kompetent lösen zu können, sollen die Lehrkräfte in allen drei Phasen der Lehrkräftebildung erwerben und dadurch die Güte ihrer Urteile erhöhen. Cordula Artelt und Cornelia Gräsel (2009) benennen im Editorial eines Sonderbandes der Zeitschrift für Pädagogische Psychologie allerdings eine offensichtliche Kluft zwischen den Forderungen der KMK und dem aktuellen Forschungsstand zu den Beurteilungskompetenzen: „Der Forderung nach Stärkung der Güte diagnostischer Urteile steht jedoch ein wissenschaftlicher Forschungsstand gegenüber, der hinsichtlich der Entstehung, der Förderbarkeit und der Auswirkungen diagnostischer Kompetenz noch unbefriedigend ist.“ (ebd., S. 157) Als offene und relevante Fragen, die wissenschaftlich zunächst geklärt werden müssten, formulierten sie vor einem Jahrzehnt, dass bislang wenig darüber bekannt sei, 1) wie und wann

diagnostische Kompetenz im Verlauf der professionellen Entwicklung entstehe, 2) wie sich diagnostische Kompetenz fördern lasse, 3) welche Auswirkungen sie für die Entwicklung der Lernleistungen bei den Schülerinnen und Schülern habe und 4) welches Wissen der diagnostischen Kompetenz zugrunde liege (ebd., S. 159).

Die im Themenheft versammelten Beiträge seien dementsprechend nur ein Anfang der notwendigen Forschungsbemühungen. Knapp zehn Jahre später erschien eine Sammelschrift, die die seitdem vollzogenen Forschungsbemühungen zur diagnostischen Kompetenz von Lehrkräften zusammenfasst (Südkamp & Praetorius, 2017). Dort wird betont, dass die Erforschung diagnostischer Kompetenzen nach wie vor ein *Desiderat der Forschung* sei, auch wenn inzwischen verschiedene Ansätze zur Erfassung und Förderung diagnostischer Kompetenz vorlägen. Zentrale Forschungsdesiderata seien aktuell 1. die Definition und Konzeptualisierung diagnostischer Kompetenz als Konstrukt, 2. die Bereitstellung umfassender Erfassungs- und Analysemethoden und 3. die weitere Entwicklung von Förderansätzen und damit verbundene Evaluationsstudien.

Diese und weitere Fragestellungen werden aktuell in verschiedenen Projekten der „Offensive Lehrerbildung“ (BMBF, 2016) oder innerhalb des eigens zur Untersuchung diagnostischer Kompetenzen ins Leben gerufenen Wissenschaftsnetzwerks NeDiKo („wissenschaftliche Netzwerk zur diagnostischen Kompetenz von Lehrkräften“, 2014-2016, Südkamp & Praetorius, 2017) untersucht. Die diagnostischen Kompetenzen scheinen dabei einer derjenigen Aspekte professioneller Kompetenz von Lehrkräften zu sein, die am stärksten diskutiert werden (s. a. Baumert & Kunter, 2006). Ingrid Hesse und Birgit Latzko (2017) sprechen gar von einer „Renaissance der pädagogisch-psychologischen Diagnostik im Lehrerberuf“ (S. 15), da das Erkennen und die Akzeptanz der Notwendigkeit pädagogischer Diagnostik aktuell einen massiven Aufschwung erlebe.

Diese Ausführungen machen deutlich, dass in der Lehrkräftebildung empirisch überprüfte Ansätze zur Erfassung und Förderung diagnostischer Kompetenzen noch fehlen. Die Ausformulierung, Definition und Modellierung diagnostischer Kompetenzen sind dabei eine besondere Herausforderung, die für die Erfassungs- und Fördermöglichkeiten weitreichende Implikationen hat (vgl. Südkamp & Praetorius, 2017). Diese Arbeit soll einen Beitrag dazu leisten, die Ausformulierung und Modellierung dieser Kompetenz voranzutreiben. Zu diesem Zweck soll zunächst ein Überblick über pädagogisch-diagnostische Aufgaben und Vorgehensweisen von Lehrkräften im Rahmen ihres pädagogischen Handelns gegeben werden.

1.2 Diagnostische Aufgaben und Vorgehensweisen von Lehrkräften

1.2.1 Diagnostische Aufgaben von Lehrkräften

Hesse & Latzko (2017) machen in ihrem Lehrbuch der Diagnostik für Lehrkräfte bereits im Vorwort deutlich, dass es Aufgabe der Lehrkräfte sei, sowohl mögliche „Schwierigkeiten beim Lernen frühzeitig erkennen und minimieren, vor allem aber die Stärken der Kinder bemerken und entwickeln“ (S. 9) zu können. Die Feststellung von Lernvoraussetzungen sowie die Beurteilung von Schulleistungen seien

verantwortungsvolle Aufgaben von Lehrkräften, die einerseits der Feststellung von Lernfortschritten und andererseits auch der „ethisch verantwortungsvollen Lenkung von Schullaufbahnen“ (S. 10) dienen.

Der Abschlussbericht der KMK (2004, 2014) zur Frage der „Perspektiven der Lehrerbildung in Deutschland“ formulierte ein Leitbild diagnostischer Tätigkeiten, das zwei Aspekte der sogenannten „*Beurteilungsaufgabe*“ von Lehrkräften herausstellt (vgl. auch Jürgens & Lissmann, 2015):

- (1) Die Beurteilungsaufgabe müsse einerseits *lern- und entwicklungsorientiert* erfolgen, d. h. *Lernvoraussetzungen und Lernprozesse* der Schüler*innen sollen festgestellt und entsprechende Fördermöglichkeiten abgeleitet werden. Hierzu müssen Lernpotentiale, Lernhindernisse und Lernfortschritte erkannt werden. Neben der Feststellung von Lernvoraussetzungen und –prozessen sollen auch Formen von Hoch- und Sonderbegabungen, Lern- und Arbeitsstörungen sowie anderer Begabungen und Möglichkeiten der Begabungsförderung erkannt werden. Zum „Kompetenzbereich Beurteilen“ gehöre auch, die Lernenden sowie deren Eltern hinsichtlich der untersuchten Fragestellungen zu beraten. Und schließlich umfasst kompetentes diagnostisches Verhalten, bei der Diagnostik, Förderung und Beratung mit inner- oder außerschulischen Kolleg*innen oder anderen Professionellen zusammenzuarbeiten.
- (2) Die Beurteilungsaufgabe müsse andererseits *methodisch orientiert* erfolgen, d. h. die *Leistungsbeurteilung* soll transparent und methodisch begründet erfolgen. Leistungsbeurteilungen sollen ggf. in unterschiedlichen Formen vorgenommen werden, sich evtl. auf verschiedene Bezugssysteme stützen und in einer Form zurückgemeldet werden, die für die Beurteilten nachvollziehbar ist. Außerdem sollen Lehrkräfte die Leistungsüberprüfungen als Rückmeldung für ihren eigenen Unterricht nutzen können.

Während sich der erste Aspekt im Wesentlichen auf die Funktion der optimalen Unterstützung der Schüler*innen bezieht, die durch die Beurteilung der Lernvoraussetzungen und –fortschritte ermöglicht werden soll, beschreibt der zweite Aspekt eine Aufgabe, die mit Diagnostik als „Feststellung“ im engeren Sinne nichts zu tun hat. Hier wird deutlich, dass die Beurteilungsaufgaben von Lehrkräften nicht nur der Unterstützung von Lernvoraussetzungen und –prozessen dienen, sondern auch der *Bewertung von erreichten Leistungen*, beispielsweise in Form von Noten (Jürgens & Lissmann, 2015; Retelsdorf & Möller, 2016; Sacher, 2009). In dieser Arbeit wird die erstgenannte Funktion, nämlich die Erfassung oder Feststellung von Lernvoraussetzungen, –prozessen und –ergebnissen, in den Blick genommen. Mit dieser Funktion sind verschiedene diagnostische Tätigkeiten innerhalb der schulischen Praxis verbunden, die der optimalen Gestaltung pädagogischer Handlungen dienen. Diese besondere Form der Diagnostik wird unter dem Begriff der „pädagogischen Diagnostik“ verstanden, deren Begriff, Aufgaben und Zielsetzungen nachfolgend näher erläutert werden.

1.2.2 Pädagogische Diagnostik als besondere Form der Diagnostik

Der *Begriff der Diagnostik im Allgemeinen* umschreibt das Erstellen von Diagnosen, das – je nach Disziplin – bestimmten Kriterien oder Regeln folgt. Die psychologische Diagnostik etwa widmet sich der Erfassung

„individueller, situationsgebundener oder entwicklungsbedingter Unterschiede im Erleben und Verhalten“ (Pawlik, 2006; S. 13), deren methodisches Kernstück psychologische Tests sind. Diagnosen sind Ergebnisse diagnostischer Handlungen, die Aussagen über einen Beurteilungsgegenstand beinhalten, welche beurteilt werden. Helmke (2009) arbeitet heraus, dass Diagnosen mehr als bloße Beschreibungen sind, denn „das charakteristische Merkmal einer Diagnose liegt darin, dass anhand vorgegebener Kategorien, Begriffe oder Konzepte *geurteilt* wird“ (S. 122, Hervorhebung im Original). Damit grenzt er den Begriff der Diagnostik von schieren Einschätzungen ab, unter denen er ein subjektives Urteil über einen Sachverhalt, z. B. in Form eines kategorienbasierten Rating-Bogens“ versteht. Jene Diagnosen, die „auf der Basis anerkannter wissenschaftlicher Theorien bewusst, gezielt und planvoll erfolgen“ (Kretschmann, 2015, S. 9), können demgegenüber als „*wissenschaftliche Diagnosen*“ bezeichnet werden und finden sich in der Medizin, der Psychologie und der Pädagogik. Unter Diagnosen können in diesem professionellen Sinne Feststellungen oder Beurteilungen verstanden werden, die von Theorien und Hypothesen geleitet werden, der Erkenntnisgewinnung oder der Entscheidungsfindung dienen und die daher wissenschaftlichen Gütekriterien genügen sollten (vgl. Helmke, 2009, S. 268). Dieser Sorgfaltsanspruch ergibt sich insbesondere dadurch, dass Diagnosen – im Kontext der Medizin oder Psychologie immer, und im Rahmen pädagogischen Handelns zumeist – auf Personen bezogen sind:

„Diagnosen stellen in der Regel explizite Aussagen über Zustände und Merkmale von Personen dar, die Ergebnis eines reflektiert und methodisch kontrolliert durchlaufenen diagnostischen Prozesses sind.“ (Schrader, 2010, S. 102)

Pädagogische Diagnosen sind spezifische Aussagen über Lernende, deren Lernvoraussetzungen, -prozesse und -ergebnisse. Das Besondere an pädagogischen Diagnosen ist, dass hier auch Aufgaben zum Beurteilungsgegenstand werden können (vgl. Artelt & Gräsel, 2009). Der Begriff der Diagnose umfasst in der *Pädagogischen Diagnostik* also nicht nur Aussagen über die Personenfähigkeit, sondern wird erweitert auf Aussagen über die Aufgabenschwierigkeit bzw. die Erfassung der Passung zwischen Personenfähigkeit und Aufgabenschwierigkeit zugunsten eines Vorgehens, das unterstützend und entwicklungsbezogen ist (Weinert, 2000 zitiert nach Jürgens & Lissmann, 2015, S. 21).

Dass sich die diagnostischen Tätigkeiten in der pädagogischen Praxis explizit auf pädagogische Fragestellungen und Entscheidungen beziehen, unterscheidet die *pädagogische Diagnostik* von der psychologischen Diagnostik (Ingenkamp & Lissmann, 2008). Während die Standards, Methoden und Verfahren in beiden Disziplinen vergleichbar sind, liegt die Besonderheit pädagogischer Diagnostik darin, dass sie im schulischen Kontext als Methode zum begründeten Fällen bestimmter Entscheidungen verstanden wird (Hesse & Latzko, 2017; Jürgens & Lissmann, 2015). Pädagogische Diagnostik kann entsprechend definiert werden als „das systematische Sammeln, Dokumentieren und Verdichten von Informationen über Lernende, um diese in Hinblick auf verschiedene Merkmale, die den Unterricht und Schulalltag betreffen (z. B. Leistungen, Verhalten, Interesse, Motivation, Emotionen), unterteilen und einordnen zu können. Pädagogische Diagnostik erfüllt hierbei jedoch keinen Selbstzweck, sondern sie soll

Entscheidungen über das pädagogische und didaktische Handeln im Rahmen der professionellen Aufgaben einer Lehrkraft *begründen und unterstützen*“ (Behrmann & van Ophuysen, 2017, S. 39, Hervorhebungen v. Verf.).

Dass pädagogisch-diagnostische Tätigkeiten üblicherweise auch auf die Förderung von Schüler*innen bezogen sind, ist eine zweite Besonderheit pädagogischer Diagnostik, die mit der Ausrichtung auf pädagogische Entscheidungen verbunden ist. Pädagogische Diagnostik lässt sich Hesse & Latzko (2017) zufolge verstehen als eine „Kategorisierung oder Einordnung des Einzelfalls“, die erstens der Optimierung von Bildungs- und Erziehungspraktiken dient, und zweitens auch der „Zuweisung zu Lerngruppen oder individuellen Förderprogrammen“. Pädagogische Diagnostik umfasst demnach vor allem die Feststellung von Lernvoraussetzungen, –prozessen und –ergebnissen mit dem Ziel, der Optimierung individuellen Lernens – Diagnostik und Förderung gehen in der Schule „Hand in Hand“ (Hesse & Latzko, 2017, S. 18).

Eine weitere Besonderheit pädagogischer Diagnostik ergibt sich schließlich dadurch, dass ein Teil der pädagogischen Entscheidungen nicht den formalisierten Regeln der psychologischen Diagnostik folgt, sondern von der regelgeleiteten Vorgehensweise abweichen kann (Helmke, 2009; Schrader, 2010). Die diagnostischen Aufgaben von Lehrkräften gestalten sich in Abhängigkeit von den spezifischen diagnostischen Zielsetzungen sogar recht unterschiedlich und ihre Umsetzung kann daher mehr oder weniger elaboriert ausfallen. Im folgenden Abschnitt wird eine Übersicht gegeben, welche diagnostischen Situationen oder Aufgabenstellungen unterschieden werden können und wie sich das diagnostische Vorgehen entsprechend unterschiedlich gestaltet.

1.2.3 Verschiedene Arten pädagogischer Diagnosen

Für eine Lehrkraft ergeben sich im Rahmen ihrer Tätigkeit verschiedenste Anlässe, in denen Diagnostik eine Rolle spielt. Ganz grob können diese Anlässe danach unterschieden werden, ob sie implizite oder explizite Diagnosen nach sich ziehen (Hesse & Latzko, 2017; Jürgens & Lissmann, 2015; Latzko & Hesse, 2009). *Explizite Diagnosen* sind objektive Beurteilungen, die bewusst reflektiert werden und für die möglichst viele Informationen zusammengetragen werden, um sie kritisch hinsichtlich ihrer Relevanz für die Entscheidung zu prüfen. *Implizite Diagnosen* wiederum sind Urteile, die nicht absichtlich oder methodisch kontrolliert entstehen, und bei denen subjektiv bedeutsame Indikatoren eher routiniert registriert werden. Zu den Anlässen, die eine explizite Diagnostik erfordern, zählen Hesse & Latzko u. a. die Feststellung von Lernvoraussetzungen, die Feststellung der Ausgangslage für längerfristige Förder- oder Nachhilfemaßnahmen, die Diagnostik bei Schullaufbahnentscheidungen oder die Diagnostik von Lernschwierigkeiten. Diese Diagnoseanlässe sind mit weit reichenden Konsequenzen verbunden.

1.2.3.1 Diagnostische Situationen und Anlässe

Diagnoseanlässe, die langfristige Folgen haben, werden von Karina Karst und Kolleg*innen (Karst et al., 2017) auch als „verbindlich“ bezeichnet. Die Verbindlichkeit ist in ihrer *Klassifikation diagnostischer Situationen* eines von vier Kriterien, anhand derer sie diagnostische Situationen strukturieren. Als relevant dafür, welcher diagnostische Weg eingeschlagen wird, erachten sie (1) den Zweck der Diagnostik, (2) die

Planbarkeit, (3) die Verbindlichkeit und Konsequenz und (4) die Perspektive auf die oder den Lernenden. Das Merkmal der Planbarkeit bezieht sich darauf, wie langfristig die Diagnostik an- und umgesetzt wird. Kurzfristige Feststellungen betreffen beispielsweise die tagtägliche Unterrichtsplanung oder die Überprüfung des Lernstandes im Rahmen einer Klassenarbeit, langfristige Untersuchungen über längere Zeiträume kommen beispielsweise bei der Lernverlaufsdiagnostik über ein Schuljahr zum Einsatz (Shavelson, 2006). Das Kriterium der *Verbindlichkeit* meint, dass die Feststellung eines Merkmals direkt oder weniger direkt an eine Entscheidung gebunden sein kann, d. h. dass eine Diagnostik entweder direkte Folgen haben kann (z. B. Zuordnung zum schulischen Förderunterricht) oder aber ausschließlich der Orientierung dient und keine Entscheidung folgt (z. B. Schulleistungsstudie ohne Implikationen). Mit der Verbindlichkeit einer Diagnose gehen unterschiedliche Grade an Konsequenz einher, d. h. die Folgen der Diagnostik für die Betroffenen sind unterschiedlich bedeutsam. Eine hohe Verbindlichkeit stehe den Autor*innen zufolge immer mit weitreichenden Konsequenzen in Zusammenhang. Weitreichende Konsequenzen hat eine Diagnose dann, wenn die darauf beruhende Entscheidung langfristige Auswirkungen hat, oder die Entscheidung sogar irreversibel ist, wie es beispielsweise bei Übergangentscheidungen der Fall ist. Die Autor*innen vermuten, dass verbindliche Entscheidungen immer dann entstehen, wenn es um Zuweisungen zu bspw. Lerngruppen oder Bildungswegen geht (Selektionsdiagnostik), und dass mit zunehmender Verbindlichkeit auch der Rechtfertigungsdruck der diagnostizierenden Person steigt.

Ein weiteres zentrales Beschreibungs- und Unterscheidungsmerkmal für diagnostische Situationen ist die Handlungsstrategie, der eine Diagnostik folgt (Jürgens & Lissmann, 2015): Diagnosen können einer *Selektions- oder Modifikationsstrategie* folgen. Das diagnostische Ziel bei der Selektionsstrategie ist die Feststellung von Lernvoraussetzungen oder Leistungen, damit die lernende Person einer Bildungsmaßnahme, -gruppe oder -institution zugewiesen werden kann. Schüler*innen mit Lernstörungen können beispielsweise auf Basis ihrer Ergebnisse in Lernstandserhebungen dem schulischen Förderunterricht zugewiesen werden. Bei der Modifikationsstrategie steht nicht die Auswahl einer passenden Lernumgebung im Zentrum, sondern die optimale Gestaltung einer vorhandenen Lernumgebung, beispielsweise die Anpassung des Regelunterrichts an die Bedürfnisse von Kindern mit Lernstörungen. Jürgens und Lissmann (ebd.) differenzieren die Selektionsentscheidungen zusätzlich dahingehend, ob die Entscheidungsalternativen als gleichwertig oder nicht gleichwertig hinsichtlich des Ziels angesehen werden können. Typischerweise sind die Entscheidungsalternativen bei bspw. Schullaufbahneempfehlungen nicht gleichwertig und daher gut gegeneinander abzuwägen.

Diese Zusammenstellung diagnostischer Situationen veranschaulicht, dass es nicht eine einheitliche Art und Weise der pädagogischen Diagnostik geben kann, sondern dass es vielmehr verschiedene diagnostische Vorgehensweisen geben muss. Je nachdem, wie weitreichend die diagnostische Entscheidung ist oder wie stark die Ziele einer Selektionsentscheidung voneinander abweichen, können die diagnostischen Prozeduren mehr oder weniger explizit, aufwändig und formalisiert ablaufen.

Nachfolgend werden drei grundlegende Diagnosearten vorgestellt, die sich hinsichtlich ihres Formalisierungsgrades unterscheiden.

1.2.3.2 Unterscheidung von Diagnosen nach Formalisierungsgrad

Für Diagnosen, die mit schwerwiegenden Entscheidungen einhergehen, sind explizite (Hesse & Latzko, 2017) oder regelgeleitete Diagnosen (Helmke, 2009) erforderlich, die entsprechend einen hohen Formalisierungsgrad aufweisen. Diese regelgeleiteten Diagnosen können von solchen Diagnosen abgegrenzt werden, die „eher beiläufig und unsystematisch im Rahmen des alltäglichen erzieherischen Handelns gewonnen werden“ (S. 122). Diese zwei grundsätzlichen Arten der pädagogischen Diagnosen werden auch als formelle vs. informelle Diagnosen bezeichnet und markieren Diagnosen höherer oder geringerer Güte (Hascher, 2008; Schrader, 2010). *Formelle Diagnosen* sind zielgerichtet, werden durch eine systematische Herangehensweise unter Verwendung evidenzbasierter Verfahren gewonnen entsprechen dem oben formulierten wissenschaftlichen Anspruch. *Informelle Diagnosen* hingegen bilden eher subjektive und implizite Eindrücke ab, die nicht zielgerichtet, systematisch oder mit erprobten Methoden gewonnen werden. Jede Form der Diagnose hat dabei ihre Stärken und Schwächen und keine ist per se zu bevorzugen: Mithilfe formeller Diagnostik können Tina Hascher (2008) zufolge ganz spezifische Fragen seriös abgeklärt werden – dies erfordere aber „fundierte Kenntnisse über erprobte Methoden“ sowie bestimmte schulische Rahmenbedingungen, die dieses Vorgehen erlauben. Die informelle Diagnostik, die überwiegend auf Beobachtungen oder Erfahrungen basiere, lasse sich einfacher in die Praxis integrieren. Sie sei aber wenig bewusst, eher routiniert und daher anfällig für Fehler. Schrader (2010) formuliert daher deutlich, dass informelle Diagnosen nicht für sich stehen können, wenn explizite Aussagen über Personen getroffen werden sollen: Sie stellen hierbei wertvolle diagnostische Hypothesen dar, die im Zuge eines hypothesengeleiteten diagnostischen Prozesses überprüft werden (Schrader, 2010, S. 108).

Hascher (2008) erweitert das Begriffskontinuum um die Kategorie der *semiformellen Diagnose*, unter der sie diagnostische Handlungen versteht, die „nicht den Kriterien formeller Diagnosen genügen, aber nicht nur zu impliziten Urteilen führen“ (S. 75). Ihr reicht die Unterscheidung in formelle und informelle Diagnosen nicht aus, um die diagnostischen Tätigkeiten von Lehrkräften differenziert zu erfassen. Häufig fände man in der Schule Mischformen aus diesen beiden Arten, die weder formell noch richtig informell seien, wie etwa gezielte Beobachtungen, die aber nicht mit erprobten Methoden erfolgen, oder der Einsatz erprobter Methoden, der aber ungezielt und unsystematisch erfolge. Auch komme es vor, dass intuitive Beobachtungen festgehalten werden und dann explizit in die Bewertung mit einfließen. Hascher formuliert recht drastisch, dass semiformelle Diagnosen eine inakzeptable Form der Diagnostik seien:

„Die Etablierung der semiformellen Diagnostik im Schulalltag ist aus meiner Sicht ein Indikator hoher Motivation bzw. Notwendigkeit in Kombination mit unzureichender Kompetenz: Lehrpersonen bemühen sich um diagnostische Schritte trotz ihres Defizits an entsprechenden Kenntnissen“ (Hascher, 2008, S. 75).

Dieses Verständnis von Diagnostik geht davon aus, dass Diagnosen eine große Reichweite für diejenigen haben können, die diagnostiziert werden. Um Fehleinschätzungen zu vermeiden, die falsche Interventionen nach sich zögen, müssen sie sorgfältig erstellt werden. Dementsprechend können formelle Diagnosen als der Königsweg der Diagnostik angesehen werden, wenn explizite Aussagen über Eigenschaften oder Zustände von Personen getroffen werden – in solchen Fällen müssen sie stets das Ergebnis eines „reflektiert und methodisch kontrolliert durchlaufenen diagnostischen Prozesses“ (Schrader, 2010, S. 102) sein. In Tabelle 1.1 werden die unterschiedlichen Arten von Diagnosen, die sich durch ihren Grad an Formalität unterscheiden, noch einmal zusammengefasst. Dort werden die genannten Diagnosearten hinsichtlich der Systematik und Zielgerichtetheit des Vorgehens sowie der eingesetzten Methoden unterschieden.

Tabelle 1.1 Unterscheidung von Diagnosearten nach dem Formalisierungsgrad

Diagnostische Information	Informationsgewinnung	Diagnosefokus	Diagnosemethoden und -verfahren
informell	unsystematisch, beiläufig	ungerichtet, ungezielt	unerprobt, subjektiv
formell	systematisch, kriteriengeleitet	gerichtet, gezielt	erprobt, objektiv
semiformell	unsystematisch, beiläufig	ungerichtet, ungezielt	erprobt, objektiv
	systematisch, kriteriengeleitet	gerichtet, gezielt	unerprobt, subjektiv

Anmerkungen: Die Darstellung der Diagnosearten orientiert sich an der Beschreibung von Hascher (2008). Semiformelle Diagnosen stellen eine besondere Kategorie dar, da sie Elemente formeller und informeller Diagnostik verbinden. Dementsprechend sind hier (mindestens) zwei Entwürfe zur Beschreibung jener Diagnosen möglich.

1.2.3.3 Automatisierte und kontrollierte Verarbeitung

Die Unterscheidung von Diagnosen nach Formalisierungsgrad geht implizit davon aus, dass es die grundlegenden Fähigkeiten einer Lehrkraft sind, die darüber bestimmen, welche Art von Diagnose sie oder er auswählt oder erzeugt. Dabei wird nicht berücksichtigt, dass in unterschiedlichen diagnostischen Situationen die kognitiven Ressourcen einer Lehrkraft unterschiedlich stark beansprucht werden. Wenn diagnostische Situationen als eine besondere Form sozialer Situationen betrachtet werden, können in Anlehnung an duale Prozessmodelle sozialer Urteilsbildung (Dual-Process-Theorien; z. B. Kahneman, 2011) auch unterschiedlich aufwändige und ressourcenintensive diagnostische Prozesse angenommen werden (Hetmanek & Van Gog, 2017). Entscheidend für die Gestaltung des diagnostischen Vorgehens könnte dabei sein, ob das Urteil schnell oder akkurat sein soll (Böhmer et al., 2015).

Schnelle Entscheidungen (Typ-1-Diagnosen) werden typischerweise im laufenden Unterrichtsgeschehen getroffen und beziehen sich beispielsweise auf die Feststellung der Lernfortschritte der Schüler*innen. In dieser Situation muss eine Lehrkraft mehrere parallele Anforderungen bewältigen, und die diagnostische

Verarbeitung muss in solchen Situationen daher mit geringem kognitiven Aufwand möglich sein. Jene Typ-1-Diagnosen entstehen im Rahmen heuristischer oder automatisierter Prozesse, die unabhängig von aktiver kognitiver Kontrolle und teilweise sogar parallel ablaufen können. Die Nachteile dieser Verarbeitung sind eine geringe Genauigkeit und eine Anfälligkeit für Fehler, die durch Übergeneralisierung entstehen.

Akkurate Entscheidungen (Typ-2-Diagnosen) hingegen werden getroffen, wenn die Entscheidungen für die betroffenen Schüler*innen hochrelevant sind, z. B. bei Übertrittsempfehlungen. In diesen Fällen werden die Informationen systematisch, methodisch kontrolliert und dem Bewusstsein zugänglich verarbeitet, um Fehldiagnosen zu vermeiden. Eine derart sorgfältige Verarbeitung läuft häufig seriell und sprachbasiert ab und ist mit hohem kognitiven wie zeitlichem Aufwand verbunden. Die kognitive Belastung entsteht dadurch, dass automatisierte Reaktionen (Typ-1-Prozesse) unterdrückt und überschrieben sowie alternative Lösungen entwickelt werden müssen.

Anhand der Beschreibung wird deutlich, dass beide Prozesse ihre jeweiligen Vorzüge und Nachteile haben und daher nicht pauschal ein Prozess für jede Situation angemessen ist. Vielmehr muss die handelnde Person in der jeweiligen Situation auswählen, welche Reaktion ein zur Situation passendes Verhalten darstellt. Um Typ-2-Diagnosen erstellen zu können, müssen Lehrkräfte außerdem über die entsprechenden kognitiven Voraussetzungen verfügen. Entsprechend des Ausmaßes kognitiver Ressourcen, die einer Lehrkraft generell sowie in der jeweiligen diagnostischen Situation zur Verfügung stehen, wird der diagnostische Prozess mehr oder weniger heuristisch vs. kontrolliert erfolgen (Böhmer et al., 2017).

1.2.4 Der diagnostische Prozess

1.2.4.1 *Verschiedene Wege führen zur Diagnose*

Die vorausgehenden Ausführungen verdeutlichen, dass der diagnostische Prozess, der in einer bestimmten diagnostischen Situation ausgelöst wird, unterschiedlich automatisiert vs. kontrolliert ablaufen kann. Eine zentrale Herausforderung für Lehrkräfte besteht darin, angesichts der Vielzahl diagnostischer Situationen (vgl. z. B. Karst et al., 2017; vgl. Kapitel 1.2.3.2) und Anlässe den passenden diagnostischen Prozess auszuführen. Stephanie Herppich und Kolleg*innen (2017; 2018) entwerfen ein diagnostisches Prozessmodell (vgl. Abb. 1.1), das diese Diversität aufgreift, indem es *verschiedene Wege der Diagnose* integriert. Das Modell beschreibt eine Reihe von Schritten im diagnostischen Prozess von Lehrkräften, die je nach Zielstellung von der diagnostizierenden Person durchgeführt bzw. ausgelassen werden können. Der kürzeste Weg zur Diagnose wird hier als Nutzung bereits vorliegender Informationen modelliert, die u. U. auch beiläufig und ohne explizite Zielstellung erhoben wurden (informelle Diagnose nach Schrader, 2010). Auf diesem kürzesten Weg werden vier Schritte durchlaufen, die bei jeder Diagnostik durchlaufen werden:

- (1) Festlegen des diagnostischen Ziels,
- (2) Verarbeitung bereits vorhandener Daten und Informationen,
- (3) Entscheidung, dass entsprechend des Ziels keine weiteren Daten erforderlich sind,
- (4) Diagnose erstellen.

Der längste Weg zur Diagnose sieht das Durchlaufen einer festen Schrittfolge vor, um zu einer reliablen, objektiven und validen Diagnose zu gelangen (formelle Diagnose nach Schrader, 2010). Diese Schrittfolge wird folgendermaßen spezifiziert:

- (1) konkrete Fragestellung ableiten,
- (2) mehrere prüfbare Hypothesen entwickeln,
- (3) Operationalisierung des Konstrukts und Auswahl geeigneter Methoden,
- (4) Erhebung diagnostischer Daten,
- (5) regelgeleitete Auswertung,
- (6) regelgeleitete Integration zur Prüfung der Hypothesen,
- (7) erneutes (teilweises) Durchlaufen des Prozesses bei widersprüchlichen Informationen, die schwer vereinbar sind, oder bei Begleitung einer Intervention.

Dass alle sieben Schritte vollständig durchlaufen werden, ist diesem Modell zufolge in der Praxis nicht zwangsläufig erforderlich. Je nach Zielstellung wählen die Lehrkräfte aus, welche Schritte essenziell sind, und gestalten dabei eine mehr oder weniger aufwändige, explizite und methodisch kontrollierte Prozedur. Die erforderliche Schnelligkeit vs. Akkuratheit wird hier als eine Dimension angenommen, die die Gestaltung des diagnostischen Prozesses maßgeblich beeinflusst. Neben der Zielstellung entscheiden auch die Anforderungen der Situation sowie die jeweiligen Fähigkeiten der Lehrkräfte darüber, wie der jeweilige Prozess konkret gestaltet wird. In diagnostischen Situationen sind damit unterschiedliche Entscheidungswege und –abläufe denkbar, mit deren Auswahl der diagnostische Prozess je nach Situation angepasst werden kann. Die Auswahl der passenden Prozedur kann dementsprechend ein Zeichen sein, an der sich die diagnostische Kompetenz von Lehrkräften messen lässt (Herppich et al., 2018; vgl. Abschnitt 2.2.2). Immer wenn der diagnostische Anlass eine explizite oder formelle Diagnose erfordert (Hesse & Latzko, 2017; Schrader, 2010), ist der längste und aufwändigste Weg im diagnostischen Prozessmodell nach Stephanie Herppich et al. (2017) erforderlich. Dieses regelgeleitete Vorgehen psychologischen Diagnostizierens wird im folgenden Kapitel (Kap. 1.2.4.2) vorgestellt.

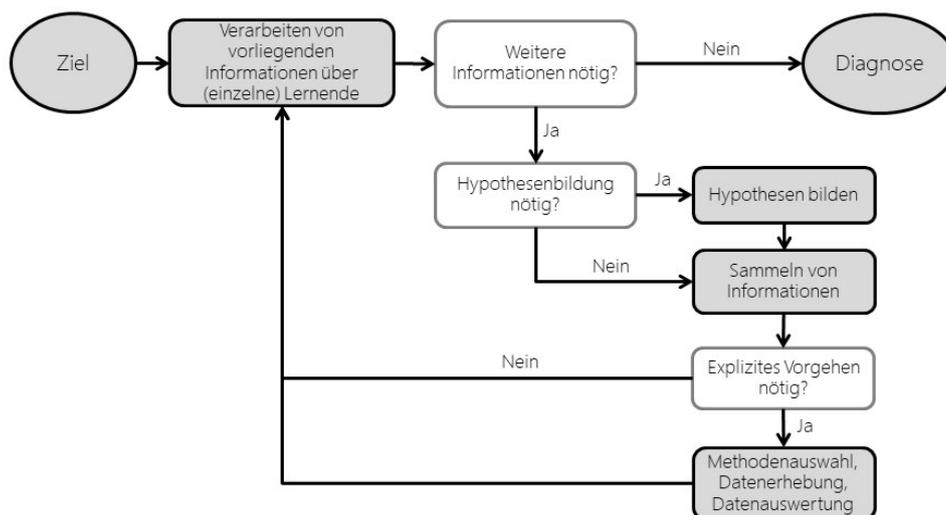


Abbildung 1.1 Der diagnostische Prozess als Modell mit flexiblen Abläufen.
Darstellung in Anlehnung an Herppich et al. (2017, S. 82)

1.2.4.2 Diagnostik als regelgeleiteter Prozess

Das Modell des diagnostischen Prozesses mit flexiblen Abläufen, das im vorausgehenden Abschnitt beschrieben wird, integriert je nach Zielstellung und Situation nicht immer explizit nur objektive, reliable oder valide Informationen – denn das Ziel des Prozesses ist es nicht immer, zu einem Urteil höchster Güte zu kommen. Aber immer, wenn eine explizite Diagnose erforderlich ist, muss sich das diagnostische Vorgehen von alltäglichen und durch Vorurteilen geprägte Urteilsprozesse abheben (Hesse & Latzko, 2017): Eine formelle Diagnose orientiert sich am *wissenschaftlichen Prozess der Hypothesenprüfung* und muss Hesse & Latzko (ebd.) zufolge zwingend bestimmten Qualitätskriterien genügen. Erstens müsse das zu messende Merkmal theoretisch präzisiert und zudem methodisch operationalisiert werden. Die Messmethoden selbst sollen sich an den wissenschaftlichen Gütekriterien der Objektivität, Reliabilität und Validität orientieren. Weiterhin sollen für die Messergebnisse objektive Vergleichsmaßstäbe zur Einordnung individueller Merkmalsausprägungen vorliegen (Normiertheit). Die Messergebnisse sollen ferner vor dem Hintergrund des Untersuchungsverlaufs hinsichtlich möglicher Fehler oder Lücken bei der Datenerhebung reflektiert werden. Das darauf begründete Urteil, das regelgeleitet und methodisch kontrolliert erstellt wurde, soll abschließend hinsichtlich seiner Richtigkeit geprüft werden. Eine Diagnostik, die diesen wissenschaftlichen Kriterien genügt, beschreibt einen komplexen diagnostischen Prozess, der eine feste und systematische Abfolge diagnostischer Schritte formuliert.

Eine prominente Modellierung dieses regelgeleiteten, an wissenschaftlichen Kriterien orientierten Vorgehens, hat Reinhold S. Jäger (2006) entwickelt. Der diagnostische Prozess nach Jäger (vgl. auch Langfeld, 2017) formuliert insgesamt vier Phasen, die einer normativen Reihenfolge nachkommen (vgl. Abb. 1.2). Der Prozess beginnt mit einer Orientierungsphase (1), in der ein zu untersuchendes Problem sowie ein diagnostisches Ziel definiert werden. Zentrales Anliegen ist hier die Ausarbeitung einer *präzisen*

Fragestellung, die das Kriterium für den Endpunkt sowie die Qualität des gesamten diagnostischen Prozesses bildet. Die Formulierung einer wissenschaftlichen Fragestellung ist Grundlage der Operationalisierung und damit auch der Methoden zur Erfassung des Sachverhalts, der geprüft werden soll. Die Fragestellung nimmt entsprechend einen besonderen Stellenwert innerhalb des psychologisch-diagnostischen Prozesses ein und erfordert daher auch besondere Aufmerksamkeit und Sorgfalt.

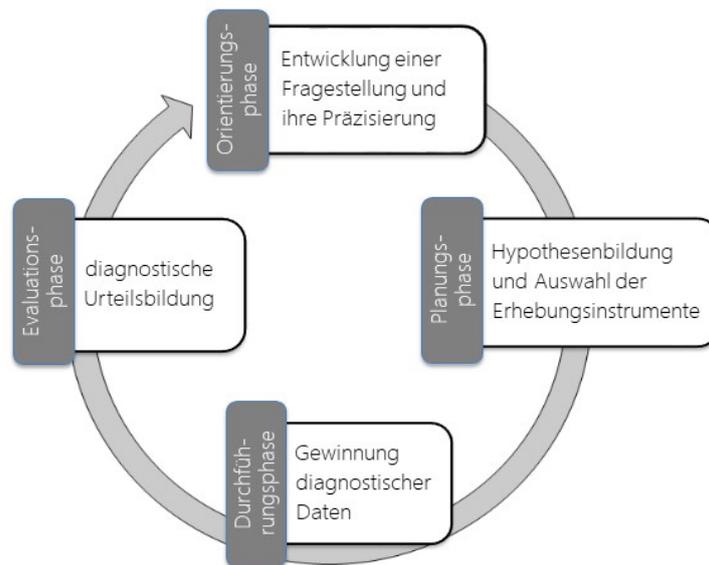


Abbildung 1.2 Der diagnostische Prozess als Modell mit regelgeleitetem Vorgehen

In der anschließenden Planungsphase (2) wird die Fragestellung in *diagnostische Hypothesen* übersetzt, die im weiteren Verlauf des Prozesses mit diagnostischen Daten überprüft werden sollen. Jäger weist darauf hin, dass in der Regel nicht nur eine, sondern mehrere Hypothesen formuliert werden, die im weiteren Verlauf des Prozesses anhand der Daten untersucht werden. Vor der Erhebung oder Sammlung dieser Daten in der eigentlichen Durchführungsphase (3) werden zunächst jene Erhebungsmethoden und –instrumente ausgewählt, die sich zur Prüfung der Hypothesen und damit zur Beantwortung der Fragestellung eignen. Die Auswahl der Daten und Verfahren richtet sich einerseits nach wissenschaftlichen Kriterien (jeweilige Gütekriterien) sowie den Theorien und Populationen für die Normierung, die den Verfahren zugrunde liegen. Andererseits erfolgt die Auswahl nach ihrer Funktion für den diagnostischen Prozess, d. h. den intendierten Zielen und Strategien. Die endgültige Auswahl solle auf Basis von Brauchbarkeit, Ökonomie, Zuverlässigkeit, Zumutbarkeit und Akzeptanz der jeweiligen Verfahren getroffen werden. Nach der *Planung der Untersuchung* erfolgt dann die *Datenerhebung* mithilfe der gewählten Verfahren.

In der Evaluationsphase (4) werden schließlich die erhobenen Daten ausgewertet und vor dem Hintergrund der eingangs aufgestellten Hypothesen interpretiert. Hier erfolgt die *diagnostische Urteilsbildung*, die das Ziel des diagnostischen Prozesses ist. In dieser zentralen Phase des diagnostischen Prozesses werden „die vorliegenden Informationen über einen Beurteilungssachverhalt (Personen, Objekte, Ereignisse, Institutionen usw.) zu einem Urteil integriert“ (Jäger, 2006, S. 94). Das Urteil kann dabei die

Form einer Diagnose oder Prognose annehmen, und bildet, wenn sich kein beratender oder therapeutischer Prozess anschließt, den Endpunkt des beschriebenen diagnostischen Prozesses.

Der diagnostische Prozess, wie ihn Jäger (2006) schematisch darstellt, beschreibt einen typischen psychodiagnostischen Prozess, der normativen und technologischen Regeln folgt, und der auf diese formalisierte Weise eine bestimmte Fragestellung zu beantworten versucht. Die Fragestellung wird ihrerseits durch einen Auftrag definiert, der i. d. R. durch eine andere Person erteilt wird. Daher beinhaltet der dargestellte psychodiagnostische Prozess Jäger zufolge sowohl kommunikative als auch dynamische Aspekte zwischen Diagnostiker*in, Diagnostikand*in und Auftraggeber*in. Diese „diagnostische Trias“ ist allerdings eine Besonderheit psychologischer Diagnostik, die im Rahmen schulischer Diagnostik i. d. R. nicht gegeben ist. Eine weitere Besonderheit psychodiagnostischer Prozesse dieser Art ist, dass der Auftrag als beendet gilt, wenn die eingangs formulierte und präziserte Fragestellung durch das diagnostische Urteil beantwortet ist. Diagnoseprozesse und pädagogische Prozesse werden damit klar voneinander getrennt (z. B. Ingenkamp & Lissmann, 2008; Karst et al., 2017; Klug et al., 2013). Die konkrete Entscheidung, Planung und Durchführung von Maßnahmen, die eventuell aus diesem Urteil resultieren, obliegt damit in der Regel also nicht mehr der/dem Diagnostiker/in. In der pädagogisch-psychologischen Diagnostik im schulischen Kontext sind diagnostische Tätigkeiten jedoch stets mit unterrichtlichen oder pädagogischen Aspekten verbunden (z. B. Hesse & Latzko, 2017; Ingenkamp, 1985; Kaiser et al., 2017). Der *pädagogisch-diagnostische Prozess* kann also nicht bei der Urteilsfindung enden, sondern muss um die Ableitung von Entscheidungen erweitert werden.

1.2.4.3 Die Entscheidung am Ende des pädagogisch-diagnostischen Prozesses

Pädagogisch-diagnostische Prozesse

Im pädagogischen Kontext leiten diagnostische Handlungen pädagogische und unterrichtliche Handlungen nicht nur ein, sondern sie begleiten sie oftmals auch oder bereiten sie nach. Hesse und Latzko (2017) formulieren in ihrem Übersichtswerk der Diagnostik für Lehrkräfte einen *pädagogisch-psychologischen diagnostischen Prozess*, der die an Diagnosen anschließenden Entscheidungen und Handlungen explizit berücksichtigt, und damit die Phasen der Treatment vorbereitenden, begleitenden und nachbereitenden Diagnostik integriert (Krapp, 1979). Dieser Prozess ist dem hypothesengeleiteten Vorgehen, wie es Jäger (1986; 2006) beschreibt, sehr ähnlich und beschreibt im Wesentlichen einen vergleichbaren Ablauf. Die prozessimmanenten Schritte werden dabei teilweise davon abweichend untergliedert und ausgeführt. Zum einen werden Hypothesen in Anlehnung an Lukesch (1998) in zwei Arten eingeteilt, die während des diagnostischen Prozesses beantwortet werden sollen: (a) Feststellungshypothesen sind auf die Frage bezogen, ob das geschilderte Problem tatsächlich vorliegt, und (b) Erklärungshypothesen beziehen sich auf Gründe oder Ursachen für das Bestehen des Problems, nachdem es bestätigt werden konnte. Die Prüfung der Erklärungshypothesen stellt einen wichtigen Schritt bei der Vorbereitung von Interventionen vor, die das festgestellte Problem beheben oder abmildern sollen. Genau an dieser Schnittstelle zwischen Diagnose und Intervention fassen Hesse und Latzko (ebd.) den diagnostischen Prozess weiter als Jäger (2006) sowie Lukesch (1998), da hier auch die Vorbereitung

der Interventionen mitgedacht und die Planung und Durchführung der konkreten pädagogischen Entscheidung mit in den pädagogisch-psychologischen Prozess aufgenommen wird. Sie begründen diese Erweiterung mit der engen Verbundenheit zwischen Diagnose und Intervention, die dem schulischen Erziehungs- und Bildungsprozess immanent ist (vgl. Kap. 1.2.2).

Ihr Entwurf des pädagogisch-psychologischen Prozess sieht folglich einen fünfschrittigen Ablaufplan vor. Die hinzugefügte fünfte Phase des Prozesses expliziert die Planung und Durchführung der Intervention, die u. a. in einer Förderung, einem Training, einer Beratung oder der Adaption des Unterrichts bestehen kann. Der „Fünfer-Schritt“ professionellen Diagnostizierens nach Hesse und Latzko (2017, S. 91 ff.) kann folgendermaßen konkretisiert werden:

- (1) Formulierung der diagnostischen *Fragestellung* und theoretische Einordnung,
- (2) Formulierung theoriegeleiteter *Hypothesen*:
 - a. Feststellungshypothesen (Liegt das Problem vor?),
 - b. Erklärungshypothesen (Wodurch ist es entstanden?);
- (3) Auswahl der *Methoden* zur Hypothesenprüfung;
- (4) *Durchführung* der Untersuchung: Wahrung von Objektivität und Ökonomie;
- (5) Planung und Durchführung einer *Intervention* und/oder Beratung.

Ihre Phasenbeschreibung ist auf die Besonderheiten pädagogisch-psychologischer Diagnostik im Kontext von Schule bedacht und formuliert daher auch zentrale Hinweise an Lehrkräfte, welche Grundsätze einer „professionellen Diagnostik“ im Verlauf des Prozesses einzuhalten sind. Bei der Formulierung der Fragestellung im ersten Schritt (1) soll bspw. auf eine theoretisch-empirische Einordnung geachtet werden. Die Hypothesen im zweiten Schritt (2) sollen entsprechend der Problembeschreibung ergebnisoffen, das heißt, nicht zu einseitig oder zu eng formuliert werden. Gleichzeitig soll die Hypothesenprüfung nach dem „Effektivitätsprinzip“ erfolgen, d. h. es sollen nur diejenigen Hypothesen geprüft werden, für die in der jeweiligen Situation Wissen verfügbar ist und die plausibel erscheinen. Auch bei der Methodenwahl im dritten Schritt (3) soll neben der Passung der Methoden zum Problem auf Effektivität geachtet werden, was „das differenzierte Wissen des Diagnostikers über diagnostische Methoden und Standards“ (ebd., S. 93) erfordere. Die Durchführung der Untersuchung im vierten Schritt (4) ist den Prinzipien der Objektivität und Ökonomie unterworfen, und erfordert zudem eine ergebnisoffene Haltung. Die Herausforderung bei der Prüfung der Hypothesen anhand der erhobenen Daten sei es, professionell mit diesen umzugehen, und die Falsifikation der eingangs formulierten Vermutungen bspw. nicht als persönlichen Misserfolg anzusehen, sondern als Anlass, alternative Hypothesen zu prüfen. In der Praxis komme es durch eigene Vorurteile der Lehrkräfte in solchen Fällen nicht selten zu einer Umdeutung von Untersuchungsergebnissen oder einer Abwertung von Untersuchungsinstrumenten. In der fünften Phase schließlich (5), die neben der Abklärung der Feststellungshypothesen auch die Prüfung der Erklärungshypothesen voraussetzt, sollen konkrete Möglichkeiten erarbeitet werden, die bestehenden Probleme zu lösen. Durch die im Prozess erhobenen Daten zur Entstehung und Aufrechterhaltung des Problems sollen nun Maßnahmen abgeleitet werden, die inner- oder außerhalb des Unterrichts liegen und

von simplen Interventionen wie der Veränderung der Sitzordnung bis hin zum Einsatz spezifischer Trainingsprogramme oder der Erstellung und Umsetzung individueller Förderpläne reichen können.

Prozess der datengestützten Unterrichtsentwicklung

Die Entscheidung für oder Ableitung von pädagogischen Interventionen auf Basis von Daten spielt nicht nur in der pädagogischen Diagnostik, sondern ganz allgemein im Rahmen *datengestützter Unterrichtsentwicklung* von Lehrkräften eine Rolle (Thoren et al., 2020). Die Forderung, dass unterrichtliches Handeln auf empirische Evidenz gründen soll, ist in den Standards für die bildungswissenschaftlichen Anteile der Lehrkräftebildung verankert (KMK, 2004; Kunina-Habenicht et al., 2012) und wird durch die Annahme begründet, dass datenbasierte Entscheidungen die Lern- und Bildungsverläufe von Schüler*innen positiv beeinflussen können (Kippers et al., 2018; Schildkamp et al., 2013). Evidenzbasierte oder datengestützte Unterrichtsentwicklung kann, ähnlich wie eine professionelle Diagnostik (Hesse & Latzko, 2017) als eine Art wissenschaftlicher Prozess beschrieben werden, in dessen Verlauf Lehrkräfte empirische Daten zu erzeugen, analysieren und interpretieren, um auf dieser Grundlage unterrichtliche oder bildungsrelevante Entscheidungen zu treffen (Gummer & Mandinach, 2015; Mandinach & Gummer, 2016). Dieser Prozess lässt sich als Abfolge aus fünf typischen Phasen beschreiben:

- (1) Probleme und Fragestellungen identifizieren,
- (2) Nutzung sowie Erhebung der zur Klärung notwendigen Daten,
- (3) Informationen über das Problem aus den Daten generieren,
- (4) Entscheidung anhand der Informationen ableiten,
- (5) Evaluation der Auswirkung der getroffenen Entscheidung.

Ähnlich wie der diagnostische Prozess (Jäger, 2006) beginnt auch der Prozess datengestützter Unterrichtsentwicklung (Mandinach & Gummer, 2016) mit der Entwicklung der konkreten Fragestellung (1). Im Unterschied dazu wird die Hypothesenbildung, die sich im diagnostischen Prozess an die Fragestellung anschließt, nicht als Schritt expliziert. Dafür ist im Prozess datengestützter Unterrichtsentwicklung wiederum die Phase der Erhebung und Nutzung der Daten (2), die auch die Auswahl der Methoden und die Planung der Untersuchung recht präzise beschreibt, zentral und sehr ausführlich beschrieben. Anders als in den Konzepten von Jäger (2006) oder Hesse & Latzko (2017) ist diese aber nicht in verschiedene Phasen untergliedert. Die Daten, die nach der Sammlung bzw. Erhebung vorliegen, werden dann in der dritten Phase des Datennutzungsprozesses (3) vor dem Problemfeld oder der Fragestellung interpretiert. Dieser Schritt der Übersetzung der Daten in Informationen könnte im diagnostischen Prozess als Urteilsbildung bezeichnet werden, die den Endpunkt des regulären psychodiagnostischen Prozesses bildet (Jäger, 2006). Gemäß des Modells datengestützter Unterrichtsentwicklung sind diese Informationen, die in Hinblick auf die Fragestellung gewonnen wurden, allerdings die Grundlage für zwei weitere Schritte. Der Datennutzungsprozess wird auch hier eng mit der unterrichtlichen und pädagogischen Praxis verbunden, und die Entscheidung für eine Intervention stellt ähnlich wie bei Hesse und Latzko (2017) einen zentralen Schritt im Prozess dar. Der Prozess datengestützter Unterrichtsentwicklung endet allerdings nicht bei der Entscheidung für eine Intervention

(4), sondern formuliert als letzten Schritt die Evaluation der getroffenen Entscheidung (5). Diese Evaluation setzt eine Beobachtung der Schüler*innen sowie des eigenen Lehrverhaltens voraus und stellt in sich einen erneuten Datennutzungsprozess dar (Mandinach & Gummer, 2016). Im Rahmen pädagogischer Diagnostik wird die Erfassung von Interventionseffekten auch als Treatment begleitende (formative) bzw. nachbereitende (summative) Diagnostik bezeichnet (Krapp, 1979).

Vorschlag eines erweiterten Prozessmodells

Zusammengenommen lässt sich das Modell des diagnostischen Prozesses, wie es durch Jäger (2006) formuliert wird, sinnvoll durch den Fünfer-Schritt von Hesse und Latzko (2017) sowie das Prozessmodell datengestützter Unterrichtsentwicklung nach Mandinach und Gummer (2016) erweitern. Im Sinne einer umfassenden pädagogisch-psychologischen Diagnostik, die die Unterrichtsentwicklung im Fokus hat, genügt diese Zusammenführung dem Anspruch an pädagogisch-psychologische Diagnostik (Hesse & Latzko, 2017; Ingenkamp & Lissmann, 2008), dass Diagnose und Förderung in der Schule stets Hand in Hand gehen sollten. Das Urteil bildet damit den Startpunkt für einen direkt anschließenden pädagogischen Prozess, der im schulischen Kontext fest in den formalen Verlauf des diagnostischen Prozesses integriert wird. Nachfolgend soll ein Vorschlag für ein erweitertes Modell eines idealtypischen diagnostischen Prozesses dargestellt werden, der auch die Ableitung pädagogischer Anschlusshandlungen integriert (vgl. Abb. 1.3). Grundlage dieses Arbeitsmodells ist die normative psycho-diagnostische Prozedur nach Jäger (2006), die durch die Modellierungen von Hesse & Latzko (2017) sowie Mandinach & Gummer (2016) an die Spezifika der pädagogisch-psychologischen Diagnostik adaptiert wird.

Die erste Phase in diesem Modell (Orientierungsphase) beinhaltet die Identifikation von Problemen und Fragestellungen, woran sich die zweite Phase der Hypothesenbildung und Untersuchungsplanung anschließt. Diese zweite Phase kann entsprechend der Modellierung von Hesse und Latzko (2017) in zwei Teilphasen untergliedert werden: die Hypothesenbildung (Planungsphase a) und die Auswahl sowie Zusammenstellung der Untersuchungsmethoden (Planungsphase b). In der dritten Phase (Durchführungsphase) werden entsprechend des Untersuchungsdesigns diagnostische Daten erhoben oder gesammelt und anschließend ausgewertet. Daran schließt sich in der vierten Phase (Evaluationsphase) der Be- und Verwertung dieser Daten an. Hesse & Latzko sowie Mandinach & Gummer (2016) folgend wird diese Evaluationsphase in drei Teilschritte untergliedert. Zunächst werden die Daten vor dem Hintergrund der Fragestellung und der Hypothesen mit dem Ziel der Urteilsbildung interpretiert (Evaluationsphase a). Damit die diagnostischen Daten auch für die Unterrichtsentwicklung nutzbar sind, erfolgt daran anschließend die Ableitung einer pädagogischen Entscheidung (Evaluationsphase b). Diese Überführung von Informationen in Entscheidungen erfolgt bspw. durch die Planung oder Entwicklung konkreter Fördermaßnahmen. Im Sinne der Vollständigkeit datengestützter Unterrichtsentwicklung nach Mandinach & Gummer (2016) folgt schließlich die Phase der Evaluation der Auswirkung der getroffenen Entscheidung bzw. eingeleiteten pädagogischen Maßnahmen (Evaluationsphase c). Diese Phase stellt im Grunde einen erneuten Prozess der Datenerhebung und –

nutzung dar, der während oder nach Abschluss des pädagogischen Prozesses durchgeführt wird (Treatment begleitende vs. abschließende Diagnostik; Hesse & Latzko, 2017).

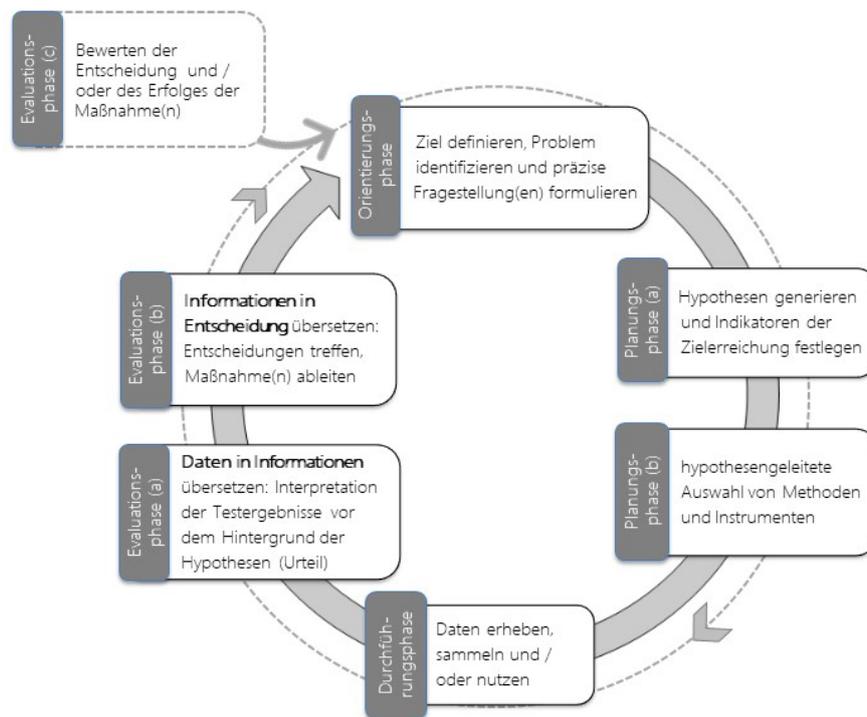


Abbildung 1.3 Der erweiterte diagnostische Prozess im pädagogisch-psychologischen Setting.

Im Rahmen der Diagnostik von Lernstörungen oder kognitiven Lernbesonderheiten führen Lehrkräfte vor allem die zwei Schritte der Evaluation der diagnostischen Daten aus. Die Erhebung und Auswertung der diagnostischen Daten erfolgt hier i. d. R. durch andere Personen, wie z. B. Sonderpädagog*innen oder Psycholog*innen.

Dieses erweiterte Prozessmodell veranschaulicht, auf welche Weise pädagogisch-psychologische Urteile idealerweise gebildet sowie mit unterrichtlichem Handeln verbunden werden sollten. Er stellt damit – ebenso wie die zugrundeliegenden Modelle – nicht unbedingt eine realistische Abbildung dessen dar, wie sich diagnostische Prozesse in der Praxis tatsächlich gestalten (vgl. Abschnitt 1.2.3.3 und 1.2.4.1). Vielmehr übernimmt es den *normativen Gedanken* des psycho-diagnostischen Prozesses (Herppich et al., 2017; Latzko & Hesse, 2009), den Jäger (2006) damit begründet, dass aufgrund der Komplexität und Verflechtungen in der Praxis eine Beschränkung auf wesentliche prozessimmanente Schritte erfolgen muss. Ein weiterer entscheidender Grund für die Formalisierung pädagogisch-psychologischer diagnostischer Prozesse ist, dass die Abklärung bestimmter Diagnosen wie bspw. Lernbesonderheiten ein formelles Vorgehen erfordert (vgl. Schrader, 2009, 2010), um Fehldiagnosen durch Urteilsfehler zu vermeiden (für eine Übersicht möglicher Urteilsfehler vgl. Helmke, 2009; Jürgens & Lissmann, 2015).

1.3 Zusammenfassung und Zwischenfazit

Diagnostik zu betreiben ist eine zentrale Aufgabe des beruflichen Alltags von Lehrkräften und diagnostische Kompetenz entsprechend eine *zentrale Kompetenz*, die Lehrkräfte für ihren beruflichen Alltag benötigen. Bei dieser Aufgabe sollen nicht nur die Leistungen oder lernrelevante Merkmale von Schülerinnen und Schülern beurteilt werden, sondern auch Lernvoraussetzungen erfasst, Lernprozesse beobachtet und ggf. Lernbesonderheiten festgestellt werden. Die diagnostischen Aufgaben einer Lehrkraft sind vielfältig, da eine *Vielzahl von diagnostischen Situationen* unterschieden werden kann, die im inner- und außerunterrichtlichen Diagnosekontext entstehen. Ein Unterscheidungsmerkmal diagnostischer Situationen im pädagogischen Kontext ist die Verbindlichkeit der Entscheidungen, die auf Grundlage der Diagnose getroffen werden. Verbindliche Diagnosen, wie bspw. Schullaufbahnentscheidungen oder die Feststellung von Lernstörungen mit dem Ziel der Zuweisung zu Fördermaßnahmen, haben in der Regel weitreichende Konsequenzen für die diagnostizierten Personen, und erfordern daher ein überlegtes und methodisch kontrolliertes Vorgehen (*formelle Diagnose*). Dieses regelgeleitete Vorgehen kann in Form eines normativen, an wissenschaftlichen Methoden orientierten Prozesses modelliert werden. Im Bereich der psychologischen Diagnostik integriert der *diagnostische Prozess* die Phasen der Orientierung und Problemlösung, Hypothesenbildung und Planung, Durchführung und Auswertung der Untersuchung und die abschließende Evaluation der gewonnenen Daten zur Urteilsbildung. Dieses Urteil informiert dann Entscheidungs- und Handlungsprozesse, die allerdings nicht in den Aufgabenbereich der Diagnostiker*innen fallen. Diagnostik im schulischen Kontext ist jedoch immer eng mit *pädagogischen Entscheidungen* und Handlungen verknüpft, sodass pädagogische Diagnostik nicht beim Urteil endet, sondern der diagnostische Prozess um das Treffen von Entscheidungen und die Ableitung oder Planung von Interventionen erweitert werden muss. Die einzelnen Schritte des pädagogisch-diagnostischen Prozesses müssen dabei nicht zwangsläufig in immer derselben Reihenfolge oder auf die gleiche Weise gestaltet sein. Vielmehr kann angenommen werden, dass die konkrete Gestaltung des Prozesses von Situation zu Situation variiert. Während im Rahmen des Unterrichtsgeschehens schnelle, automatisierte und oftmals heuristische Urteile gefällt werden (*informelle Diagnose*), kommt bei hochrelevanten Entscheidungen für die Diagnostikand*innen ein kontrollierter, kognitiv aufwändiger analytischer Diagnoseprozess zur Anwendung. Die Herausforderung für die Lehrkraft besteht darin, den passenden Prozess auszuwählen.

— Diagnostische Situationen sind mannigfaltig und diagnostische Abläufe daher unterschiedlich. Diagnosen, die mit weitreichenden Konsequenzen verbunden sind, erfordern aber stets eine kontrollierte Urteilsbildung. Das Urteil dient in der pädagogischen Diagnostik der Ableitung von Anschlusshandlungen. —

2 Indikatoren diagnostischer Kompetenz von Lehrkräften

Im ersten Kapitel wurde herausgearbeitet, dass je nach diagnostischer Situation verschiedene Abläufe des diagnostischen Prozesses möglich sind, die zu unterschiedlich genauen Diagnosen führen können. In diesem Kapitel soll veranschaulicht werden, dass genaue Diagnosen einen zentralen Aspekt diagnostischer Kompetenzen von Lehrkräften darstellen, und dass die konkrete Gestaltung des diagnostischen Prozesses dementsprechend auch als Aspekt diagnostischer Kompetenz betrachtet werden kann. Diagnostische Kompetenz wird schließlich als eine Kompetenz verstanden, die sowohl *Prozess- als auch Produktmerkmale* vereint: Diagnostische Kompetenz lässt sich sowohl an „guten“ Urteilen als auch an professionellen Prozessabläufen erkennen.

Ganz allgemein bezeichnet diagnostische Kompetenz „die Kompetenz von Lehrkräften [...], Merkmale ihrer Schülerinnen und Schüler angemessen zu beurteilen und Lern- und Aufgabenanforderungen adäquat einzuschätzen“ (Artelt & Gräsel, 2009, S. 157). Die Verankerung diagnostischer Kompetenz an diesen zwei zentralen Gegenständen (Personen und Aufgaben) bezeichnen Hesse & Latzko (2017) als notwendigen Konsens gemäß der Definition und Aufgaben pädagogischer Diagnostik (vgl. Kapitel 1.2.2). Die kompetente Beurteilung dieser beiden Diagnosegegenstände ist nämlich zentral, wenn es um die Passung von Personen- und Aufgabenmerkmalen im Rahmen adaptiven Unterrichtens geht, die ihrerseits einen dritten Beurteilungsgegenstand darstellen kann (Helmke, 2007, 2009). In dieser Arbeit soll insbesondere dem Aspekt der Beurteilung von Personenmerkmalen Aufmerksamkeit geschenkt werden.

2.1 Urteilsgenauigkeit als Kern diagnostischer Kompetenz?

Das Fällen von korrekten Urteilen über Personen wird auch als *Urteilsgenauigkeit* bezeichnet, und beschreibt den Kern einer prominenten Definition diagnostischer Kompetenz, die bereits vor 30 Jahren formuliert wurde (Schrader, 1989):

„Mit diagnostischer Kompetenz („diagnostic competence“ / „accuracy of judgment“) bezeichnet man die Fähigkeit eines Urteilers, Personen zutreffend zu beurteilen“
(Schrader, 2010, S. 102.)

Diagnostische Kompetenz ist im Sinne dieser Definition mit Urteilsgenauigkeit gleichzusetzen. In dieser Definition diagnostischer Kompetenz spielt die Güte oder *Qualität einer Diagnose* eine zentrale Rolle, für die dementsprechend klassische psychometrische Gütekriterien eingefordert werden (vgl. Schrader, 2010; Hesse & Latzko, 2017). Eine Diagnose sollte erstens objektiv sein, d. h. verschiedene Lehrkräfte sollten dieselbe Leistung auch auf dieselbe Weise bewerten. Sie sollte zweitens reliabel sein, sodass wiederholte Beurteilungen desselben Gegenstandes gleich ausfallen. Drittens sollte sie valide, d. h. nur durch das zu beurteilende Merkmal und nicht durch andere Faktoren beeinflusst sein. Es kann allerdings angenommen werden, dass Lehrer*innenurteile tatsächlich oftmals nicht den geforderten Gütekriterien entsprechen, denn ein und dieselbe Leistung werde häufig von verschiedenen Lehrkräften unterschiedlich bewertet, von derselben Lehrkraft zu zwei verschiedenen Zeitpunkten ungleich bewertet oder die Beurteilung komme

nicht nur durch das zu beurteilende Merkmal zustande, sondern auch durch andere Einflüsse, die mit dem Merkmal nichts zu tun haben (vgl. Ingenkamp, 1995 in Hesse & Latzko, 2017). Diese Vermutung bestätigte sich in diversen Studien, die die Urteilsgenauigkeit von Lehrkräften empirisch überprüften (Metaanalysen: Hoge & Coladarci, 1989; Südkamp et al., 2012).

Zur *Bestimmung der Güte von Urteilen* wird in diesen Studien in der Regel berechnet, wie hoch die Übereinstimmung zwischen der Leistungseinschätzung einer Lehrkraft und objektiven Ergebnissen aus einem standardisierten Leistungstest ist. Als Maße der Übereinstimmung werden drei Akkuratheitskomponenten herangezogen, die auf unterschiedlicher Grundlage die Einschätzungen der Lehrkräfte mit einem empirischen Pendant vergleichen (Schrader & Helmke, 1987):

- (1) Die *Niveauebene* gibt an, ob die Höhe des durchschnittlichen Leistungsniveaus einer Lerngruppe korrekt eingeschätzt wird und kann auf diese Weise Über- oder Unterschätzungen sichtbar machen;
- (2) Die *Streuungs- oder Differenzierungs-komponente* erfasst die korrekte Einschätzung der Streuung bzw. Verteilung der Leistungen innerhalb einer Lerngruppe und zeigt an, ob das Ausmaß der Homogenität richtig beurteilt wird;
- (3) Die *Rangordnungs- oder Vergleichs-komponente* ermittelt die Richtigkeit der Rangordnung der Leistungen innerhalb der Lerngruppe und gibt die diagnostische Sensitivität einer Lehrkraft an.

Die Rangordnungskomponente, auch als Korrelationskomponente bezeichnet, gilt als die zentrale Komponente diagnostischer Kompetenz (Spinath, 2005), da sie empirisch besonders gut erfasst werden kann und interindividuelle Unterschiede der Lerngruppe widerspiegelt. Im folgenden Abschnitt werden einige Befunde zusammengetragen, die anhand der korrelativen Übereinstimmung zwischen der von der Lehrkraft erstellten Rangfolge und der Verteilung in einem objektiven Leistungstest ermitteln, inwiefern die Urteilsgenauigkeit von Lehrkräften gegeben oder eingeschränkt ist.

2.1.1 Ausgewählte Befunde zur Urteilsgenauigkeit

Anna Südkamp und Kolleg*innen (2012) zeigen in einer umfangreichen Meta-Analyse, dass die *mittlere Übereinstimmung* zwischen Schüler*innenleistungen und Lehrer*innenurteilen mit $r = .63$ nicht perfekt, aber doch zufriedenstellend hoch sei. Beunruhigender sei hingegen die Streuung der Urteilsgenauigkeit der Lehrkräfte, die mit einer Spannweite von $r = -.05$ bis $r = .97$ von quasi keiner Übereinstimmung bis zu perfekter Genauigkeit reicht. Zwischen den einzelnen Lehrkräften bestehen demzufolge große Unterschiede (vgl. auch Metaanalyse von Hoge & Coladarci, 1989). Südkamp und Kolleg*innen führen diese Unterschiede zwischen den Urteilsleistungen der Lehrkräfte auf vier verschiedene Faktoren zurück, die einschränkend auf die Übereinstimmung zwischen Lehrer*innen-Einschätzungen und Schüler*innen-Leistungen einwirken: (1) Merkmale der Lehrkraft, (2) Merkmale des Urteils, (3) Merkmale der Schüler*innen und (4) Merkmale des Tests.

Wenn die *Merkmale des Urteils* als Einflussfaktor auf die Urteilsgenauigkeit analysiert werden, zeigen sich Unterschiede in Bezug auf unterschiedliche Fächer, wie etwa Deutsch vs. Mathematik (Lorenz & Artelt,

2009). Auch die Spezifität der Domäne, d. h. ob ein Urteil über einen umschriebenen Fähigkeitsbereich, wie bspw. arithmetische Fähigkeiten, oder die akademische Leistung allgemein beurteilt werden soll, wirkt auf die Höhe der Urteilsgenauigkeit ein (Karing, 2009). Ferner fällt die Urteilsgenauigkeit bei der Beurteilung von Schulleistungen grundsätzlich höher aus als die Beurteilung schulleistungsrelevanter Fähigkeiten (Karing, 2009; Spinath, 2005), die als Voraussetzung für das Erbringen von Schulleistungen angesehen werden (Hosenfeld et al., 2006). Birgit Spinath (2005) zeigte für alle drei von Schrader und Helmke (1987) entwickelten Maße der Urteilsgenauigkeit mehr oder weniger deutliche Verzerrungen, wenn die Lehrkräfte die allgemeinen kognitiven Fähigkeiten (Intelligenz), die Fähigkeitsselbstwahrnehmung (Selbstkonzept), die Lernmotivation oder die Leistungsängstlichkeit ihrer Schüler*innen beurteilen sollten. Die höchste Güte der Beurteilungen zeigte sich hierbei noch für die Intelligenz, bei der die Niveau- und Streuungskomponente auf hohe Treffsicherheit hinwies, die Rangordnungskomponente mit durchschnittlich nur $r = .40$ aber nicht zufriedenstellend ausfiel. Die Einschätzung intellektueller Lernvoraussetzungen ist insbesondere dann eine Herausforderung für Lehrkräfte, wenn sie die Einschätzung anhand von tatsächlichen kognitiven Potentialen vornehmen sollen, wie Spinath (ebd) verlangt hat. Wenn sich die Lehrkräfte jedoch an der allgemeinen schulischen Performanz orientieren können, zeigt sich eine höhere mittlere Urteilsgenauigkeit von $r = .67$ (Wild, 1991). Insbesondere das Erkennen von sogenannten „Underachievern“, d. h. Schüler*innen mit hohem kognitivem Leistungspotenzial mit widersprüchlich schwachen Schulleistungen, fällt den Lehrkräften aufgrund dieser Leistungsorientierung allerdings schwer (Rost & Hanses, 1998).

Während der Einfluss von Urteilsmerkmalen schon sehr gut untersucht ist, gibt es bislang wenige Erkenntnisse über den Zusammenhang zwischen *Merkmalen der Lehrkräfte* sowie deren Auswirkungen auf die Güte der Urteile (vgl. Südkamp et al., 2012). Ein Faktor, der die Urteilsgenauigkeit von Lehrkräften moderiert, scheinen die kognitiven Fähigkeiten von Lehrkräften zu sein (Kaiser et al., 2012). Die Berufserfahrung einer Lehrkraft wiederum scheint keinen maßgeblichen Einfluss auf die Treffsicherheit zu haben (Schrader & Helmke, 1987; Südkamp et al., 2012). Allgemein überschätzen Lehrkräfte jeder Erfahrungsstufe das Niveau der schulischen Leistungen ihrer Schüler*innen (vgl. Hesse & Latzko, 2017; Kaiser et al., 2015). Ferner ist bei Lehrkräften im Allgemeinen zu beobachten, dass sie sich an einem klasseninternen Bezugssystem orientieren und die klassenübergreifende, d. h. objektive, Beurteilung von Leistungen oft schwer fällt (Ingenkamp & Lissmann, 2008; Südkamp & Möller, 2009; Weinert, 2001). Selbst in simulierten Umgebungen, d. h. bei virtuellen Schüler*innen in einem simulierten Klassenraum, konnte gezeigt werden, dass bei der Beurteilung von Schüler*innen ein sog. Referenzgruppeneffekt wirkt, sodass bei objektiv gleichen Leistungen der/des Einzelnen ihr oder sein Leistungsniveau je nach Ausprägung der übrigen Schüler*innen in der Klasse beurteilt wird (Südkamp et al., 2008; Südkamp & Möller, 2009).

Was sich in den Studien im simulierten Klassenzimmer ebenfalls nachweisen ließ, ist die Tatsache, dass die Leistungen von Schüler*innen nicht nur auf Basis der Merkmale, die tatsächlich Gegenstand der Beurteilung sind, eingeschätzt werden, sondern dass bei der Urteilsbildung auch andere Merkmale

herangezogen werden, die mit der Beurteilung eigentlich nichts zu tun haben sollten – und dass diese urteilsirrelevanten Merkmale die Urteilsgenauigkeit drastisch reduzieren können (Schülerinventar; Kaiser et al., 2015). Bevor es im Rahmen von Simulationen die Möglichkeit zur systematischen Variation von *Schüler*innen-Eigenschaften* gab, warnten Lehrbücher für Lehrkräfte (z. B. Helmke, 2009; Ingenkamp & Lissmann, 2008; Jürgens & Lissmann, 2015) bereits vor dem Umstand, dass Leistungsurteile durch unsystematische wie auch systematische Urteilsfehler und –tendenzen bei der Schätzung von Leistungen verzerrt sein können. Zu den Quellen für systematische Urteilsfehler zählen u. a. implizite Persönlichkeitstheorien, die Lehrkräfte auf Basis verschiedener allgemeiner Merkmale ihrer Schüler*innen über diese bilden, und die dann bei der Beurteilung spezifischer Leistungen verzerrend einwirken (vgl. Hesse & Latzko, 2017). Zu den Schüler*innen-Merkmalen, die einen empirisch gut belegten Einfluss auf die Bildung impliziter Persönlichkeitstheorien haben, gehören u. a. die allgemeine Intelligenz sowie die Motivation von Schüler*innen (Hesse & Latzko, 2017; Kaiser et al., 2012). Selbst Oberflächenmerkmale wie die physische Attraktivität (Ritts et al., 1992), das Vorhandensein eines Migrationshintergrunds (Sprietsma, 2009) oder der Name (Kube, 2009; Rudolph et al., 2007) haben Einfluss auf die Wahrnehmung und Beurteilung einzelner Schüler*innen. Den genannten Studien zufolge werden attraktive Schüler*innen positiver beurteilt als weniger attraktive, Mädchen tendenziell positiver als Jungen (hier sind Domänenunterschiede zu beachten, vgl. z. B. Kaiser et al., 2015). Die Namen der Schüler*innen sind wiederum eine Quelle für Verzerrungen, da diese mit Vorurteilen über die Attraktivität, Intelligenz oder Verhaltensauffälligkeiten assoziiert sind. Möglicherweise steht dies in Zusammenhang dem assoziierten familiären Hintergrund – denn Schüler*innen aus bildungsfernen Schichten oder mit Migrationshintergrund werden schlechter bewertet, bei objektiv vergleichbaren Leistungen.

Diese Studienlage belegt eindrücklich, dass die Urteilsgenauigkeit von Lehrkräften durch verschiedene Faktoren beeinträchtigt wird. Die individuell schwankenden Fähigkeiten der Lehrkräfte (Hoge & Coladarci, 1989; Kaiser et al., 2015; Südkamp et al., 2012) sind ein Hinweis darauf, dass die Urteilsgenauigkeit der Lehrkräfte erhöht bzw. auf ein homogen hohes Niveau gebracht werden sollte. Um dies zu erreichen, ist die bloße Betrachtung von Einflussfaktoren auf das Urteil jedoch nicht ausreichend, sondern es muss darüber nachgedacht werden, wie die Urteilsgenauigkeit konkret erhöht werden kann. Hierfür ist es notwendig, die rein „produktorientierte“ Sichtweise (vgl. Schrader, 2010, S. 106), die in den Studien zur Urteilsgenauigkeit eingenommen wird, zu erweitern. Dem Paradigma der Urteilsgenauigkeit, das sich mit der Genauigkeit diagnostischer Urteile und deren Bedingungen beschäftigt, muss die *Betrachtung von Prozessmerkmalen* gegenübergestellt werden, bei der der Urteilsprozess selbst zum Gegenstand der Untersuchungen wird (Schrader, 2009, 2013). Eine erste Erweiterung der Konzeptualisierung diagnostischer Kompetenz stellt das Konzept der diagnostischen Expertise dar. Dort werden zwar noch keine konkreten, operationalisierbaren Prozessmerkmale beschrieben – der Einfluss des diagnostischen Prozesses auf die Urteilsgenauigkeit allerdings bereits explizit mitgedacht. Außerdem wird die Fähigkeit inbegriffen, aus der Diagnose passende pädagogische Anschlusshandlungen abzuleiten.

2.1.2 Diagnostische Expertise als Erweiterung der Urteilsgenauigkeit

Ein Ansatzpunkt zur Verbesserung der Urteilsgenauigkeit ist die Orientierung am Ideal der formellen Diagnose, d. h. die Einhaltung eines systematischen, methodisch-kontrollierten Prozesses, welche Urteilsverzerrungen entgegen wirken soll (vgl. Abschnitt 1.2.4.2). Die *Fokussierung auf den diagnostischen Prozess*, der zu akkuraten Diagnosen führt, erweitert die enge Definition diagnostischer Kompetenz, die mit der Umschreibung als Urteilsgenauigkeit einhergeht. Zur Abgrenzung der produktorientierten Sichtweise auf diagnostische Kompetenz führte Andreas Helmke (2007, 2009) den *Begriff der diagnostischen Expertise* ein. Dieser Begriff soll deutlich machen, dass das Verständnis diagnostischer Kompetenz – neben der Genauigkeit von Urteilen – auch die Auswahl und Durchführung professioneller diagnostischer Handlungen beinhalten muss. Die Treffsicherheit eines Urteils ist Helmke (2007) zufolge zwar ein wichtiges, aber bei Weitem nicht das einzige Kriterium diagnostischer Kompetenz. Neben der Diagnosegenauigkeit müssten bspw. Fähigkeiten zur Feststellung von Personenmerkmalen vorhanden sein, um den Unterricht adaptiv an die Voraussetzungen der Schüler*innen anzupassen. Hierzu sind auch die Fähigkeit zur Feststellung von Ist-Zuständen und die Identifikation von Potentialen („Zone der nächsten Entwicklung“; Hesse & Latzko, 2017, S. 30) notwendig oder die Wahl einer passenden Bezugsnorm. Insgesamt umfasse diagnostische Kompetenz im Sinne einer Expertise folgende Fähigkeiten (Helmke, 2009; hier dargestellt in Anlehnung an Jürgens & Lissmann, 2015: „Standards for Teacher Competence in Educational Assessment of Students“):

- (1) für Fragestellung angemessene Diagnoseverfahren auswählen können;
- (2) Diagnoseverfahren selbst entwickeln können, die für bestimmte Unterrichtsentscheidungen angemessen sind;
- (3) Standardisierte und informelle Tests anwenden, auswerten und interpretieren können;
- (4) Diagnostische Informationen für pädagogische Entscheidungen einzelne Schüler betreffend, die Unterrichtsplanung, die Entwicklung des Curriculums oder Schulentwicklungsprozesse nutzen können;
- (5) Valide Verfahren zur zusammenfassenden Beurteilung von Einzelbewertungen entwickeln können;
- (6) Diagnostische Ergebnisse ggü. Lehrkräften, Eltern und Schüler*innen verständlich kommunizieren können;
- (7) Ethisch oder rechtlich inakzeptable oder anderweitig ungeeignete Diagnoseinstrumente erkennen können.

Diese Aufzählung beinhaltet einerseits Fähigkeiten, die die professionelle Durchführung diagnostischer Prozeduren betreffen. Andererseits wird hervorgehoben, dass die diagnostische Expertise auch Fähigkeiten zur Nutzung diagnostischer Daten einschließt. Diagnostische Kompetenz wird umfassend verstanden und als Fähigkeit zur *Ableitung von pädagogischen Handlungen* auf der Grundlage der Beurteilung von Lernvoraussetzungen, –fortschritten und –schwierigkeiten (Weinert, 2001). Das Wissen darüber,

„welche Aspekte in die Diagnostik einzufließen haben, wie dies zu erfolgen hat und was sich daraus ableiten lässt“ (Hascher, 2008, S. 77) können als Kernmerkmale diagnostischer Kompetenz festgehalten werden, wenn diese als Expertise verstanden wird. Jene Expertise zeigt sich zusammengefasst darin, dass Lehrkräfte Leistungen und Fähigkeiten erfassen, diese Daten interpretieren und auf der Datenbasis sinnvolle Fördermaßnahmen ableiten sowie bewerten können (vgl. „Assessment Literacy“: Webb, 2002; „Data Literacy“: Mandinach & Gummer, 2016).

Zur Ausbildung diagnostischer Kompetenzen gehört diesem Verständnis zufolge insbesondere, über informelle Diagnosen hinauszugehen und *formelle Diagnosen erstellen* zu können (Schrader, 2010). Diese Fähigkeit beinhaltet beispielsweise, dass empirisch überprüfte Erhebungsverfahren eingesetzt werden, um möglichst akkurat zu diagnostizieren. Diese Ausrichtung der diagnostischen Prozeduren am Ideal der formellen Diagnose ist allerdings auch mit Schwierigkeiten verbunden. Zum einen reicht für eine*n Expert*in nicht nur „mentale Informiertheit und das instrumentelle Wissen zur Verwendung diagnostischer Verfahren und Materialien“ (Jürgens & Lissmann, 2015, S. 21) aus. Zum anderen muss bedacht werden, dass diese formelle Form der Diagnostik nicht für jeden Anlass passend ist (vgl. Kap. 1.2.3.3 und 1.2.4.1). Formelle Diagnostik bedeutet einen hohen Aufwand, der einerseits nicht immer passend, und andererseits auch nicht immer vorteilhaft ist: Diagnosen im Unterricht könnten sogar „durch eine pädagogisch günstige Voreingenommenheit getragen“ sein (Hesse & Latzko, 2017, S. 32), da gewisse Wahrnehmungsverzerrungen förderlich sein können für die Lern- und Leistungsentwicklung von Schüler*innen. Hesse und Latzko (ebd.) regen Lehrkräfte daher an, zwischen subjektiven und pädagogisch fruchtbaren Diagnosen, die handlungsleitend sind, sowie objektiven, erkenntnisleitenden Urteilen, die auf Anwendung standardisierter Verfahren beruhen, zu unterscheiden. Mit dieser Unterscheidung wollen sie sensibilisieren, dass die Genauigkeit von Urteilen nur ein, aber nicht das einzige Kriterium diagnostischer Kompetenz ist. Dieser Gedanke wird im folgenden Abschnitt genauer ausgeführt. Dort werden konkrete Prozessmerkmale diagnostischer Kompetenz vorgestellt, die einbeziehen, dass es je nach Anlass unterschiedliche Formen der Beurteilungen geben kann.

2.2 Prozessmerkmale als Indikator diagnostischer Kompetenz

Im vorausgehenden Abschnitt wurde herausgestellt, dass die Urteilgenauigkeit als ein Kernmerkmal diagnostischer Kompetenz anzusehen ist (vgl. Schrader, 2010, 2013). Da das Urteil oder die Diagnose ein Ausgangspunkt für pädagogische Entscheidungen ist, ist sie innerhalb eines professionellen pädagogischen Handelns zweifelsfrei von großer Bedeutung (Herppich et al., 2017; Latzko & Hesse, 2009). Allerdings betont diese Betrachtung diagnostischer Kompetenz – wie bereits erwähnt – ausschließlich das Produkt der Diagnose. Durch den Begriff der diagnostischen Expertise kann dieser Limitation begegnet werden, indem formuliert wird, auf welchem Wege ein*e Diagnostiker*in zu genauen Urteilen kommt. Vor dem Hintergrund, dass es verschiedenste diagnostische Anlässe und Zielstellungen gibt, die verschiedene Abläufe des Prozesses implizieren (vgl. Abschnitt 1.2.4.1), müssen die Analyse des Prozesses und seine konkreten Merkmale noch mehr in den Vordergrund gerückt und in Zusammenhang mit dem Produkt des Prozesses gebracht werden.

2.2.1 Die Diagnose als Produkt und als Prozess

2.2.1.1 Vier-Komponenten-Modell der Diagnosequalität

Lars Behrmann und Stefanie van Ophuysen (2017) stellen ein heuristisches Modell vor, das sowohl für die Diagnose als Produkt als auch für den zugrundeliegenden Prozess Merkmale formuliert, die die Qualität einer Diagnose beschreiben können. Sie machen darauf aufmerksam, dass selbst für die *Diagnose als Produkt* nicht nur die Genauigkeit oder Akkuratheit ausschlaggebend ist – allein schon deshalb, weil es gar nicht unter allen Umständen eine korrekte Diagnose gäbe. Vor allem im Rahmen komplexer Entscheidungsprozesse, bei denen eine Vielzahl von Informationen eine Rolle spielt, sei es oftmals schwierig, ein eindeutig richtiges diagnostisches Urteil abzugeben. Die Qualität eines Urteils sei in solchen Fällen bereits dann gewährleistet, indem hochwertige diagnostische Informationen und Daten generiert werden. Ein „gutes Urteil“ ist dementsprechend nicht nur ein akkurates Urteil, sondern vornehmlich eines, das durch einen diagnostischen Prozess mit hoher Qualität entstanden ist. In ihrem *Vier-Komponenten-Modell der Diagnosequalität* (vgl. Abb. 2.1) definieren sie vier Merkmale guter Diagnosen, die einerseits auf die Ergebnisse des diagnostischen Prozesses und andererseits auf den diagnostischen Prozess selbst bezogen sind.

Zur Beurteilung der *Qualität des Prozesses* unterscheiden sie Merkmale der Informationserfassung sowie der Informationsverarbeitung. Der Prozess der Informationserfassung spiegelt von außen sichtbares diagnostisches Handeln wider, das die Lehrkraft hier konkret im Rahmen des Unterrichts zeigt. Um die Informationserfassung professionell zu gestalten, bedarf es eines klaren Ziels, das die zu beurteilenden Merkmale definiert. Neben der Auswahl und Implementation der ausgewählten Untersuchungsmethoden im Unterricht ist es den Autor*innen zufolge ein Qualitätsmerkmal, dass die im Rahmen unterrichtlichen Handelns gewonnenen Beobachtungen systematisch über einen längeren Zeitraum dokumentiert werden, damit sie für spätere Entscheidungen verfügbar sind. Der an die Informationserfassung gekoppelte Prozess der Informationsverarbeitung ist dem Modell zufolge dann qualitativ hochwertig, wenn er einerseits flexibel und revidierbar ist, was eine gewisse ergebnisoffene Haltung während des Prozesses impliziert (vgl. auch diagnostischer Prozess nach Hesse & Latzko, 2017). Die Verarbeitung soll ferner unverzerrt, d. h. mit möglichst wenigen Urteilsfehlern belastet, und außerdem ressourcenschonend gestaltet sein. Aufgrund ihrer begrenzten Ressourcen müsse eine Lehrkraft diese Kosten-Nutzen-Entscheidungen dahingehend treffen, welche Informationen sie zur Urteilsfindung berücksichtigt und welche sie außen vor lässt. Die spezifischen Merkmale der beiden Prozesskomponenten wirken sich schließlich auf die Prozessergebnisse aus.

Zur Bemessung der *Güte der Ergebnisse des Prozesses* werden zum einen die Qualität der diagnostischen Daten und zum anderen ein hochwertiges Urteil herangezogen. Die Qualität der diagnostischen Daten oder Informationen werden hier nicht nur anhand der klassischen Gütekriterien (Objektivität, Reliabilität, Validität) festgemacht, sondern auch an der Relevanz der Daten und ob diese verschiedene Perspektiven widerspiegeln. Diese Multiperspektivität sichert ab, dass die Einschätzungen nicht nur auf eigenen

Einschätzungen beruhen, sondern z. B. die Einschätzungen anderer Lehrkräfte implizieren. Ein Faktor, der die Datenqualität praktisch immer limitiert und daher für die Bemessung der Güte berücksichtigt werden muss, ist der Umstand, dass den Lehrkräften in der Praxis nur eine eingeschränkte Menge an Daten verfügbar ist. Die Güte des Urteils, das die zweite Ergebniskomponente darstellt, ist dem Modell zufolge neben der Akkuratheit daran zu bemessen, ob die Kriterien der Transparenz und der Fairness erfüllt sind. Während es bei der Akkuratheit darum geht, dass die Diagnose das betreffende Merkmal einer Person zutreffend widerspiegelt, fokussieren die Merkmale der Transparenz und der Fairness darauf, dass diese Einschätzung auch durch andere Personen nachvollziehbar sein soll. Das Vorhandensein objektiver Daten und Kriterien erhöht die Transparenz und die Fairness des Urteils erheblich.

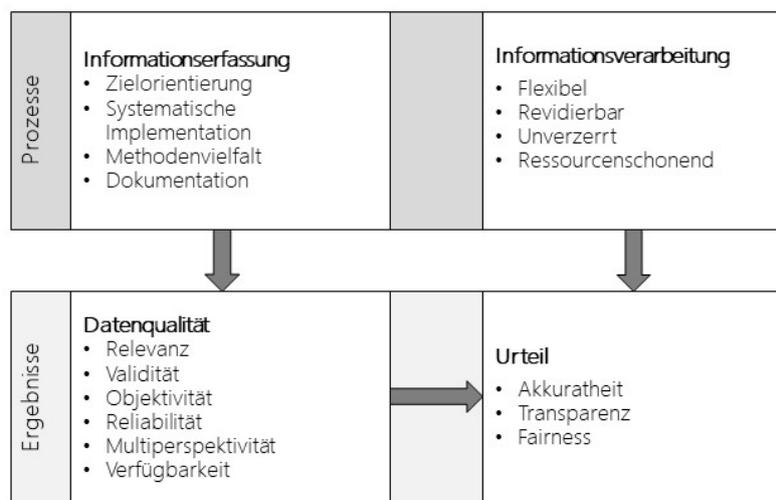


Abbildung 2.1 Vier-Komponenten-Modell der Diagnosequalität.
Darstellung in Anlehnung an Behrmann & Van Ophuysen (2017, S. 39).

Die zentrale im Modell von Behrmann & van Ophuysen (2017) ist, dass sich die Merkmale des diagnostischen Prozesses auf die Qualität der Produkte auswirken. Die empirische Überprüfung des Modells steht zwar bislang noch aus. Für die Gültigkeit der einzelnen Komponenten finden sich jedoch einige empirische Belege, die im folgenden Abschnitt vorgestellt werden. Diese exemplarischen Studien zeigen, dass Lehrpersonen mit unterschiedlich ausgeprägten diagnostischen Kompetenzen unterschiedliche Muster der Verarbeitung diagnostischer Informationen aufweisen – und damit, dass die Merkmale des Prozesses tatsächlich einen Einfluss auf die Güte der Produkte haben. Diese Befunde sprechen für die Plausibilität der im Modell angenommenen Komponenten.

2.2.1.2 Empirische Befunde zu Prozessen der Informationserfassung und –verarbeitung

Dass Diagnoseprozesse nicht einheitlich verlaufen und entsprechend die Diagnosen beeinflussen, zeigen bereits *Studien zur Urteilsgenauigkeit*. Spinath (2005) konnte in ihrer Untersuchung zur Urteilsgenauigkeit für lernrelevante Merkmale feststellen, dass die drei Komponenten, die als Maß der Urteilsgenauigkeit herangezogen werden können, relativ unabhängig voneinander zu sein scheinen. Sie fand nicht nur Unterschiede für die einzelnen Beurteilungsgegenstände, sondern auch eine beträchtliche intraindividuelle

Varianz für die einzelnen Genauigkeitskomponenten innerhalb desselben Beurteilungsgegenstandes. Während beispielsweise die Beurteilung der Intelligenz ein homogenes und relativ wenig verzerrtes Bild ergab, wenn die Höhe des mittleren Fähigkeitsniveaus (Niveauelemente) sowie die Streuung der Leistungen (Streuungskomponente) eingeschätzt werden sollten, waren in den drei anderen Merkmalsbereichen stärkere Verzerrungen und auch größere Inkonsistenzen bei diesen beiden Komponenten zu verzeichnen. Diese Studie lässt vermuten, dass je nach Zielstellung der Diagnose hier unterschiedliche Prozesse der Informationsverarbeitung ablaufen. Spinath arbeitet heraus, dass Verzerrungen bezüglich der Niveauelemente (mittlere Einschätzung des Niveaus der Leistungen) grundsätzliche, schüler*innenübergreifende Unter- oder Überschätzung durch Urteilstendenzen oder inadäquate Vergleichsmaßstäbe anzeigen können. Verzerrungen hinsichtlich der Streuungskomponente (Verteilung der Leistungen) können ebenfalls ein Zeichen für systematische Urteilstendenzen sein. Und Ungenauigkeiten bei der Beurteilung der Rangordnungskomponente (Bestimmung der Rangordnung der Schüler*innen) zeigen schüler*innenspezifische Verzerrungen an.

Studien, die Lehrkräfte im Verlauf ihrer professionellen Entwicklung mithilfe von Experten*innen-Noviz*innen-Vergleichen untersuchen, zeigen, dass das Ausmaß an Berufserfahrung einen Einfluss auf die Anfälligkeit für Urteilsverzerrungen hat. Erfahrene Lehrkräfte und Lehramtsstudierende unterscheiden sich hinsichtlich ihrer Anfälligkeit für Urteilsverzerrungen entsprechend der *Zielorientierung der Diagnostik*. Sofern das diagnostische Ziel ein systematisches und kriteriengeleitetes Vorgehen erfordert, weil das diagnostische Ziel bedeutsamere Konsequenzen hat, können sich erfahrene Lehrkräfte – im Gegensatz zu Lehramtsstudierenden – teilweise von Vorurteilen freimachen (Krolak-Schwerdt et al., 2013). Auch die Nutzungswege von Vorinformationen zur Urteilsbildung in einer Deutscharbeit unterscheiden sich für die erfahrenen Lehrkräfte je nach Zielorientierung, nicht jedoch bei den Lehramtsstudierenden (Dünnebier et al., 2009). Jene Vorinformationen dienten erfahrenen Lehrkräften in dieser Studie nur dann als Anker für ihr Urteil, wenn das Ziel die allgemeine Eindrucksbildung war. Lehramtsstudierende hingegen zeigten auch in der Bedingung, dass eine Schullaufbahempfehlung abgegeben werden sollte, eine Verzerrung durch die Vorinformationen. Eine ähnlich angelegte Studie (Krolak-Schwerdt et al., 2009) belegte den Einfluss der sozialen Zielorientierung bei der Bearbeitung von Fallvignetten. Erfahrene Lehrkräfte arbeiteten unter dem Ziel der Eindrucksbildung weniger genau als unter der Bedingung, in der sie eine Schullaufbahempfehlung abgeben sollen. Die Elaboriertheit ihres Vorgehens war wieder abhängig vom Ziel der Beurteilung, während die Lehramtsstudierenden als Noviz*innen keine Unterschiede gemäß der Zielorientierung bei der Verarbeitung zeigten.

Diese Studien zur sozialen Zielorientierung illustrieren eindrücklich, dass pädagogische Diagnosen auf heuristischen (kategoriebasierten) oder analytischen (merkmalsbasierten) Wegen erstellt werden können (vgl. Schrader, 2013), und damit eine unterschiedliche Anfälligkeit für verzerrende Einflüsse haben. Die Differenzierung analytischer und heuristischer Prozesse der diagnostischen Informationsverarbeitung, die hier bei der Interpretation der Ergebnisse vorgenommen wird, deckt sich mit der Unterscheidung zweier grundlegender Arten der diagnostischen Urteilsbildung (Hetmanek & Van Gog, 2017).

Die Wirkung verzerrender Einflüsse auf die Urteilsbildung wird seit jüngerer Zeit im Rahmen von *Vignettenstudien* untersucht, die sich zur Erforschung des diagnostischen Prozesses von Lehrkräften besonders eignen. Hier konnte gezeigt werden, dass Lehrkräfte bei der Beurteilung von Schulleistungen (vgl. Kaiser et al., 2015) sowie Lernprozessen (Klug et al., 2013) Informationen heranziehen, die für das aktuelle Urteil gar nicht relevant sind. Die Nutzung irrelevanter Informationen wurde als Zeichen heuristischer Verarbeitung gedeutet und verschlechterte in den genannten Studien die Güte der Beurteilung substantiell. Darüber hinaus konnte auch hier herausgestellt werden, dass sich die diagnostische Kompetenz von Lehrkräften in verschiedenen Phasen ihrer Ausbildung entsprechend des Ausmaßes an analytischer vs. heuristischer Verarbeitung unterscheidet (Klug et al., 2016). Diese Studien können erste Hinweise dafür liefern, dass die Betrachtung der Informationserfassung und –verarbeitung die Qualität diagnostischer Urteile erklären kann. Zur Bestätigung dieser Annahme sind allerdings weitere Studien erforderlich, die die gefundenen Effekte replizieren und bestenfalls auch konkrete Ausprägungen der Prozessmerkmale mit denen der Produkte in Verbindung bringen. Ein Modell, das Prozessmerkmale explizit als Indikator diagnostischer Kompetenz aufgreift und damit für die wissenschaftliche Untersuchung strukturieren will, wird im folgenden Abschnitt vorgestellt.

2.2.2 Modellierung diagnostischer Kompetenz anhand von Prozessmerkmalen

Durch den Entwurf des Vier-Komponenten-Modells (Behrmann & van Ophuysen, 2017) konnte der Diskurs zur Definition diagnostischer Kompetenzen bereichert werden, indem darin relevante Prozess- und Produktmerkmale zusammenstellt werden, die weit differenzierter sind als beim Konstrukt der Urteilsgenauigkeit. Was bei dieser Betrachtung offen bleibt, ist allerdings die Frage, auf welche Weise die jeweiligen Prozessmerkmale auf die Produkte einwirken und inwiefern Einflussfaktoren für die Prozessmerkmale angenommen werden können. Im Vier-Komponenten-Modell werden also die Dimensionen und Merkmale von Diagnosen differenziert, aber noch keine Merkmale diagnostischer Kompetenz definiert. Zur Definition jener Merkmale bedarf es der Modellierung der Wirkzusammenhänge – ein solches Modell ist allerdings ein Desiderat, zumindest in der deutschen Forschungslandschaft (vgl. Südkamp & Praetorius, 2017). Die 2015 ins Leben gerufene Forschungsgruppe NeDiKo aus 15 deutschen Wissenschaftler*innen (Herppich et al., 2017) hat einen Versuch unternommen, die Merkmale und Wirkzusammenhänge zur Entstehung diagnostischer Kompetenz zu modellieren, um damit eine Grundlage für die Erfassung sowie die Erarbeitung von Ansätzen zur Förderung dieser Kompetenz(en) zu schaffen. Als Herausforderung bei der Modellierung beschreibt die Arbeitsgruppe dabei die Vielzahl diagnostischer Situationen und Tätigkeiten (vgl. Abschnitt 1.2.3.1), die unterschiedliche diagnostische Prozesse auslösen können (vgl. Abschnitt 1.2.4.1). Ungeklärt sei in diesem Zusammenhang die Frage, ob diagnostische Kompetenz als ein- oder mehrdimensionales Konstrukt zu verstehen sei: In der empirischen Untersuchung diagnostischer Kompetenzen von Lehrkräften zeigte sich, dass die diagnostischen Einschätzungen in unterschiedlichen Situationen kaum miteinander korrelieren (z. B. Spinath, 2005, vgl. Abschnitt 2.2.1.2), und dass sich diagnostische Kompetenz in verschiedenen Facetten ausdrückt oder gar von unterschiedlichen diagnostischen Kompetenzen ausgegangen werden

muss (Karst et al., 2017). Das NeDiKo-Kompetenzmodell wurde entwickelt, um diese Fragen im Rahmen zukünftiger Forschung zu betrachten. Das Kompetenzmodell erlaubt dabei unterschiedliche Zugänge der Betrachtung diagnostischer Kompetenz (Herppich et al., 2017, 2018), indem es die Differenzierung analytischer und heuristischer Prozesse der diagnostischen Informationsverarbeitung aufgreift.

Diagnostische Kompetenz kann im Spielraum dieser Pole als korrekte „Auswahl des jeweils angemessenen Diagnose-Modus sowie die Qualität der diagnostischen Prozedur“ (Hetmanek & Van Gog, 2017, S. 215) bezeichnet werden. Ob der gewählte Diagnosemodus angemessen ist, ergibt sich aus der Übereinstimmung der situationalen Anforderungen (vgl. Karst et al., 2017) sowie dem gewählten Formalisierungsgrad der Diagnose (Hesse & Latzko, 2017; Schrader, 2009; 2010). Damit können „die Anforderungen an kompetentes Diagnostizieren [...] je nach Situation stark voneinander abweichen“ (Hetmanek & Van Gog, 2017, S. 209).

Im Zentrum des NeDiKo-Kompetenzmodells (Herppich et al., 2017, 2018) steht entsprechend der *diagnostische Prozess*, der je nach diagnostischer Situation unterschiedlich lang und unterschiedlich explizit ausfallen kann (vgl. Kap. 1.2.3.2). Die Abbildung des Modells (s. Abb. 2.2) verdeutlicht, dass die je nach diagnostischer Situationen ausgelösten Prozesse ein mehr oder weniger vollständiges Durchlaufen des normativen diagnostischen Prozesses beinhalten (Modell des diagnostischen Prozesses mit flexiblen Abläufen nach Herppich et al., 2017; vgl. Abb. 1.1 in Kap. 1.2.3.2). Im Modell werden formelle und informelle Diagnoseprozesse unterschieden, die sich durch das Ausmaß an Kontrolliertheit, Systematik und Bewusstheit unterscheiden. Entsprechend des Formalisierungsgrades unterscheiden sich die skizzierten Prozesse auch hinsichtlich ihres Umfangs sowie des erforderlichen kognitiven Aufwands (Hetmanek & Van Gog, 2017; vgl. auch Brühweiler et al., 2017). Die Auswahl der diagnostischen Prozedur erfolgt dementsprechend nicht nur auf Grund der Erfordernisse der jeweiligen diagnostischen Situation und Zielstellung, sondern auch auf Basis der vorhandenen kognitiven Ressourcen, über die eine Lehrkraft verfügt. Die *kognitiven Leistungsdispositionen*, die im Modell als Voraussetzung diagnostisch kompetenten Handelns berücksichtigt werden, spiegeln dabei ein speziell auf Diagnostik bezogenes Wissen wider und bestimmen die Handlungsmöglichkeiten einer Person. Dieses spezifische Wissen ist den Autor*innen zufolge im Rahmen anderer Leistungsdispositionen der Person zu betrachten, wie z. B. allgemeine kognitive Fähigkeiten, Motivation, diagnostikrelevante Selbstkonzepte oder Überzeugungen. Das Ingesamt der spezifischen und unspezifischen kognitiven Fähigkeiten wird als grundlegend für die mögliche Qualität des Prozesses und seiner Ergebnisse erachtet, die wiederum als „quantifizierbare Performanz“ einer diagnostizierenden Lehrperson messbar sind. Das Ausmaß diagnostischer Kompetenz zeigt sich den Annahmen des Modells entsprechend also erstens anhand der kognitiven Voraussetzungen, zweitens an den Handlungen im Prozess und drittens anhand der jeweiligen Prozessergebnisse, d. h. der Diagnosen.

Zusammenfassend wird *diagnostische Kompetenz* im Rahmen des Modells verstanden als eine *kognitive Leistungsdisposition*, „die es den Lehrkräften ermöglicht, diagnostische Anforderungen in verschiedenen pädagogischen Handlungssituationen zu erfüllen“ (vgl. Herppich et al. 2018). Das Arbeitsmodell folgt

damit der Annahme, dass sich Kompetenz nicht nur in messbarem Verhalten äußert, sondern dass für dieses Verhalten bestimmte Fähigkeiten benötigt werden, die je nach Situation sichtbar werden, und die ihrerseits als Teil der diagnostischen Kompetenzen verstanden werden müssen (vgl. allgemeiner Kompetenzbegriff nach Blömeke et al., 2015; Weinert, 2001; Kap. 3). Sowohl diese ganzheitliche Sichtweise von Kompetenz als auch die kognitiven Voraussetzungen, die konkret für diagnostisch kompetentes Handeln benötigt werden, werden im folgenden Kapitel 3 noch ausführlicher beschrieben.

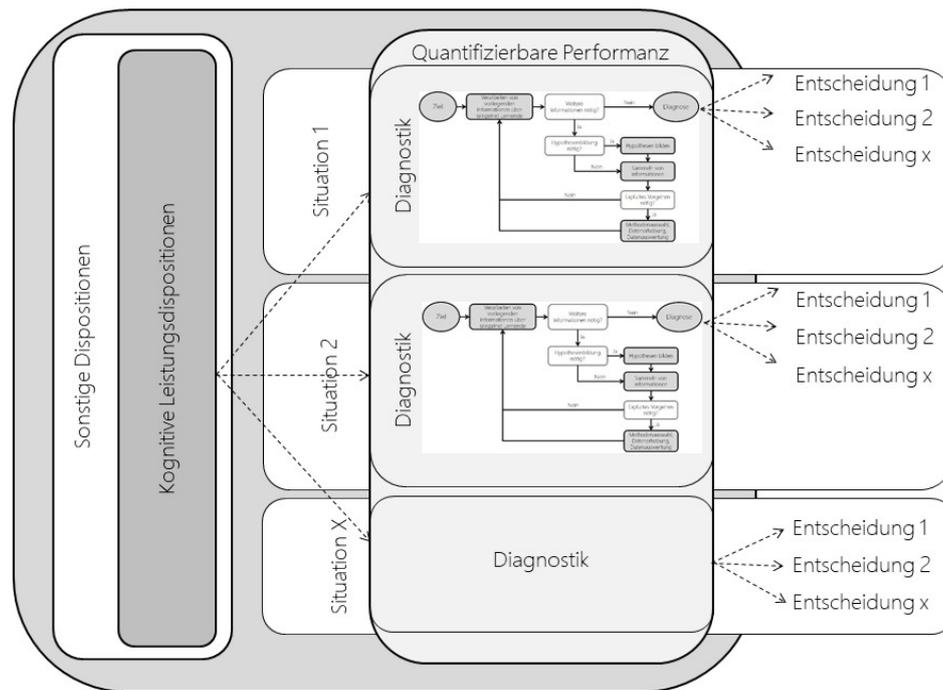


Abbildung 2.2 Schematische Darstellung des NeDiKo-Arbeitsmodells diagnostischer Kompetenz.
(Darstellung in Anlehnung an Herppich et al., 2017, 2018)

2.2.3 Integration von Anschlusshandlungen als Desiderat der Modellierung

Das NeDiKo-Arbeitsmodell diagnostischer Kompetenz (Herppich et al., 2017, 2018) stellt den diagnostischen Prozess in das Zentrum bei der Betrachtung diagnostischer Kompetenz und greift damit einen zentralen Teilaspekt diagnostischer Expertise auf, wie bspw. Andreas Helmke (2009) oder Tina Hascher (2008) sie definieren: Das kompetente Handeln von Lehrpersonen wird einerseits durch die Urteile und andererseits durch Urteilsprozesse sichtbar. Der zweite zentrale Teilaspekt, der im Konzept diagnostischer Expertise hervorgehoben wird, wird in diesem Modell allerdings jedoch nicht als zentraler Teil diagnostischer Kompetenz definiert: die pädagogische Entscheidung. Zwar wird verdeutlicht, dass die Diagnosen bestimmte pädagogische Entscheidungen einleiten, diese Entscheidungen werden aber außerhalb des diagnostischen Prozesses angesiedelt und damit eindeutig nicht mehr als Teil diagnostischer Kompetenz verstanden. Die Anschlusshandlungen, die als Besonderheit pädagogischer Diagnostik (vgl. Kap. 1.2.2; erweitertes diagnostisches Prozessmodell: Abb. 1.3) und als so wichtiger Aspekt diagnostischer

Expertise herausgearbeitet wurden (vgl. auch Schrader, 2013), spielen hier bei der Bemessung diagnostischer Kompetenz keine Rolle.

Die Trennung zwischen diagnostischen und pädagogischen Situationen wird üblicherweise in der klassischen psychologischen Diagnostik vorgenommen, die in der Regel eine Treatment vorbereitende Diagnostik ist (Jäger, 2006; Krapp, 1979; Latzko & Hesse, 2009). Ziel dieser Diagnostik ist es, die Entscheidung einer Auftraggeberin oder eines Auftraggebers zu informieren. Die Ausführung der Handlung, die auf Basis der Entscheidung initiiert wird, wird dort in der Regel nicht in den diagnostischen Prozess eingeordnet. Auch wenn diese strikte Trennung bei der Konzeptualisierung diagnostischer Kompetenz von Lehrkräften teilweise übernommen wird (vgl. Ingenkamp & Lissmann, 2008; Karst et al., 2017), fällt die Trennung beider Prozesse in der Praxis schwer: Die enge *Verwobenheit diagnostischer und pädagogischer Situationen* stellt die empirische Erfassung diagnostischer Kompetenz sogar vor große Herausforderungen (Praetorius et al., 2017). Pädagogische Diagnostik, wie sie in der Schule stattfindet, ist grundsätzlich eng mit den pädagogischen sowie unterrichtlichen Entscheidungen und Handlungen verbunden (Kaiser et al., 2017) und wird außerdem in der Regel durch die Lehrkraft selbst initiiert – statt extern beauftragt zu sein.

Die Entscheidung oder Handlung sollte daher in diagnostischen Prozessen, die Lehrkräfte durchlaufen, mitgedacht werden oder sogar Teil des Prozesses sein. Ein Kompetenzmodell, das sowohl den diagnostischen Prozess als auch die Anschlusshandlungen mitdenkt, wurde unlängst für die Beurteilung von Lernprozessen vorgelegt (Klug et al., 2013, 2016). Der Ansatz von Julia Klug und Kolleg*innen modelliert diagnostische Kompetenz entlang der Prozessphasen nach Jäger (2006) und enthält eine sogenannte *postaktionale Phase*. In dieser werten die Lehrkräfte zuvor erhobenen Daten aus, melden die Ergebnisse zurück und planen auf deren Basis Anschlusshandlungen. Auch wenn dort nicht spezifiziert wird, wie genau die Ableitung der Anschlusshandlungen von Statten gehen kann, wird die Anschlusshandlung erfreulicherweise als Teil diagnostischer Kompetenz mitgedacht. Was im Modell jedoch offen bleibt, sind wiederum die Fragen nach den Wirkzusammenhängen und den Voraussetzungen, die die Bewältigung der einzelnen Schritte im Prozess ermöglichen. Die Formulierung eines umfassenden Kompetenzmodells, das die Wirkzusammenhänge, die Voraussetzungen, die Angemessenheit des Vorgehens im diagnostischen Prozess und die diagnostischen Produkte – die neben Diagnosen auch die Qualität der Daten (vgl. Behrmann & van Ophuysen, 2017) sowie der Entscheidungen für pädagogisch-diagnostische Anschlusshandlungen beinhalten sollten – ist Aufgabe zukünftiger theoretischer wie empirischer Arbeiten.

2.3 Zusammenfassung und Zwischenfazit

In diesem Kapitel wurde herausgearbeitet, anhand welcher Merkmale diagnostische Kompetenz von Lehrkräften festgemacht oder erkannt werden kann. Die Beschreibung jener Indikatoren von diagnostischer Kompetenz ist eng verbunden mit der Frage der Definition oder Modellierung diagnostischer Kompetenz. Die Definition diagnostischer Kompetenz als *Urteilsgenauigkeit* hebt hervor,

dass diagnostische kompetente Personen in der Lage sein sollen, akkurate Diagnosen zu vergeben. Empirische Untersuchungen konnten zeigen, dass die Urteilsgenauigkeit von Lehrkräften durch verschiedene Faktoren verzerrt sein kann. Diese Befunde sind hilfreich, um Bedingungen für genaue Diagnosen zu identifizieren, machen aber keine Aussagen darüber, wie der Prozess der Diagnosestellung gestaltet sein muss, um genaue Diagnosen zu erstellen. In Abgrenzung zur Urteilsgenauigkeit wird daher auch der Begriff der *diagnostischen Expertise* verwendet, der betont, dass nicht nur korrekte Urteile Zeichen diagnostischer Kompetenz sind. Neben korrekten Diagnosen wird hier auch der *Prozess der Diagnosestellung als Merkmal* diagnostischer Kompetenz betrachtet, der auch die *Ableitung von Anschlussbehandlungen* beinhaltet. Diagnostische Kompetenz als Expertise betrachtet wird klar handlungsbezogen verstanden, setzt jedoch ausschließlich am Ideal der formellen Diagnose an. In Anbetracht des Spektrums unterschiedlicher Diagnoseanlässe und –aufgaben, mit denen eine Lehrkraft in ihrem Unterricht konfrontiert ist, muss davon ausgegangen werden, dass es für die verschiedenen Situationen auch verschiedene Abläufe des Diagnoseprozesses geben muss. Diesem Gedanken kann Rechnung getragen werden, indem diagnostische Kompetenz anhand der *Auswahl der jeweils passenden Diagnoseprozedur* definiert wird. Neben Diagnosen als beobachtbare Performanz müssen auch die kognitiven Voraussetzungen (Dispositionen) als Teil diagnostischer Kompetenz betrachtet werden, da diese bestimmen, inwiefern eine Person kompetente Handlungen überhaupt ermöglichen kann. Da sich Lehrkräfte in verschiedenen Phasen ihrer professionellen Entwicklung hinsichtlich ihrer Informationsverarbeitungsprozesse unterscheiden, muss das Verständnis diagnostischer Kompetenz zudem die *Qualität der diagnostischen Daten* berücksichtigen. Eine derart umfassende Modellierung diagnostischer Kompetenz als Prozess- und Produktkompetenz kann einen wertvollen Rahmen für zukünftige Studien liefern, die Möglichkeiten der Erfassung und Förderung diagnostischer Kompetenz(en) erforschen. Dabei wäre es wünschenswert, dass die Entscheidungen, die auf Basis der Urteile erstellt werden, mit in die Modellierung aufgenommen und nicht getrennt hiervon betrachtet würden.

— Diagnostisch kompetente Personen sind in der Lage, Diagnosen hoher Güte zu erstellen, indem sie einen analytischen diagnostischen Prozess zur Anwendung bringen. Sie sind auch in der Lage, ihre diagnostische Prozedur anzupassen, wenn die Situation ein weniger formelles Vorgehen erlaubt oder erfordert. Merkmale diagnostischer Kompetenz sind daher sowohl diagnostische Prozeduren als auch deren Ergebnisse. —

3 Vorüberlegungen zur Förderung diagnostischer Kompetenz

Das NeDiKo-Arbeitsmodell diagnostischer Kompetenz (Herppich et al., 2017, 2018; vgl. Abschnitt 2.2.2) verdeutlicht, dass kompetentes Handeln an das Vorhandensein bestimmter kognitiver Voraussetzungen gebunden ist. Diese Modellierung diagnostischer Kompetenz orientiert sich an einem Kompetenzbegriff, der Kompetenz als *kontextspezifische kognitive Dispositionen* versteht, die erlernbar sind und erworben werden müssen, um in umschriebenen Situationen kompetent handeln zu können (Klieme & Leutner, 2006). Bei der Modellierung von Kompetenz können demnach das beobachtbare Verhalten selbst (Performanz) als auch die dafür notwendigen Voraussetzungen (Dispositionen) betrachtet werden. In diesem Kapitel soll zunächst diese umfassende Sichtweise von Kompetenz näher erläutert werden (Abschnitt 3.1).

Entsprechend dieses Kompetenzverständnisses werden anschließend zwei zentrale Voraussetzungen diagnostischer Kompetenz vorgestellt: Neben kognitiven Aspekten (Abschnitt 3.2) werden auch affektiv-motivationale Faktoren vorgestellt, die dazu beitragen können, dass die kognitiven Voraussetzungen einer Lehrperson tatsächlich handlungswirksam werden (Abschnitt 3.3).

3.1 Überlegungen zum Begriff der Kompetenz

3.1.1 Kompetenz als Voraussetzung für Performanz

Kompetenz kann umschrieben werden als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2001, S. 27f). In dieser umfassenden Definition von Kompetenz fasst Franz E. Weinert verschiedene Aspekte zusammen, die gegeben sein müssen, damit eine Person ein professionelles Verhalten zeigen kann: Auf der einen Hand zählen hierzu kognitive Voraussetzungen, die nicht als feste Persönlichkeitsmerkmale aufgefasst, sondern als veränderbare und damit erlernbare Kompetenzen verstanden werden. Auf der anderen Hand zählt Weinert (ebd.) verschiedene Rahmenfaktoren oder „Bereitschaften“ auf, die die Umsetzung der kognitiven Voraussetzungen in tatsächliches Verhalten begünstigen. All jene Aspekte werden im Rahmen dieser einflussreichen Konzeptualisierung von Kompetenz als Dispositionen verstanden, die eine Vorbedingung für kompetentes Verhalten sind. Anhand dieser Darstellung wird deutlich, dass Kompetenz als ein komplexes Konstrukt zu verstehen ist, das sich aus verschiedenen Facetten zusammensetzt (Kunter, 2014). In dieser Definition von Kompetenz werden allerdings nur die Dispositionen für ein Verhalten als Kompetenz verstanden, nicht aber das Verhalten selbst. Unklar bleibt hierbei, wie sich das Zusammenspiel der Dispositionen genau gestaltet oder wie es gestaltet werden muss, damit schließlich professionelles Verhalten resultiert. Um Hinweise für die Förderung von Kompetenzen ableiten zu können, müssen sowohl das beobachtbare Verhalten als auch die Wirkzusammenhänge zwischen den einzelnen Faktoren spezifiziert werden.

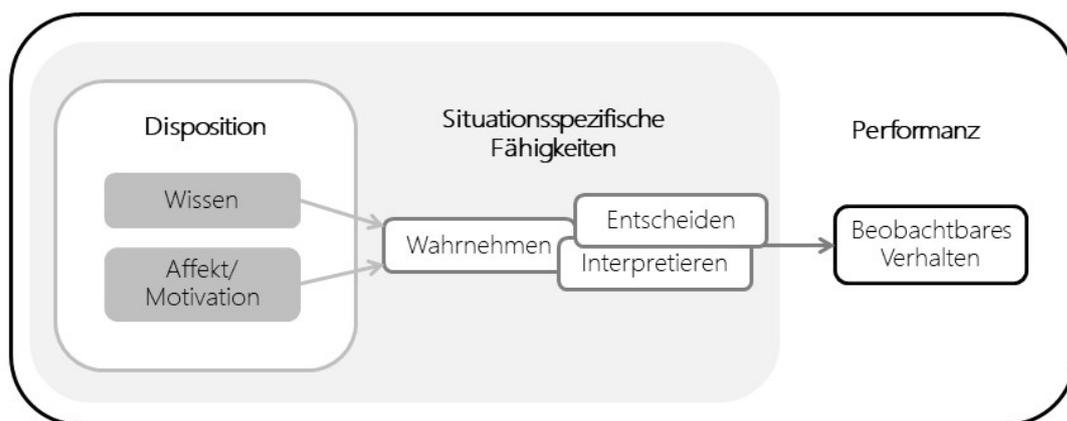
Eine erste Ausformulierung der Zusammenhänge zwischen Voraussetzungen professionellen Handelns und dem Handeln selbst geben Kunter und Kollg*innen (Baumert & Kunter, 2006; Kunter et al., 2011) in ihrem Modell professioneller Kompetenz. Kompetenz wird dort als Zusammensetzung aus Professionswissen, Überzeugungen, motivationalen Merkmalen und Fähigkeiten zur Selbstregulation beschrieben, die durch Lerngelegenheiten vor, während und nach der Ausbildung erworben werden, und damit eine Voraussetzung für professionelles Verhalten sind. In diesem Modell wird angenommen, dass Lerngelegenheiten dazu führen, dass Kompetenzen aufgebaut werden, und dass diese entsprechend ihrer Ausprägung mehr oder weniger direkt in die entsprechenden Handlungen überführt werden können. Im Rahmen von Lerngelegenheiten müssten daher auch Möglichkeiten geschaffen werden, um praktische Erfahrungen zu sammeln, damit sowohl Wissen als auch Können fokussiert werden (Baumert & Kunter, 2006). Neben *Wissen und Können* misst Mareike Kunter (2011, 2014) insbesondere der *Motivation* einen großen Stellenwert für die Entwicklung professioneller Kompetenz bei, da diese darüber entscheide, „wie viele kognitive Ressourcen Lehrkräfte im Unterrichtsverlauf investieren, ob sie ihr Wissen in einer bestimmten Unterrichtssituation anwenden und ob sie überhaupt bereit sind, Kenntnisse über unterschiedliche Strategien zu erwerben“ (Kunter, 2014, S. 535). Damit habe insbesondere die Motivation neben dem Wissen einer Lehrkraft einen entscheidenden Einfluss auf die unterrichteten Schüler*innen (Hattie, 2003). Dass motivationale Aspekte nicht nur im Rahmen unterrichtlichen Handelns eine Rolle bei der Überführung von Kompetenz in Performanz spielen, sondern auch für das diagnostische Handeln, zeigten unlängst erste Studien für diagnostische Aufgaben außerhalb des Unterrichts (Klug et al., 2016; Ohle et al., 2015; Westphal et al., 2018). In diesen Ansätzen werden motivationale Faktoren als Prädiktoren für diagnostische Verhaltensweisen herangezogen, die als mehr oder weniger kompetent bezeichnet werden. Im Zentrum steht dabei die Frage, inwiefern die untersuchten Prädiktoren zur Etablierung professioneller Handlungskompetenzen beitragen können – eine wichtige Frage im Zusammenhang mit Lerngelegenheiten für Lehrkräfte (Lipowsky, 2014). Entsprechend wird hier das Verhalten der Lehrkraft als Gradmesser von Kompetenz verstanden, welches wiederum an verschiedene Voraussetzungen geknüpft ist.

3.1.2 Kompetenz als Kontinuum

Ein Kompetenzmodell, das sowohl das Verhalten als auch die jeweiligen Voraussetzungen für dieses Verhalten als Aspekte von Kompetenz versteht, ist das „Modell der Kompetenz als Kontinuum“ von Sigrid Blömeke und Kolleg*innen (Blömeke et al., 2015). Kompetenz setzt sich in ihrem Modell aus Dispositionen, situationsspezifischen Fähigkeiten und aus beobachtbarem Verhalten (Performanz) zusammen – alle drei Aspekte werden von dem Begriff der „Kompetenz“ geeint (vgl. Abb. 3.1). Dabei ist den Autor*innen einerseits wichtig, Kompetenz ganzheitlich zu denken und damit Dichotomien zwischen den Begriffen „Kompetenz“ vs. „Performanz“ zu überwinden. Andererseits wird hier auch ein kontinuierlicher Weg von den Voraussetzungen zu den Handlungen skizziert. Die situationsübergreifenden Voraussetzungen von Kompetenz können in kompetentes Verhalten überführt werden, indem situationsspezifische Fähigkeiten eine Vermittlerrolle spielen. Diese situationsspezifischen

Fähigkeiten ermöglichen es der Person – basierend auf zuvor erworbenen Voraussetzungen – eine Situation professionell wahrzunehmen, zu interpretieren und dementsprechend in dieser Situation professionell zu entscheiden. Die Entscheidung wird dann in Form von mehr oder weniger kompetentem Verhalten beobachtbar. Wenn Kompetenz bemessen oder gefördert werden soll, können entsprechend alle drei genannten Aspekte herangezogen werden: Dispositionen, situationsspezifische Fähigkeiten und Performanz. Die Erfassung der situationsspezifischen Fähigkeiten ist dabei mit den größten Herausforderungen verbunden, da sich diese selbst nicht unbedingt beobachten lassen – vielmehr handelt es sich hierbei um psychologische Prozesse innerhalb einer Person in einer konkreten Situation.

Die Dispositionen für die situationsspezifischen Fähigkeiten sind wiederum als situationsübergreifend und als eine notwendige Voraussetzung für die Entwicklung situationsspezifischer Fähigkeiten zu verstehen. Ihnen kommt daher im Rahmen des Modells eine besondere Bedeutung zu, da sie am Anfang des Kompetenzerwerbs stehen – sie spielen jedoch auch in späteren Phasen der Professionalisierung noch immer eine zentrale Rolle. Zu den Dispositionen zählt einerseits das Wissen einer Person, was die Annahmen gängiger Modelle der Kompetenzentwicklung in der Lehramtsausbildung (vgl. Baumert & Kunter, 2006) explizit aufgreift. Andererseits betonen Blömeke und Kolleg*innen neben der Relevanz des Wissens auch den Stellenwert affektiv-motivationaler Aspekte, die ebenfalls als Handlungsvoraussetzungen zu bezeichnen sind. Letztere erzeugen die bereits bei Weinert (2001) formulierte Bereitschaft, das erworbene Wissen anzuwenden. Diese beiden zentralen Dispositionen werden in den folgenden Abschnitten näher erläutert sowie konkret für diagnostische Kompetenzen ausformuliert.



*Abbildung 3.1 Modell der Kompetenz als Kontinuum.
(Darstellung in Anlehnung an Blömeke et al. (2015, S. 7))*

Die im Modell formulierten Kompetenzaspekte und Wirkzusammenhänge lassen sich aus meiner Sicht mit dem umfassenden Ansatz zur Modellierung diagnostischer Kompetenz im vorausgehenden Kapitel (vgl. Abschnitt 2.2.2) gut zusammenfügen: Für diagnostische Kompetenz wird in dem Rahmenmodell von

Herppich und Kolleg*innen (2017, 2018) ebenfalls angenommen, dass gewisse Voraussetzungen – dort vornehmlich kognitiver Art – gegeben sein müssen, um einen professionellen diagnostischen Prozess durchführen zu können. Die einzelnen Schritte und Entscheidungen im diagnostischen Prozess, die gemäß des Anlasses und der Situationen unterschiedlich gestaltet werden, können als situationsspezifische Fähigkeiten verstanden werden, in der diagnostische Informationen wahrgenommen und interpretiert werden und schließlich die Grundlage für eine Entscheidung sind. Die Entscheidung am Ende des Prozesses ist eine konkrete, manifeste Diagnose, die nach Abschluss des diagnostischen Prozesses in weitere pädagogische Entscheidungen überführt wird. Je nachdem, wie explizit oder formell die Diagnose im Rahmen des Prozesses erstellt wird, können die situationsspezifischen Fähigkeiten bei diagnostischen Tätigkeiten mehr oder weniger gut sichtbar gemacht werden: Während sie bei informellen (Schrader, 2010) oder schnellen Typ-1-Diagnosen (Hetmanek & Van Gog, 2017) eher unbewusst verlaufen und entsprechend auch von außen kaum beobachtet werden können, können die Verarbeitungsprozesse der diagnostischen Informationen bei formellen oder analytischen Typ-2-Diagnosen explizit gemacht und damit vermutlich besser nachvollzogen werden.

3.2 Kognitive Voraussetzungen diagnostischer Kompetenz

Bereits in den Ausführungen zu den Kompetenzmodellen wurde verdeutlicht, dass Wissen eine wichtige kognitive Voraussetzung ist, damit Lehrkräfte in der Praxis kompetentes Verhalten zeigen. Gemeinhin wird *Wissen als Voraussetzung für Können* angesehen (Köller et al., 2016). In diesem Begriffspaar, Wissen und Können, spiegelt sich eine Dualität von Theorie und Praxis wider (Bromme, 2008), die in jüngster Zeit viel diskutiert und aufzuheben versucht wird: die Verbindung von Theorie und Praxis steht im Zentrum der Qualitätsoffensive Lehrerbildung (BMBF, 2016; Rothland, 2019). Damit diese Verbindung gelingt, müssen Wissen und Können, d. h. die Voraussetzungen und das Handeln gemeinsam gedacht werden. Auch das Modell der Kompetenz als Kontinuum (Blömeke et al., 2015) greift diesen Gedanken auf. Baumert und Kunter (2006) verdeutlichen, dass „professionelle Handlungskompetenz aus dem Zusammenspiel von spezifischem, erfahrungsgesättigten deklarativen und prozeduralen Wissen“ (S.481) entsteht. Mareike Kunter (2011) definiert die beiden Wissensarten genauer, indem sie deklaratives Wissen als grundlegendes, nicht-situationsspezifisches Wissen, und prozedurales Wissen als situationsspezifisches, ablauforientiertes Wissen, das theoriegeleitetes, professionelles Handeln ermöglicht, beschreibt.

Die Begriffe des *prozeduralen und deklarativen Wissens* werden auch bei der Beschreibung diagnostischer Expertise aufgegriffen: Hascher (2008) beschreibt, dass diagnostische Expertise eine handlungsbezogene Kompetenz sei, und dass diese nicht nur deklaratives, sondern auch prozedurales Wissen erfordere. Konkret brauche man zum Diagnostizieren Wissen darüber, welche Aspekte in die Diagnose einfließen und auf welche Weise zu diagnostizieren sei (vgl. Hascher, 2008). In diesen Ausführungen bleibt allerdings unklar, was genau unter deklarativem und prozeduralem Wissen in Bezug auf Diagnostik zu verstehen ist. Hesse & Latzko (2017) ergänzen, dass umfängliches Wissen im Rahmen praktischer Erfahrungen erweitert werden muss, um regelgerechte diagnostische Prozeduren ausführen zu können. Herppich und Kolleg*innen (2017, 2018) nennen eine ganze Reihe von verschiedenen Wissensarten, die im Rahmen der

Ausbildung von Lehrkräften erworben werden sollten: Konzeptuelles, situationales, prozedurales, strategisches und methodisches Wissen seien in unterschiedlichem Ausmaß Einflussfaktoren auf den diagnostischen Prozess. Dabei bleibt unklar, was die Autor*innen unter den von ihnen genannten Wissensbereichen genau verstehen und inwiefern diese konzeptuell strukturiert werden können. Anhand dieser Ausführungen wird deutlich, dass diagnostisches Wissen eine wichtige Voraussetzung für diagnostisch kompetentes Handeln ist, und dass deklaratives und prozedurales Wissen dabei eine Rolle spielen könnten – die Ausdifferenzierung und konzeptuelle Einordnung des diagnostischen Wissens stellt allerdings noch ein Arbeitsfeld bei der Konzeptualisierung diagnostischer Kompetenz dar (vgl. Herppich et al., 2017). Im Folgenden soll der Versuch unternommen werden, verschiedene Aspekte diagnostischen Wissens systematisch zusammenzustellen.

Dabei werden zunächst deklaratives (Abschnitt 3.2.1) und prozedurales Wissen (vgl. Abschnitt 3.2.2) als zwei grundlegende Arten diagnostischen Wissens unterschieden (Baumert & Kunter, 2006; Hascher, 2008), die im nachfolgenden Kapitel 4 bei der Vorstellung verschiedener Förderansätze diagnostischer Kompetenz ebenfalls berücksichtigt werden (Karing & Seidel, 2017). Eine Herausforderung bei der Definition besteht darin, dass die Begriffe in unterschiedlichen Domänen teilweise unterschiedlich verwendet werden und sich dadurch teilweise unterschiedliche Bedeutungen der Begriffe ergeben. Bei beiden Wissensarten werden das *kognitionspsychologische sowie bildungswissenschaftliche Verständnis* gegenübergestellt und auf Grundlage beider Verständnisse schließlich Arbeitsdefinitionen für die vorliegende Arbeit entworfen. Nachfolgend werden Möglichkeiten vorgestellt, wie der Aufbau deklarativen und prozeduralen Wissens gelingen kann (Abschnitt 3.2.3). Abschließend werden inhaltliche Dimensionen diagnostischen Wissens beschrieben, die als Bezugspunkte für den Erwerb diagnostischen Wissens beider Arten darstellen können (Abschnitt 3.3.3).

3.2.1 Deklaratives Wissen

In der *kognitiven Psychologie* wird deklaratives Wissen gemeinhin als verbalisierbares, erklärbares, abrufbares und explizites Wissen bezeichnet, während prozedurales Wissen ein nicht verbalisierbares Können, einen impliziten Prozess oder ein dem Gedächtnis nicht zugängliches Wissen darstellt (Cohen et al., 1985). Damit hat deklaratives Wissen neben der Explizierbarkeit eine weitere charakteristische Eigenschaft: Es ist “im Unterschied zu prozeduralem Wissen ‚manipulierbar‘[...], das heißt, dass deklaratives Wissen in Gedanken bewegt werden kann. Diese Fähigkeit zur Manipulation von Wissen ist vermutlich besonders in redundanten Situationen wichtig, in denen man sich also zwischen verschiedenen möglichen Verhaltensweisen entscheiden muss.“ (Stangl, 2018a) Dieses sprachbasierte Wissen kann in zwei Unterarten unterteilt werden. Es enthält einerseits Fakten- oder Begriffswissen, das mehr oder weniger unverbunden vorliegt, und andererseits konzeptuelles Wissen, das die Verbindungen der Fakten und Begriffe repräsentiert (Stern et al., 2016).

Die Abgrenzung und das Zusammenwirken beider Wissensarten werden insbesondere im Bereich der Arithmetik verständlich: *Faktenwissen* beschreibt ein schnell verfügbares Wissen, das direkt aus dem

Langzeitgedächtnis abgerufen werden kann (z. B. Additionsfakten bis 10, vgl. Landerl et al., 2017).

Konzeptwissen kann genauer umschrieben werden als die Beziehungen zwischen einzelnen „Wissensstücken“ sowie das Verständnis grundlegender Prinzipien einer Domäne (Rittle-Johnson & Alibali, 1999). Konzeptuelles Wissen entsteht durch die Verbindungen von einzelnen Informationen oder Fakten, deren Stärke wiederum moderiert die Abrufgeschwindigkeit aus dem Gedächtnis, wenn Antworten auf mathematische Probleme gefunden werden sollen (Goldman & Hasselbring, 1997). Konzeptuelles Wissen ist das Wissen über Gesetzmäßigkeiten und Prinzipien, durch das ein Verständnis für Operationen oder Prozeduren gelinge – es ist folglich die Grundlage dafür, dass Fakten- und Prozedurenwissen flexibel und sinnvoll, auch auf komplexe und unbekannte Aufgabenstellungen, angewendet werden können (vgl. Landerl et al., 2017).

Auch im *hochschuldidaktischen Bereich* werden in enger Anlehnung an die Lernzieltaxonomie nach Anderson und Kolleg*innen (Anderson et al., 2001) zwei Dimensionen von Wissen unterschieden, die dem deklarativen Wissen zugeordnet werden können. *Faktenwissen* kann als ein „Wissen um fundamentale Gegebenheiten eines Faches“ wie etwa Daten, Namen, Größen, einfache Formeln und fachspezifische Darstellungen verstanden werden (Glessmer & Lüth, 2016, S. 214). *Konzeptwissen* ist das „Wissen um eine Gesamtheit von Eigenschaften, die in einem begrifflichen, attributiven und kausalen Zusammenhang zueinander stehen“ (Glessmer & Lüth, ebd., S. 215). Es entsteht durch eine „methodische Verknüpfung beobachtbarer Gegenstände und Ereignisse unter allgemeinen Begriffen und Gesetzmäßigkeiten“ (S. 215) und bezieht sich somit – in Abgrenzung zu Faktenwissen – auf Modelle und Theorien. Konzeptuelles Wissen ermöglicht eine systematische Organisation von Informationen und die Abgrenzung verschiedener Wissensbereiche, es sei damit die Grundlage zum Differenzieren. Der Vermittlung konzeptuellen Wissens kommt insbesondere in der Phase des beginnenden Kompetenzerwerbs von Lehrkräften eine besondere Bedeutung zu: Dadurch, dass Lehramtsstudierenden durch ihre eigene Schulzeit bereits über ein breites konzeptuelles Wissen verfügen, das allerdings eher intuitiv und nicht mit fundiertem wissenschaftlichen Wissen verknüpft oder oft sogar widersprüchlich dazu ist, kommt der Professionalisierung dieses Wissens allgemein (vgl. Baumert & Kunter, 2006) wie auch hinsichtlich der Förderung diagnostischer Kompetenzen von Lehrkräften (Glogger-Frey & Renkl, 2017, vgl. 4.3) eine besondere Bedeutung zu.

Diese Zusammenschau zur Beschreibung deklarativen Wissens zeigt, dass die kognitionspsychologischen und bildungswissenschaftlichen Definitionen weitestgehend übereinstimmen. Deklaratives Wissen wird daher zusammenfassend definiert als verbalisierbares oder explizierbares Begriffswissen, das sich aus Fakten- und Konzeptwissen zusammensetzt (Stern, Schalk & Schumacher, 2016). *Faktenwissen, auch Was-Wissen genannt*, beschreibt das Wissen über fundamentale Gegebenheiten des Faches (vgl. Glessmer & Lüth, 2016). Es umfasst das Wissen um einzelne Begriffe, Fakten und Eigenschaften, die nicht unbedingt in Begriffsnetzwerke oder übergeordnete Strukturen eingebunden sind (vgl. Stern et al., 2016).

Konzeptwissen, auch Wozu-Wissen genannt, hingegen beschreibt genau dieses Wissen über Zusammenhänge von Fakten und Begriffen. Zentral ist hier das Wissen über eine Gesamtheit von Eigenschaften, die in Beziehung zueinander stehen (Glessmer & Lüth, 2016). Auf diese Weise wird ein Verständnis für

bestimmte Prinzipien einer Domäne möglich (Rittle-Johnson & Alibali, 1999). Deklaratives Wissen ist also grundlegendes Wissen in einer Domäne, das eine Voraussetzung für die Anwendung dieses Wissens darstellt.

3.2.2 Prozedurales Wissen

Prozedurales Wissen wird in der *Kognitionspsychologie* *gemeinbin* als ein unbewusstes Wissen bezeichnet, das durch Erfahrungen entsteht (Cohen et al., 1985). „Prozedurales Wissen wird auch als ‚Wissen wie‘ verstanden, also automatisiertes Handlungswissen, das durch Wiederholung entsteht und zumindest im Detail nicht kommuniziert werden kann“ (Stern, Schalk & Schumacher in Basiswissen Lehrerbildung S. 111). Damit beinhaltet es implizite „Fertigkeiten, die automatisch, ohne Nachdenken eingesetzt werden“ (Stangl, 2018b). Üblicherweise werden mit prozeduralem Wissen zuvor erlernte Bewegungs- oder Handlungsmuster, wie bspw. Fahrradfahren, in Verbindung gebracht, die unbewusst während der Ausführung abgerufen werden. Dieses Wissen über die Ausführung bestimmter Handlungen kann allerdings auch bewusst und damit verbalisiert werden und so auf andere Bereiche übertragen werden (Pezzulo, 2011).

In der *Mathematik* wird prozedurales Wissen als „flüssiges Wissen“ über Regeln, Algorithmen oder Prozeduren zur Lösung mathematischer Probleme verstanden (Goldman & Hasselbring, 1997): Es beschreibt das Wissen darüber, wie mathematische Berechnungen ausgeführt werden und auch die Anwendung der nötigen Schritte in der richtigen Reihenfolge. Prozedurales Wissen ist dabei nicht nur auf Rechenprobleme beschränkt, sondern umfasst auch die Lösung von Sach- oder Textaufgaben (Miller & Hudson, 2007). In dieser Domäne wird davon ausgegangen, dass sich prozedurales Wissen in sehr enger Verbindung mit konzeptuellem Wissen entwickle (Rittle-Johnson & Alibali, 1999; Landerl et al., 2017). Prozedurales Wissen wird dabei als ein bewusstes Wissen verstanden, das sich auf der Grundlage deklarativen Wissens etabliert. Prozedurales Wissen stelle Handlungssequenzen zum Problemlösen dar, welches nicht von deklarativem Wissen getrennt werden könne - vielmehr setze das Entwickeln und Durchführen von Prozeduren das konzeptuelle Verständnis arithmetischer Prozeduren voraus (Rittle-Johnson & Alibali, 1999).

Im *hochschuldidaktischen Bereich* wird prozedurales Wissen als ein Wissen über Prozeduren oder Prozesse verstanden, das an die Definitionen aus dem arithmetischen Bereich erinnert: Prozedurales Wissen kennzeichnet eine „systematische Problemlösung anhand vorgegebener Schritte“ oder die schrittweise Anwendung von Konzepten auf vorgegebene Art zur Lösung eines Problems (Glessmer & Lüth, 2016, S. 213). Prozedurales Wissen umfasse Kenntnisse über fachspezifische Kompetenzen, Algorithmen, Techniken, Methoden sowie Kriterien zur Bestimmung angemessener Prozeduren, z. B. Wissen um Schritte des wissenschaftlichen Prozesses. Hier bleibt unklar, ob prozedurales Wissen Handlungssequenzen beschreibt, die konzeptuelles Wissen voraussetzen (z. B. Anderson et al., 2001), oder ob beide Wissensarten unabhängig voneinander aufgebaut werden können (z. B. Glessmer & Lüth, 2016; für den Bereich der Arithmetik: Landerl et al., 2017). Grundsätzlich können bei prozeduralem Wissen

Wirkzusammenhänge gegeben sein - müssen aber nicht (vgl. Stern et al., 2016), während konzeptuelles Wissen eben genau dieses explizite Wissen zu Wirkzusammenhängen darstellt. Im Rahmen der Definition diagnostischer Expertise nach Helmke (2009) ist prozedurales Wissen auch ein „Wissen über Methoden zur Einschätzung von Schülerleistungen und zur Selbstdiagnose“, das durch konzeptuelles Wissen, das beispielsweise Kenntnisse über das Wirken von Urteilsfehlern und –tendenzen beinhaltet, ergänzt werden muss (ebd., S. 122). Auch wird eine enge Beziehung zwischen konzeptuellem und prozeduralem Wissen angenommen. Angesichts dieser Nähe zu konzeptuellem Wissen wird prozedurales Wissen im Kontext diagnostischer Kompetenz eher als explizites Wissen verstanden.

Zusammenfassend kann *prozedurales Wissen, auch Wie-Wissen* genannt, als handlungsbezogenes Wissen über Prozeduren oder Prozesse, die zur systematischen Problemlösung benötigt werden, verstanden werden. Aufgrund der Unterschiede in den zitierten Definitionen dahingehend, inwiefern dieses Wissen bewusst oder implizit handlungsleitend ist, wird hier das Wissen über Handlungssequenzen in den Vordergrund gestellt: Prozedurales Wissen meint demzufolge die schrittweise Anwendung von Konzepten zur Lösung eines Problems (Glessmer & Lüth, 2016), das sowohl automatisiert als auch explizit abgerufen werden kann. In der vorliegenden Arbeit wird prozedurales Wissen als Wissen über die Anwendung diagnostischer Methoden und Verfahren sowie die situationsangemessene Auswahl diagnostischer Prozeduren verstanden, die explizit oder in automatisierter Form erfolgen kann (vgl. auch automatisiertes vs. analytisches Diagnostizieren, Kap. 1.2.3.3).

3.2.3 Erwerb deklarativen und prozeduralen Wissens

Als Grundlage für Ansätze zur Förderung kognitiver Voraussetzungen diagnostischer Kompetenz (vgl. Kapitel 4) sollen nachfolgend grundlegende Möglichkeiten zum Aufbau deklarativen sowie prozeduralen Wissens vorgestellt werden. Für beide Wissensarten gestalten sich die Lernerfahrungen grundsätzlich in Abhängigkeit von den kognitiven Voraussetzungen sowie vom Vorwissen der Lernenden (Köller et al., 2016) und führen dazu, dass Wissen erweitert, (neu) strukturiert und geordnet wird (s. a. Stern et al., 2016, S. 108). Bei erfolgreichen Lernprozessen werden in der Regel deklaratives wie prozedurales Wissen etabliert, die – vermittelt über kognitive Prozesse und affektive Faktoren – situationsspezifisch abgerufen werden können (für eine Übersicht siehe Stern et al., 2016). Sowohl bei der Einspeicherung als auch beim Abruf spielt die Tiefe der Wissensrepräsentation eine Rolle. Um eine möglichst hohe Verarbeitungstiefe zu erreichen, sollten deklaratives und prozedurales parallel aufgebaut und gleichermaßen verknüpft werden. Der Leitgedanke der Verarbeitungstiefe, der für Lernzieltaxonomien charakteristisch ist (Bloom et al., 1956), wird im Abschnitt 3.2.3.2 noch einmal aufgegriffen, wenn der Wissenserwerb im Hochschulsetting näher ausgeführt wird. Zunächst sollen grundsätzliche Prinzipien zum Aufbau deklarativen und prozeduralen Wissens vorgestellt werden.

3.2.3.1 Prinzipien des Erwerbs deklarativen und prozeduralen Wissens

Deklaratives Wissen und prozedurales Wissen entstehen einerseits in wechselseitiger Interaktion, für sich genommen jedoch auf eine unterschiedliche Weise: Ein fundiertes *deklaratives Wissen* entsteht durch

Verbindungen zwischen einzelnen Begriffen, Fakten und Konzepten in Form von Netzwerken (vgl. Stern et al., 2016). Die entstehenden Begriffsnetzwerke erleichtern einerseits die Informationsverarbeitung und setzen damit andererseits kognitive Ressourcen für komplexere kognitive Prozesse frei. Die Bildung dieser Netzwerke kann dabei auf unterschiedliche Weise von Statten gehen: Es können sich etwa zwei Informationen verbinden, die bereits im Gedächtnis abgespeichert waren, oder neue Informationen können mit bereits bestehenden Informationen in Verbindung gebracht werden (vgl. z. B. Goldman & Hasselbring, 1997). Begriffsnetzwerke entstehen also immer auf Grundlage bereits vorhandenen Wissens, welches erweitert oder umstrukturiert werden kann. Um deklaratives Wissen auch anwenden zu können, ist eine *konzeptuelle Umstrukturierung* erforderlich (Helmke & Lenske, 2013; Schnotz, 2006). Elsbeth Stern und Kollegen (2016) betonen, dass eine Umstrukturierung des Begriffswissens nur dann erfolgen kann, wenn kognitiv aktivierende Lernformen gewählt werden, was beispielsweise über Selbsterklärungen (Lernende anregen, sich Dinge in eigenen Worten zu erklären), metakognitive Fragen (selbstständige Kontrolle des Lernstandes) oder die Arbeit mit kontrastierenden Einzelfällen (selbstständige Bearbeitung von kognitiv herausfordernden Aufgabe unter Vorgabe einer Musterlösung) erreicht werden kann. Jene Lernformen zeigten sich auf der empirischen Ebene überlegen im Vergleich zum herkömmlichen Vorgehen, bei dem Lerninhalte in Form von einem Input mit anschließenden Übungsaufgaben bearbeitet werden. Insbesondere bei Transferaufgaben können die Lernenden mit diesen Methoden bessere Ergebnisse erzielen – vermutlich deshalb, weil sie ein tieferes Verständnis der Lerninhalte erreichten.

Während deklaratives Wissen also vor allem durch konzeptuelle Umstrukturierungen und Netzwerkbildungen aufgebaut werden kann, die sich durch verschiedene instruktionale Methoden steuern lassen, wird der *Aufbau von prozeduralem Wissen* erst durch eine aktive Anwendung dieses Wissens möglich (vgl. Stern et al., 2016). Hierzu müssen Fakten- und Konzeptwissen derart gebündelt werden, dass die Bewältigung neuer Aufgaben möglich wird. Dieser Argumentation ist die Logik inhärent, dass Wissen eine Voraussetzung für Können darstellt (vgl. auch Abschnitt 3.1.1), was insbesondere für komplexe Handlungsmuster angenommen wird (z. B. Pezzulo, 2011). Unabhängig von der Domäne, in der prozedurales Wissen gebildet wird, entsteht und verfestigt sich dieses stets durch Wiederholung, die schließlich zu einer Automatisierung führt. Dabei werden Wissensteile verknüpft: „Prozeduralisierung kann als Verdichtung verstanden werden: Einzelne Wissens Elemente werden zu größeren Einheiten zusammengefasst“ (Stern et al., 2016, S. 111). Als typisches Beispiel für eine solche Verdichtung nennen Stern und Kollegen den Leseerwerb: Während ungeübte Leser*innen sich Wörter über eine serielle Buchstaben-Laut-Zuordnung mühsam erarbeiten müssen, können diese Zuordnungen mit zunehmender Übung besser und schneller erfolgen, bis geübte Leser*innen schließlich die entsprechenden Wortbilder mit der zugehörigen Bedeutung automatisch assoziieren können. Diese Automatisierung ist wichtig, weil dadurch Arbeitsgedächtniskapazitäten frei werden, die dann für komplexere Prozesse wie z. B. das Textverständnis genutzt werden können. Anhand dieses Beispiels wird deutlich, dass der Aufbau deklarativen und prozeduralen Wissens nicht unabhängig voneinander gedacht, sondern bestenfalls interaktiv gefördert werden sollten.

Fakten-, Konzept- und Prozedurenwissen sind drei zentrale Dimensionen von Wissen, die im Kontext schulischen Lernens (z. B. Stern et al., 2016) sowie beim Wissenserwerb in der Hochschullehre (z. B. Glessmer & Lüth, 2016) eine zentrale Rolle spielen. Wenn sie umfassend gefördert werden sollen, müssen zwei Prinzipien bedacht werden: Einerseits setzen sie einander voraus, d. h. Faktenwissen kann als Voraussetzung konzeptuellen Wissens angenommen werden und dieses wiederum kann durch Anwendung den Aufbau prozeduralen Wissens begünstigen (vgl. Abb. 3.2). Andererseits entwickeln sie sich in Interaktion miteinander, d. h. nur ein paralleler Aufbau aller drei Wissensarten führt zu einem tiefen Verständnis und ermöglicht Wissenstransfer. Damit belastbare und flexibel einsetzbare Wissensnetzwerke entstehen können, sollte in beiden Wissensbereichen eine möglichst hohe Stufe kognitiver Verarbeitung angestrebt werden – dies kann nur erreicht werden, wenn alle Wissensarten miteinander verbunden werden: Im Sinne einer umfassenden Kompetenzentwicklung „geht es darum, prozedurales *und* deklaratives Wissen aufzubauen, um die [...] kognitiven Prozesse zu ermöglichen. Dabei darf man sich den Aufbau von Wissen jedoch nicht wie das Besteigen einer Treppe oder einer Leiter vorstellen, sondern eher als einem [sic!] nicht kontinuierlich verlaufenden Prozess der Optimierung von Netzwerken aus deklarativem und prozeduralem Wissen.“ (Stern et al., 2016, S. 114, Hervorhebung v. Verf.)

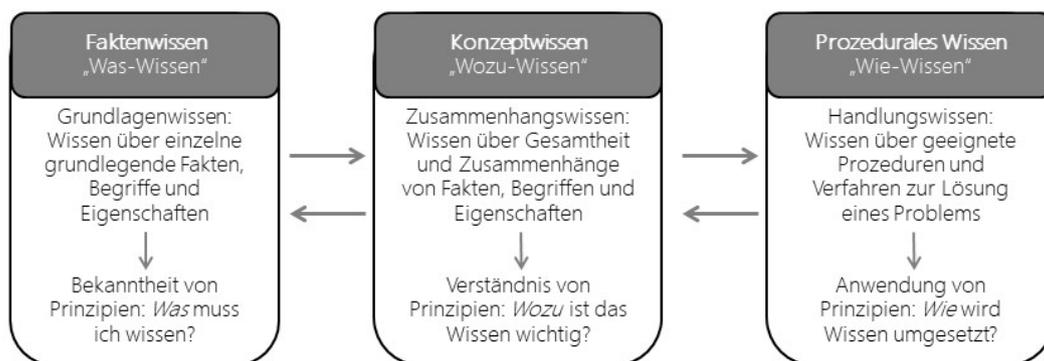


Abbildung 3.2 Wissensarten im Rahmen des Erwerbs professioneller Kompetenz

3.2.3.2 Wissenserwerb im Hochschulkontext

Bereits im vorausgehenden Abschnitt wurde angedeutet, dass die Verarbeitungstiefe für den Wissenserwerb eine zentrale Rolle spielt. Der Aspekt der Verarbeitungstiefe wird in Lernzieltaxonomien beim schulischen Lernen (z. B. Bloom et al., 1956; Anderson et al., 2001) als Voraussetzung für die erreichbare Wissenstiefe angesehen. Jene *Lernzieltaxonomien* bringen unterschiedlich komplexe kognitive Prozesse in eine Rangfolge und kommen inzwischen auch im Rahmen hochschuldidaktischer Überlegungen zur Anwendung, um deklaratives und prozedurales Wissen gezielt zu fördern (Glessmer & Lüth, 2016). Lernzieltaxonomien formulieren unterschiedlich hoch angesetzte Lernziele und dienen als Orientierung für die Entwicklung von Lern- und Prüfungsumgebungen, die eine gewisse Verarbeitungstiefe implizieren (Halbherr et al., 2016; Weidlich & Spannagel, 2014). In Tabelle 3.1 wird

eine für den hochschuldidaktischen Bereich ausformulierte Lernzieltaxonomie zusammengefasst, bei der sechs hierarchisch aufeinander aufbauende Prozessdimensionen angenommen werden (Glessmer & Lüth, 2016, in Anlehnung an Anderson et al., 2001): Die Prozesse des Erinnerns, Verstehens, Anwendens, Analysierens, Evaluierens und Erschaffens spiegeln unterschiedlich komplexe kognitive Verarbeitungsvorgänge wider und können daher auch als Kompetenzstufen aufgefasst werden, die Bezug nehmen zur Verarbeitungstiefe des Wissens. Ein hohes Kompetenzniveau bedeutet, dass komplexe kognitive Prozesse von den Lernenden bewältigt werden können. Der am wenigsten komplexe Prozess und damit das niedrigste Lernziel ist auf Stufe I das Erinnern. Dass man sich an bestimmte Lerninhalte erinnert, ist der Beginn jeder kognitiven Verarbeitung in einer Anforderungssituation. Um das Gelernte anwenden zu können, reicht Erinnern allein jedoch nicht aus, sondern es muss auch verstanden werden. Um schließlich die höchste Stufe des Erschaffens zu erreichen, müssen alle darunter liegenden Stufen erreicht worden sein. In ihrer Lernzieltaxonomie gehen Glessmer & Lüth (2016; abweichend von Anderson et al., 2001) davon aus, dass innerhalb jeder der drei Wissensarten unterschiedlich komplexe Verarbeitungsprozesse stattfinden können. Dementsprechend entwerfen sie in ihrer Arbeit einen Entscheidungsbaum, der 3 (Wissensarten) mal 6 (Lernzieldimensionen) mögliche Einordnungen von Übungs- oder Prüfungsfragen vorsieht. In allen drei Wissensdimensionen sollen möglichst komplexe kognitive Prozesse ausgelöst werden, um eine möglichst tiefe Verarbeitung der Wissensinhalte zu erreichen. Dass die sechs Prozessdimensionen tatsächlich in dieser Rangfolge etabliert werden, konnte dabei teilweise empirisch an Studierenden bestätigt werden (Kreitzer & Madaus, 1994), bedarf allerdings noch weiterer Forschungsbemühungen.

Anhand der Tabelle wird deutlich, dass die verschiedenen Lernziele i. d. R. mit bestimmten Aufgaben oder Arbeitsaufträgen verbunden sind. Die Einordnung von Übungs- wie Prüfungsaufgaben innerhalb der kognitiven Prozessdimensionen ist im Sinne der Validität von Prüfungsaufgaben relevant: Halbherr et al. (2016) fordern, dass Dozierende den Kompetenzaufbau von Studierenden stufenweise mit gezielten Fragen fördern und im Sinne des „Constructive Alignment“ die Übungs- und Prüfungsinhalte aufeinander abstimmen sollten. Dabei könne allerdings keine Gewissheit darüber erreicht werden, welche Denkprozesse jeweils genau mit den gestellten Fragen angeregt werden. Lernzieltaxonomien hätten sich zwar grundsätzlich bewährt, gäben aber wenig Hilfestellung für die korrekte praktische Umsetzung. Trotz des Versuchs, typische Verben für die einzelnen Stufen zu sammeln, um die wenig trennscharfen Definitionen der Niveaustufen zu kompensieren (vgl. Glessmer & Lüth 2016), sei es häufig Realität an Hochschulen, dass Übungs- und Prüfungsfragen nicht zusammenpassen, was einerseits die Validität der Prüfungen und andererseits die Lernmotivation der Studierenden reduziere (Halbherr et al., 2016).

Tabelle 3.1 Lernzielebenen im Überblick

Prozess	Definition	Anforderung	Typische Verben
Ebene I Erinnern	„abgespeicherte Inhalte aus dem Gedächtnis abrufen“	Wiedergabe von Inhalten: Frage und Antwort sind bekannt; z. B. Übersetzung eines Wortes, Definition eines Begriffes, Angabe einer Formel	definieren, reproduzieren, schildern, bezeichnen, aufsagen, angeben, benennen
Ebene II Verstehen	„aus vorliegendem Material direkt ablesbare Inhalte wiedergeben können“	Beschreibung von vorgegebenem Material, z. B. Werte aus einer Grafik ablesen (meint noch nicht die Interpretation oder den Vergleich solcher Werte, da dies weitere Denkschritte voraussetzt)	darstellen, beschreiben, bestimmen, formulieren, zusammenfassen, lokalisieren, erläutern
Ebene III Anwenden	„ein gegebenes Problem auf eine vorgegebene Art lösen“	Problemlösung durch Anwendung von Wissen mit bereitgestelltem Lösungsweg, z. B. Anwendung einer Rechenregel (Faktenwissen), Nachkochen eines Rezepts (prozedurales Wissen)	durchführen, berechnen, benutzen, herausfinden, anwenden, lösen, planen
Ebene IV Analysieren	„die Beziehung von Teilen zueinander und zu einer übergeordneten Struktur erläutern können“	Anhand vorgegebener Kriterien zu einer Einschätzung oder Lösung eines Problems kommen – dabei wird der Weg der Problemanalyse oder -lösung nicht mehr vorgegeben	testen, kontrastieren, vergleichen, isolieren, auswählen, unterscheiden, experimentieren, kategorisieren
Ebene V Evaluieren	„auf Basis von selbstgewählten Kriterien ein begründetes Urteil treffen“	Während bei der Analyse die Kriterien vorgegeben werden, anhand derer zu einem Urteil gelangt wird, sind diese Kriterien hier selbst zu wählen	beurteilen, argumentieren, voraussagen, wählen, begründen, prüfen, kritisieren, klassifizieren
Ebene VI Erschaffen	„Inhalte für einen selbst neu weiterentwickeln“	Inhalte oder Konzepte werden von Studierenden erstmalig (nicht generell) selbst erdacht oder neu zusammengeführt; z. B. Sammlung von Fakten, Entwicklung von Konzepten, Zusammenstellung von Methoden	sammeln, konstruieren, entwerfen, verbinden, konzipieren, zusammenstellen, entwickeln

Anmerkungen: Die Tabelle ist eine Zusammenfassung der Darstellungen von Glessmer & Lüth (2016, S. 211-213). Die einzelnen Lernziele bauen aufeinander auf, d. h. Erinnern ist die Grundlage für alle anderen Lernziele. Die nächsthöhere Lernzielebene schließt die vorausgehenden immer mit ein. Zur Lösung von Analysefragen müssen beispielsweise Grundbegriffe erinnert, der Sinn aus typischen Darstellungen entnommen und Probleme auf eine vorgegebene Art gelöst werden können. Je nachdem, welche Lernzielebene in Übungen oder Prüfungen intendiert wird, fallen Aufgabenstellungen und die Auswahl entsprechender Verben in den Arbeitsaufträgen unterschiedlich aus. Der Bezug auf Kompetenzen, die die Studierenden zur Lösung bestimmter Aufgabenstellungen befähigen sollen, gelingt durch Überlegungen zu konkreten Anforderungen von Arbeitsaufträgen aber besser.

Eine Lösung für die Problematik des Misalignments zwischen Prüfungs- und Übungsaufgaben kann Halbherr und Kolleg*innen zufolge in der Besinnung auf kompetenzorientierte Überlegungen liegen: Im Sinne der Kompetenzorientierung sollten Lernziele formuliert werden, deren Erreichen die Studierenden dazu befähigt, bestimmte Probleme erfolgreich lösen zu können. Dieser Gedanke findet sich in den

Definitionen sowie Anforderungen wieder, die Glessmer & Lüth (2016) in ihrer hochschuldidaktischen Schrift formulieren und die in Tabelle 3.1 ebenfalls wiedergegeben werden.

3.2.4 Inhaltsbereiche des diagnostischen Wissens

Nachdem in diesem Kapitel drei Arten oder Aspekte von Wissen (Fakten-, Konzept- und Prozedurenwissen) sowie sechs Dimensionen kognitiver Verarbeitungstiefe dieser Wissensaspekte (Lernzielebenen) vorgestellt wurden, soll abschließend zusammengetragen werden, welche konkreten Inhalte diese Wissensstrukturen in Bezug auf pädagogische Diagnostik abbilden. Da diagnostische Aufgaben von Lehrkräften umfangreich und anspruchsvoll sind, muss einerseits ein umfangreiches und fundiertes Wissen angestrebt werden (vgl. Hascher, 2008; Hesse & Latzko, 2017). Andererseits sollten gewisse Inhalte als Standards festgelegt werden, um sicherstellen, dass Handlungsweisen in der Praxis auf Basistheorien beruhen und damit mehr als „Handlungsrezepte“ sind (Oser, 1997). Die Festlegung von Wissensbereichen, die als *Basiswissen zur Entwicklung diagnostischer Kompetenz* gelten können, ist anhand des aktuellen Forschungsstandes sicher nicht erschöpfend möglich. Gerade vor dem Hintergrund, dass die Modellierung diagnostischer Kompetenz noch ein Arbeitsfeld ist, ist auch die Untersuchung der Fragen, *welche Wissensaspekte in welcher Weise und in welchem Ausmaß zur Bildung diagnostischer Kompetenz beitragen*, ein Desiderat (vgl. Herppich et al., 2017). Herppich und Kolleg*innen (ebd.) nennen selbst eine Reihe von unterschiedlichen Wissensfacetten, die zur Ausbildung diagnostischer Kompetenz notwendig sein könnten – und betonen dabei, dass zum gegenwärtigen Zeitpunkt unklar ist, welche Wissensaspekte tatsächlich maßgeblich für die Entwicklung diagnostischer Kompetenz sind und wie sie miteinander zusammenhängen. Auch verschiedene andere Autor*rinnen, die sich auf theoretischer oder praktischer Ebene mit dem Kompetenzbereich auseinandersetzen, nehmen ähnliche Auflistungen essentieller Inhalte diagnostischen Wissens vor (z. B. Helmke, 2009; Hesse & Latzko, 2017; Jäger, 2006; KMK, 2004; Jürgens & Lissmann, 2015; Oser, 2001; Retelsdorf & Müller, 2016).

Behrmann und Van Ophuysen (2017) machen im Rahmen des Vier-Komponenten-Modelles der Diagnose einen Vorschlag zur Strukturierung diagnostischen Wissens in zwei Domänen, die nachfolgend für die Zusammenstellung verschiedener Wissensfacetten übernommen werden. Um qualitativ hochwertig diagnostizieren zu können, brauche eine Person erstens ein hohes methodisches Wissen, und zweitens benötige sie ein fundiertes didaktisches und pädagogisch-psychologisches Wissen. *Methodisches Wissen als erste Dimension* umfasst Kenntnisse darüber, wie diagnostische Daten in bestimmten Situationen gewonnen und verarbeitet werden (Behrmann & Van Ophuysen, ebd.). Hierzu können folgende Wissensfacetten gezählt werden, die von anderen Autor*innen hervorgehoben wurden:

- Merkmale einer professionellen Diagnostik im Unterschied zu automatischen Urteilsprozessen und den Ablauf des diagnostischen Prozesses kennen (Hesse & Latzko, 2017; Retelsdorf & Möller, 2016)

- verschiedene diagnostische Methoden wie Tests, Beobachtungen und Gespräche kennen (Hesse & Latzko, 2017) und für Fragestellung angemessene Verfahren auswählen können (Helmke, 2009; Jürgens & Lissmann, 2015)
- messtheoretische Grundlagen zu Skalierung, Gütekriterien und Itemkennwerte kennen (Hesse & Latzko, 2017; Retelsdorf & Möller, 2016)
- ausgewählte Test- und Fragebogenverfahren, Quellen zur Beschaffung (Helmke, 2009) und Kriterien für die Auswahl geeigneter Erhebungs- und Auswertungsmethoden kennen (technologisches Wissen; Jäger, 2006)
- einen (eigenen) Test einsetzen, auswerten und rückmelden können (Helmke, 2009)
- Kenntnisse über die „Einordnung des individuellen Verhaltens unter dem Blickwinkel einer Vergleichsgruppe“ (Vergleichswissen; Jäger, 2006, S.90)
- Grundlagen der Lernprozessdiagnostik kennen (KMK, 2004)

Helmke (2009) weist darauf hin, dass Lehrkräfte zwar einerseits wissen müssen, welche Testverfahren es gibt und wie mit ihnen umzugehen ist, da der Einsatz von standardisierten Testverfahren bereits erheblich dazu beitragen kann, zu einer Diagnose hoher Güte zu gelangen. Daneben sollten Lehrkräfte aber auch wissen, dass Testleistungen nicht per se „bessere oder verlässlichere Messungen als das Lehrerurteil [...]“ (S. 123) seien, sondern dass die eingesetzten Methoden „entsprechend des jeweiligen Urteilsgegenstandes durchaus unterschiedlich ausfallen“ (S. 123; vgl. auch Oser, 2001). Vor dem Hintergrund des NeDiKo-Arbeitsmodells diagnostischer Kompetenz (Herppich et al., 2017, 2018) werden durch solide Kenntnisse im methodisch-technologischen Bereich drei Kernmerkmale einer hochwertigen pädagogischen Diagnostik sichergestellt: a) eine angemessene Methodenauswahl unter Beachtung der Gütekriterien von Testinstrumenten, b) eine systematische Datenerhebung und c) eine sinnvolle Datenauswertung (vgl. Förster & Souvignier, 2017). Dieses Wissen reicht jedoch noch nicht aus, um die gewonnenen diagnostischen Informationen situationsangemessen zu selektieren oder zu integrieren und sie zudem für das pädagogische Handeln nutzbar zu machen.

Erst das *didaktische und pädagogisch-psychologische Wissen als zweite Dimension* ermöglicht es, die für die jeweilige Entscheidung relevanten Informationen erkennen und verarbeiten zu können (Behrmann & Van Ophuysen, 2017). Dieses Wissen benötigen Lehrkräfte, um den entwicklungspsychologischen Stand von Schüler*innen hinsichtlich verschiedener Merkmale (u. a. Schulleistungen, Intelligenz) zu diagnostizieren, spezifische Lernschwierigkeiten zu erkennen und passende Interventionen ableiten zu können (Oser, 2001). Diesem Wissensbereich können u. a. folgende konkreten Kenntnisse zugerechnet werden:

- Wissen über Merkmale von Heterogenität, um Entwicklungsstände, -hindernisse und –fortschritte erkennen können (KMK, 2004)
- Wissen über Einflüsse auf Erleben und Verhalten (Bedingungswissen; Jäger, 2006) sowie konkrete Einflussfaktoren auf Lernprozesse und mögliche Auswirkungen auf Leistungen kennen (KMK, 2004)

- Wissen über Strategien, die zu einer Veränderung des Erlebens und Verhaltens führen können (Änderungswissen; Jäger, 2006)
- Fördermöglichkeiten für verschiedene Lernausgangslagen kennen – auch Begabungen erkennen und Möglichkeiten der Begabungsförderung kennen (KMK, 2004)
- unterschiedliche Kooperationspartner, auch anderer Professionen und Einrichtungen, kennen, um bei Fragestellungen mit inner- und außerschulischen Ansprechpartnern kooperieren zu können (KMK, 2004)

Zusätzlich zu diesen konkreten Wissensinhalten benötigen Lehrkräfte ein *Metawissen über ihre eigenen Kompetenzen und Grenzen*, das im Falle eines Kompetenzdefizits die Delegation des diagnostischen Auftrags an eine kompetente Person nach sich zieht (Kompetenzwissen; Jäger, 2006). Hierzu bedarf es eines Bewusstseins darüber, dass die eigenen diagnostischen Leistungen fehleranfällig und verzerrbar sind. Zu wissen und präsent zu haben, welche typischen Fehler und Verzerrungen bei Urteilen beobachtet werden können und welche Gütekriterien diagnostische Leistungen daher erfüllen sollten, sind daher ebenfalls essentielle Wissensbereiche zur Entwicklung diagnostischer Kompetenz (vgl. Helmke, 2009; Retelsdorf & Möller, 2016). Alle genannten Inhalte können grundsätzlich als notwendige Bedingung dafür angesehen werden, dass Lehrkräfte ein diagnostisch kompetentes Verhalten zeigen – sie sind jedoch keine hinreichende Bedingung. Damit das bestenfalls umfangreiche Wissen einer Lehrperson tatsächlich in kompetentes diagnostisches Handeln überführt wird, braucht es, wie bereits erwähnt (vgl. Abschnitt 3.1.2) weiterer Voraussetzungen, und zwar affektiv-motivationale Voraussetzungen. Diese werden im folgenden Abschnitt näher erläutert.

3.3 Affektiv-motivationale Voraussetzungen diagnostischer Kompetenz

Wissen allein ist noch kein hinreichendes Merkmal für Kompetenz, sondern nur *eine* Voraussetzung, die zu kompetentem Handeln führen kann: „Kompetenzfacetten entwickeln sich im Zusammenspiel der kognitiven und motivationalen Eingangsvoraussetzungen der (angehenden) Lehrkräfte mit den Inhalten und Rahmenbedingungen aus Aus- und Weiterbildung.“ (Köller et al., 2016, S. 20). Damit einerseits die Bereitschaft entsteht, an Lernangeboten teilzunehmen und sie zu nutzen – eine Vorbedingung dafür, dass Lernprozesse angestoßen werden – sind affektiv-motivationale Aspekte wichtig (Khan & Weiss, 1973; Kittel et al., 2017; Lipowsky, 2014). Andererseits spielen affektiv-motivationale Voraussetzungen wiederum eine Rolle dabei, dass das erworbene Wissen in Handeln überführt wird (vgl. Blömeke et al., 2015; Kunter, 2014; Lipowsky, 2014). Allgemein lässt sich *Motivation* als ein Wunsch oder eine „Energie“ beschreiben, die Menschen dazu bewegt, zu handeln. Da diese „Energie“ aus vielen unterschiedlichen Quellen stammen kann, kann Motivation nicht als singuläres Konstrukt, sondern als eine grundlegende Bereitschaft definiert werden, Handlungen zu vollziehen und diese Handlungen in eine bestimmte Richtung zu lenken (Ryan & Deci, 2000). Dieses „Sammelkonstrukt“ beschreibt einerseits psychologische Prozesse und andererseits Merkmale von Personen, sodass Motivation mehrdimensional verstanden und daher verschiedene Formen von Motivationen unterschieden werden müssen (Kunter, 2014).

Nachfolgend sollen zwei affektiv-motivationale Aspekte erläutert werden, die im Rahmen der Lehrkräftebildung allgemein sowie spezifisch für diagnostisches Verhalten herausgearbeitet wurden (vgl. Klug et al., 2016; Ohle et al., 2015; Westphal et al., 2018).

3.3.1 Einstellungen als Prädiktor für professionelles Handeln

Einstellungen bzw. Überzeugungen von Lehrkräften wird ein zentraler Stellenwert bei der Entwicklung professioneller Kompetenz beigemessen. Sie können einerseits ein Zeichen professioneller Entwicklung sein (z. B. Baumert & Kunter, 2006; Kunter et al., 2011). Andererseits spielen sie auch eine Rolle dabei, dass Interventionen für Lehrkräfte ihre Wirkung entfalten: Einstellungen bzw. Überzeugungen von Lehrkräften sollten daher ein zentrales Ziel von Bildungsprogrammen für Lehrkräfte sein, da sie erstens die Informationsverarbeitung positiv beeinflussen und zweitens im Rahmen von Fortbildungen verändert werden können (Lipowsky, 2014; Richardson, 1996). Einstellungen beeinflussen die Wahrnehmung und Verarbeitung von Begegnungen und daher auch die Informationsverarbeitung, das Denken, Schlussfolgern, Lernen und die Motivation einer Lehrkraft in besonderem Maße (Köller et al., 2000). Einstellungen bzw. Überzeugungen werden daher auf der einen Seite als eine Voraussetzung für die Nutzung von Fortbildungsangeboten durch Lehrkräfte betrachtet und andererseits ist die Veränderung dieser häufig ein zentrales Interventionsziel, das die Grundlage für Wissens- oder Verhaltensänderungen darstellt (Kittel et al., 2017; Kittel et al., 2018).

Einstellungen können allgemein bezeichnet werden als „eine psychologische Tendenz, die dadurch zum Ausdruck kommt, dass man einen bestimmten Gegenstand mit einem gewissen Grad an Zustimmung oder Ablehnung bewertet“ (Bohner, 2003; S. 267) oder eine Art erfahrungsgeladene Bereitschaft, die einen direkten oder zumindest sehr dynamischen Einfluss auf die Reaktion hat, die eine Person gegenüber Objekten oder Situationen zeigt (Richardson, 1996). Einstellungen haben also eine starke affektive Komponente und sind mehr oder weniger „zusammenfassende Bewertungen eines Gegenstands“ (Bohner ebd.). Einstellungen zeigen sich außerdem im kognitiven Bereich, und zwar in Form verbalisierter Überzeugungen oder Meinungen, sowie im behavioralen Bereich, d. h. in Form beobachtbaren Verhaltens oder anhand von Selbstaussagen über eigenes Verhalten (Dreikomponenten-Modell; Rosenberg & Hovland, 1960; Verplanken et al., 1998).

Überzeugungen werden im Rahmen dieses Dreikomponenten-Modells als ein Teilaspekt von Einstellungen betrachtet, nämlich als kognitive Komponente der Bewertung eines Einstellungsgegenstandes. In Arbeiten zur Lehrkräftebildung werden Überzeugungen (engl. „beliefs“) häufig als persönliche Glaubenssätze, Annahmen, Prinzipien oder Ideologien definiert, die eigene Entscheidungen beeinflussen (Richardson, 1996; Tillema, 1995). Überzeugungen beeinflussen außerdem den Wissenserwerb von Lehrkräften maßgeblich und haben so einen substantiellen prädiktiven Wert für die Vorhersage des Wissens einer Lehrkraft (für Mathematiklehrkräfte: Ernest, 1989; für Grundschullehrkräfte: Tillema, 1995). Auf der anderen Seite beeinflusst wiederum das Ausmaß des Wissens die Ausprägung von Überzeugungen: Bei der Überprüfung der Hypothese, dass die Vermittlung von konzeptuellem Wissen besonders dann gut gelinge,

wenn die Trainingsinhalte mit den Überzeugungen der Lehrkräfte übereinstimmten (Akzeptanz der Trainingsinhalte durch Kongruenz), ergaben sich Hinweise für die umgekehrte Wirkrichtung – nämlich eine Wissensänderung zu einer Überzeugungsänderung führen kann (Tillema, 1995). Diese Studie repliziert damit einen Befund von Ernest (1989), der für Mathematiklehrkräfte zeigen konnte, dass die Entwicklung von Wissen und Überzeugungen in einer Wechselwirkung stehen. Überzeugungen werden daher in verschiedenen Ansätzen der Lehrkräftebildung als ein zentraler Aspekt professioneller Kompetenz angesehen (z. B. Baumert & Kunter, 2006; Lipowsky, 2014).

Ernest (1989) modelliert in seiner Studie – neben dem Einfluss der Überzeugungen – auch den *Stellenwert von Einstellungen* für das Unterrichten. Einstellungen definiert er stark affektiv konnotiert: Einstellungen umfassen, ob eine Lehrkraft Mathematik als Fach gern möge, ob er/sie Unterhaltung und Interesse bei dem Lösen von Matheaufgaben empfinde und auch, ob die Lehrkraft Vertrauen in die eigenen mathematischen Fähigkeiten zeige. Sie seien das Ergebnis persönlicher Erfahrungen mit dem Unterrichtsfach und korrelieren positiv mit der Breite des Wissens einer Lehrkraft (allgemein über Unterricht sowie speziell über Mathematik). Der „Enthusiasmus für das eigene Fach“ sowie für Pädagogik im Allgemeinen spielt auch in aktuellen Angebot-Nutzungs-Modellen des Unterrichtens (z. B. Kunter et al., 2011; Lipowsky, 2014; Kittel & Rollett, 2017) eine doppelte Rolle: Die positiven Affekte führten dazu, dass Bildungsangebote genutzt werden, wodurch Kompetenzen für kompetentes Verhalten entwickelt werden. Positiven Einstellungen wird daher im Allgemeinen ein großer Stellenwert für das professionelle Handeln einer Lehrkraft zugeschrieben (Kunter et al., 2013). Für die Entwicklung diagnostischer Kompetenz bzw. Expertise (im Speziellen) können Einstellungen – neben anderen motivationalen Variablen und in Ergänzung zu Wissen – ebenfalls eine wichtige Rolle spielen. Einstellungen haben etwa einen entscheidenden Einfluss darauf, in welcher Frequenz und Qualität eine Lehrkraft diagnostische Handlungen zeigt (Ohle et al., 2015). Auch im Kontext der Nutzung datenbasierter Rückmeldungen konnten positive Einstellungen zur Vorhersage der Qualität diagnostischen Handelns einen bedeutsamen prädiktiven Wert einnehmen (Westphal et al., 2018). Es liegen allerdings auch erste Hinweise dafür vor, dass Einstellungen nicht in allen Stadien der Professionalisierung eine gleich große Rolle spielen: In einer Studie von Julia Klug und Kolleg*innen (2016) waren Einstellungen nur bei Referendar*innen ein bedeutsamer Prädiktor für die Vorhersage diagnostischer Kompetenz, bei Studierenden und Lehrkräften in der Praxis hatten sie hingegen keinen prädiktiven Wert. Klug et al. argumentieren, dass die Untersuchung des Zusammenhangs von Einstellungen im Zusammenspiel mit anderen Prädiktoren diagnostisch kompetenten Verhaltens ein wichtiges Desiderat für die zukünftige Forschung darstelle.

Ein weiterer Gegenstand zukünftiger Forschung betrifft die Frage, auf welche Weise positive Einstellungen, als ein Aspekt diagnostischer Expertise (vgl. Helmke, 2009), im Rahmen von Interventionen gefördert werden können. Wertvolle Hinweise hierzu geben Studien aus der Inklusionsforschung, die sich intensiv mit der *Veränderung von Einstellungen* beschäftigen. Der Grund hierfür ist, dass Lehrkräfte, die positiver zur Inklusion eingestellt sind, in stärkerem Maße eine inklusionsförderliche Unterrichtsgestaltung zeigen als weniger positiv eingestellte Kolleg*innen (Hartinger

et al., 2006). Das Ausmaß positiver Einstellungen gegenüber Inklusion ist wiederum von dem Ausmaß der Auseinandersetzung mit dem Thema abhängig: Bosse und Spörer (2014) konnten zeigen, dass einerseits die Studienphase angehender Lehrkräfte bzw. die bisherige Studienzeit ein bedeutsamer Prädiktor für die Höhe von inklusionsbezogenen Einstellungen ist (vgl. auch Egener, 2018). Andererseits demonstrieren sie, dass auch das Ausmaß persönlicher Erfahrungen und Kontakte mit behinderten Personen sich auf die Höhe inklusionszugewandter Einstellungen auswirkt (vgl. Forlin et al., 2010). Neben persönlichen Erfahrungen konnten insbesondere auch eigene Unterrichtserfahrungen als Auslöser von Einstellungsänderungen herausgestellt werden (Hellmich & Görel, 2014).

Die besondere Kraft persönlicher oder praktischer Erfahrungen als auslösendes Moment von affektiv bedingten Einstellungsänderungen konnte in Studien außerhalb des Inklusionskontextes nachgewiesen werden (z. B. Ernest, 1989; Richardson, 1996). Um Einstellungsänderungen nachhaltig zu induzieren, müsse Richardson (ebd.) zufolge stets auch ein kognitives Element hinzukommen: Das Nachdenken und Diskutieren über eigene Überzeugungen und Praktiken sowie mögliche Alternativen ermögliche erst eine intensive Auseinandersetzung, die zu Einstellungsänderungen führen könne, und zwar insbesondere dann, wenn sie mit praktischen Erfahrungen kombiniert werde. Die Verknüpfung von kognitiven Inputs mit praktischen Erfahrungen ist auch im Rahmen der Forschung zur Reflektionskompetenz von Lehrkräften als eine Methode zur wirksamen Reflektion hervorgehoben worden (siehe auch Lipowsky, 2014; Hascher, 2011). Richardson (ebd.) führt weiter aus, dass praktische Erfahrungen nicht unbedingt tatsächliche Praxiserfahrungen im Schulkontext voraussetzen, sondern dass diese Erfahrungen auch im Rahmen von beispielsweise Video-Fallbeispielen oder Diskussionen mit Lehrkräften möglich seien. Diese Annahme findet Unterstützung in Studien zur Inklusion, die zeigen konnten, dass bereits die Sensibilisierung für ein Thema im Rahmen von Lehrveranstaltungen in der Lage ist, Einstellungen zu verändern (Opalinski & Scharenberg, 2018).

3.3.2 Selbstwirksamkeitserwartungen und kompetenzbezogene Selbsteinschätzungen

In der Lehrkräftebildung allgemein wird den Selbstwirksamkeitserwartungen von Lehrkräften ein zentraler Stellenwert innerhalb der motivationalen Orientierungen zugeschrieben (z. B. Kunter, 2011, 2014; Schwarzer & Warner, 2014). Konkret für den Bereich der pädagogischen Diagnostik kann ebenfalls angenommen werden, dass Selbstwirksamkeitserwartungen von Lehrkräften als subjektive Einschätzungen ihrer eigenen Kompetenz einen entscheidenden Einfluss bei der Initiation diagnostischer Handlungen haben (vgl. Ohle et al., 2015). Auch wenn die Erforschung *diagnostikbezogener Selbstwirksamkeitserwartungen* gerade erst am Anfang steht und gesicherte Hinweise hierzu bislang ausstehen, konnte in einigen Studien gezeigt werden, dass diese selbstbezogenen Kognitionen in Bezug auf Diagnostik sowohl zur Vorhersage der Leistungen bei der Diagnostik von Lernprozessen (Klug et al., 2016) als auch für die Nutzung von Vergleichsarbeiten zur datengestützten Unterrichtsentwicklung herangezogen werden können (Westphal et al., 2018). Bereits frühere Studien, die sich mit datengestützter Unterrichtsentwicklung oder der Entwicklung professioneller Kompetenzen von Lehrkräften im Allgemeinen beschäftigten, setzten

Selbsteinschätzungsmaße zur diagnostischen Kompetenz als einen potentiellen Prädiktor ein (Helmke & Hosenfeld, 2005; Kunter et al., 2014).

Die Selbstwirksamkeit kann allgemein als Überzeugung über die eigene Kompetenz verstanden werden (vgl. Schwarzer & Warner, 2014). Der Begriff der Selbstwirksamkeitserwartung (Bandura, 1997, 2010) definiert selbstbezogene Kognitionen, die beschreiben, wie überzeugt eine Person von ihrem eigenen Können ist. Sie beschreiben ein Zutrauen in eigene Handlungskompetenzen, und zwar in Situationen, die mit hohen Anforderungen und ggf. hinderlichen Faktoren verbunden sind (vgl. Langfeld, 2017, Schwarzer & Warner, 2014). Schwarzer und Warner (ebd.) betonen, dass selbstwirksame Personen schwierige Aufgaben häufiger angehen und diese mit mehr Ausdauer durchführen als Personen mit einem geringen Selbstwirksamkeitserleben (Jerusalem & Schwarzer, 1992). Allgemeine Selbstwirksamkeitserwartungen (Schwarzer & Jerusalem, 1999) seien für Lehrkräfte eine wichtige Voraussetzung für deren Motivation und Stressbewältigung. Spezifische Selbstwirksamkeitserwartungen, die auf eine bestimmte Handlung umschrieben sind, können Lehrkräfte dazu bewegen, spezifische Anforderungen des Berufslebens zu bewältigen, auch bei Widerständen, und haben damit einen bedeutsamen Einfluss auf das unterrichtliche Verhalten der Lehrkraft und infolge dessen auch auf das Verhalten der Schüler*innen, die sie unterrichtet. Auch wenn gezeigt werden konnte, dass die Selbstwirksamkeit einen Einfluss auf die Leistungen einer Lehrperson hat, können sich tatsächliche Fähigkeiten und selbstwahrgenommene Fähigkeiten jedoch durchaus unterscheiden (Bandura, 1997; Schwarzer & Warner 2014). Diese selbstbezogenen Überzeugungen geben also an, wie kompetent eine Person sich gegenüber bestimmter Handlungen fühlt, und nicht unbedingt, wie kompetent sie tatsächlich ist (zur Diskrepanz zwischen selbsteingeschätzten und tatsächlichen Kompetenzen im Rahmen der Lehrkräftebildung vgl. z. B. Schladitz et al., 2015; Thoren et al., 2020).

Zur *Förderung von Selbstwirksamkeit* oder subjektivem Kompetenzerleben im Rahmen von Interventionen sollten Lehrkräfte die Möglichkeit bekommen, ihre Kompetenzen selbstständig und selbstgesteuert zu erwerben (vgl. Angebot- Nutzungsmodell beruflichen Lernens von Lipowsky, 2014). Im Bereich der Inklusionsforschung zeigte sich, dass die Anzahl von Begegnungen, d. h. durch persönliche Kontakte (hier mit behinderten Schüler*innen), sowohl inklusionszugewandte Einstellungen als auch die inklusionsbezogene Selbstwirksamkeit gesteigert werden können (Hellmich & Görel, 2014). Zur Entwicklung von Selbstwirksamkeit bräuchten die Lehrkräfte, laut Hellmich und Görel (ebd.), jedoch zusätzlich konkrete Übungsmöglichkeiten und Methoden, mit denen sie die Konzepte in die Praxis umsetzen können. Darüber hinaus könnte die Erhöhung von Wissen bspw. im Unterrichtskontext dazu führen, dass sich – in Interaktion mit den Einstellungen – auch die Selbstwirksamkeit steigern ließe: Je mehr Wissen beispielsweise eine Mathematiklehrkraft hat, desto höher sind ihre mathematikbezogenen Einstellungen und auch ihre Überzeugungen, wie gut sie das Fach unterrichten kann (Ernest, 1989). Schwarzer und Warner (2014) betonen in ihrem Überblick darüber, wie Selbstwirksamkeit grundsätzlich entsteht, dass die erfolgreichste Art, den Aufbau von Selbstwirksamkeitserwartungen voranzutreiben, „wohltdosierte Erfolgserfahrungen“ seien. Wenn es nicht möglich sei, eigene Erfahrungen zu machen,

könne auch erfolgreich anhand von Modellen gelernt werden, um Selbstwirksamkeit zu erwerben. Lehrkräfte bräuchten zur Entwicklung von Selbstwirksamkeit v. a. eine gezielte Unterstützung in genau den Bereichen, in denen sie Unsicherheiten verspüren. Anhand dieser recht allgemeinen Hinweise zeigt sich, dass die Entwicklung und der Wirknachweis konkreter Methoden zur Erhöhung der Selbstwirksamkeit – insbesondere im Bereich der pädagogischen Diagnostik – ein Desiderat für zukünftige Forschung im deutschsprachigen Raum ist.

3.4 Zusammenfassung und Zwischenfazit

Diagnostische Kompetenz, die als Expertise oder Prozesskompetenz verstanden wird, setzt deklaratives und prozedurales Wissen als kognitive Dispositionen voraus. Dieses *deklarative Wissen* beschreibt explizites Wissen über Fakten, Begriffe und Konzepte. Zum Aufbau einer soliden Wissensbasis von Lehrkräften, insbesondere zu Beginn des Kompetenzerwerbs und hinsichtlich der späteren Anwendbarkeit des erworbenen Wissens, muss zunächst konzeptuelles Wissen aufgebaut und konsolidiert werden, denn konzeptuelles Wissen ermöglicht das Verständnis von Fakten- und Begriffswissen. Die Anwendung konzeptuellen Wissens auf Probleme unter Nutzung bestimmter Methoden oder Schrittfolgen wird im hochschuldidaktischen Kontext als *prozedurales Wissen* bezeichnet. Prozedurales Wissen ist ein – mehr oder weniger bewusstes – Anwendungswissen, das durch Übung entsteht, und Kenntnisse über Prozeduren, Abläufe und die Anwendung von Methoden beinhaltet. Prozeduralem Wissen wird im Kontext der Entwicklung diagnostischer Expertise eine zentrale Bedeutung beigemessen. Neben einer soliden Wissensbasis ist im Sinne einer erweiterten Kompetenzdefinition auch das Vorhandensein *affektiv-motivationaler Aspekte* wichtig, damit das erworbene Wissen in professionelles Verhalten übersetzt wird. Als relevante motivationale Voraussetzungen des professionellen Handelns von Lehrkräften wurden in diesem Kapitel einerseits Einstellungen und andererseits Selbstwirksamkeit erläutert. Für den Bereich der diagnostischen Kompetenz von Lehrkräften liegen erste Hinweise darauf vor, dass die ausgewählten Konstrukte in Zusammenhang mit diagnostisch kompetentem Verhalten stehen. Welche Rolle die diagnostikbezogene Einstellungen und diagnostikbezogene Selbstwirksamkeit konkret für die Ausübung diagnostischer Tätigkeiten und damit für die Entwicklung diagnostischer Kompetenz spielen, ist ein Gegenstand für zukünftige Forschung.

— Diagnostisch kompetentes Handeln ist an Voraussetzungen geknüpft. Zu diesen Voraussetzungen gehören ein fundiertes Wissen, als kognitive Voraussetzung, und die Motivation, dieses Wissen in Handlungen umzusetzen. Quellen von Motivation für Diagnostik können Einstellungen und Selbstwirksamkeitsüberzeugungen sein. —

4 Ansätze zur Förderung diagnostischer Kompetenz

Im vorausgehenden Kapitel wurden grundsätzliche Überlegungen angestellt, welche Aspekte Interventionen zur umfassenden Förderung diagnostischer Kompetenzen berücksichtigen können und sollten. Bezugnehmend auf Modelle professioneller Kompetenz im Allgemeinen (Blömeke et al., 2015; vgl. Kap. 3.1.2) sowie diagnostischer Kompetenz im Speziellen (Herppich et al., 2017; vgl. Kap. 2.2.2) können grundsätzlich a) die Voraussetzungen, b) die prozessbezogenen Fähigkeiten sowie c) die Produkte des diagnostischen Prozesses als Ansatzpunkte für Interventionen herangezogen werden. In diesem Kapitel wird eine heterogene Sammlung bisheriger Förderansätze im deutschsprachigen Raum vorgestellt, die entlang dieser drei Ansatzpunkte strukturiert werden. Dabei wird deutlich, dass bislang kaum Förderansätze vorliegen, die spezifisch auf die *Zielgruppe der Lehramtsstudierenden* zugeschnitten sind und dass die Entwicklung von Interventionen für die Lehramtsausbildung ein Desiderat der Forschung darstellt.

Die Entwicklung und Evaluation von geeigneten Fördermaßnahmen ist allerdings nicht nur für die Zielgruppe der Lehramtsstudierenden, sondern in allen Phasen der Lehrkräftebildung, ein Desiderat der Forschung zu diagnostischen Kompetenzen. Bis vor zehn Jahren lagen kaum Ansätze zur Förderung diagnostischer Kompetenzen vor (vgl. Artelt & Gräsel, 2009; Schrader, 2009), obwohl dieses Anliegen seit spätestens Anfang der 2000er Jahre bildungspolitisch explizit verfolgt wird (z. B. KMK, 2004). Der aktuelle Forschungsstand kann den Bedarf an evaluierten Interventionen nicht angemessen decken (Hetmanek & Van Gog, 2017) – ein Missstand, der bereits vor zehn Jahren bemängelt wurde: Artelt & Gräsel (2009, S. 157) stellten bereits zu diesem Zeitpunkt im Editorial eines Themenheftes zu diagnostischer Kompetenz fest, dass der Forderung nach Förderung diagnostischer Kompetenz ein *unzureichender Kenntnisstand* gegenüberstehe, *wie und wann* diagnostische Kompetenz im Verlauf der professionellen Entwicklung entstehe und wie sie sich gezielt fördern lasse. Erste Ansätze zur Förderung diagnostischer Kompetenzen beziehen sich üblicherweise auf die Verbesserung der *Urteilsgenauigkeit* und damit auf einen eng umrissenen Aspekt diagnostischer Kompetenz, nämlich der Diagnoseleistung (Karing & Seidel, 2017). Nachfolgend wird zunächst der Ansatz zur Verbesserung der Urteilsgenauigkeit nach Helmke (z. B. Helmke, 2009) als Ansatzpunkt zur Verbesserung diagnostischer Produkte vorgestellt (Kap. 4.1.). Anschließend werden Ansätze zur Verbesserung konkreten diagnostischen Vorgehens im diagnostischen Prozess skizziert (Kap. 4.2) und abschließend werden Ansätze zur Förderung der Voraussetzungen kompetenten Diagnostizierens erläutert (Kap. 4.3). Insbesondere die Förderung der kognitiven Voraussetzungen diagnostischer Kompetenzen wird hier als vielversprechender Ansatzpunkt für die Zielgruppe der Lehramtsstudierenden betrachtet.

4.1 Verbesserung der Urteilsgenauigkeit

Die Verbesserung der Urteilsgenauigkeit ist einer der ersten Ansatzpunkte zur Förderung diagnostischer Kompetenz, die sich an den Produkten des diagnostischen Prozesses orientiert. Ausgehend von der

Bedeutsamkeit akkurater Urteile für die optimale Förderung ihrer Schüler*innen wurde dieser Ansatz im Zuge der Einführung von Vergleichsarbeiten entwickelt (Helmke, 2003; Helmke et al., 2004). Das zentrale Element dieses Ansatzes ist der bewusste Abgleich der Einschätzungen von Lehrkräften mit den realen Testergebnissen (vgl. Kap. 2.1). Der Ablauf der einzelnen Phasen kann wie folgt skizziert werden: (1) Klärung des Bereichs zur Prüfung eigener Diagnosekompetenz, z. B. Lesekompetenz; (2) Prognose der Schülerleistung oder des Personenmerkmals einschließlich schriftlicher Fixierung der Einschätzung für jede/n einzelne/n Schüler/in; (3) Erhebung der tatsächlichen Leistung oder des Merkmals anhand eines standardisierten Tests, einer Klassenarbeit oder der Selbsteinschätzung der Schüler*innen; (4) Vergleich der Vorhersage mit den empirischen Daten. Erste Erfahrungen mit diesem Vierersschritt zeigten bald, dass der reine Abgleich der prognostizierten mit den tatsächlichen Ergebnissen allein noch nicht zu einer substantiellen Erhöhung der Urteilsgenauigkeit der Lehrkräfte beitragen konnte. Folglich wurde dieser Ansatz um ein weiteres zentrales Element erweitert: (5) die Reflektion der auftretenden Diskrepanzen (Helmke, 2007). Dieses fünfschrittige Konzept, an dessen Ende die ausführliche Analyse des Vergleichs und die damit verbundene Suche nach Gründen für die Über- oder Unterschätzungen der Lehrkräfte steht, soll Helmke (ebd.) zufolge dafür sensibilisieren, dass informelle Diagnosen oftmals ungenau sind und daher stets hinterfragt werden sollten. Durch die festgestellten Diskrepanzen können auch systematische Urteilstendenzen der Lehrkraft, ein Mangel an differenzierten Informationen über die betreffenden Schüler*innen oder auch ein Mangel an theoretischem Wissen der Lehrkraft offenkundig werden, und somit auch Möglichkeiten zur Optimierung der diagnostischen Kompetenzen abgeleitet werden. Offen bleibt hierbei, welche Methoden genau und auf welche Weise zur Verbesserung der diagnostischen Kompetenzen beitragen können. Außerdem wird anhand des Vorgehens zur Förderung der Urteilsgenauigkeit deutlich, dass dieses Vorgehen nur für (angehende) Lehrkräfte geeignet ist, die sich in der Praxis befinden und im Rahmen dieser Praxistätigkeit Eindrücke über eine Gruppe von Schüler*innen gewinnen konnten, die sie nachfolgend mit objektiven diagnostischen Daten abgleichen. Für Lehramtsstudierende kommt ein solches Vorgehen allenfalls im Rahmen des Praxissemesters in Frage (vgl. Thoren et al., 2020).

4.2 Verbesserung kompetenten Vorgehens im diagnostischen Prozess

4.2.1 Ansätze zur Förderung von Fähigkeiten zur Erstellung formeller Diagnosen

Ausgehend davon, dass die Reflexion der eigenen Urteilsgenauigkeit allein noch nicht zu einer substantiellen Verbesserung diagnostischer Produkte beitragen kann, sondern zusätzlicher Interventionen bedarf, um Urteilsverzerrungen in Zukunft zu vermeiden, wurden für die praktische Aus- und Fortbildung von Lehrkräften weitere Förderansätze entwickelt. In ihrem Übersichtswerk zum Thema „Diagnostik für Lehrkräfte“ stellen Ingrid Hesse und Brigitte Latzko (2017) eine Reihe von Möglichkeiten zur *Förderung praxisrelevanter diagnostischer Tätigkeiten* vor, die sich in der Regel an Lehrkräfte aus der Berufspraxis richten und an der Verbesserung der Fähigkeiten zur Durchführung eines regelgeleiteten diagnostischen Prozesses ansetzen – mit dem Ziel, die Genauigkeit bei expliziten Diagnosen zu erhöhen. Fähigkeiten zur

Vorbereitung, Durchführung, Auswertung und Interpretation diagnostischer Verfahren können dem Konzept der diagnostischen Expertise zugeordnet werden (vgl. Kap. 2.1.2). Nachfolgend werden einige Fördermethoden dieser Sammlung vorgestellt, die m. E. zum Einsatz im Lehramtsstudium geeignet sein könnten.

Um praktische Fähigkeiten für die Anwendung von diagnostischen Verfahren zu erwerben, eignet sich beispielsweise das *selbstständige Erarbeiten von Testverfahren* in einer Kleingruppe. Durch die theoretische Erarbeitung des Manuals kann methodisches Grundlagenwissen vertieft und praxisorientiert angewendet werden. Im Rahmen von Rollenspielen bietet sich hier auch die Gelegenheit zur Reflexion der Testdurchführung. Eine weitere Möglichkeit zur Förderung praxisnaher diagnostischer Kompetenz ist die *Konstruktion eines eigenen informellen Tests*, wie er in Form von Klassenarbeiten in der Schulpraxis regelmäßig zum Einsatz kommt. Diese Übung ermöglicht es den Studierenden, Aspekte der Validität bei der Prüfung von Schulleistungen zu reflektieren. Eine weitere Methode zur Begleitung oder Vorbereitung der Bearbeitung diagnostischer Fragestellungen ist die *angeleitete Fallanalyse*. Bei diesem Ansatz wird ein in einer textbasierten Falldarstellung geschildertes Praxisproblem systematisch analysiert und reflektiert, indem die Vorgabe und Besprechung der jeweils relevanten fallbezogenen Informationen in einer klar strukturierten Reihenfolge erfolgen. Diese Struktur verfolgt die einzelnen Phasen des diagnostischen Prozesses, der hier gewissermaßen in verkürzter Form simuliert wird. Neben dem Aufstellen und Prüfen von Hypothesen und der hypothesengeleiteten Interpretation von Testergebnissen ist auch das Ableiten von pädagogischen Empfehlungen eine diagnostische Handlung, die im Rahmen dieser Übung mitbedacht wird.

Die genannten Maßnahmen beziehen sich auf relevante Aspekte, die insbesondere bei der Diagnostik von Lernbesonderheiten eine zentrale Rolle spielen könnten: Sie fokussieren den Umgang mit standardisierten diagnostischen Verfahren sowie formellen diagnostischen Informationen. Sie könnten damit zur beginnenden Förderung diagnostischer Expertise im Studium beitragen. Die umfängliche Entwicklung diagnostischer Expertise setzt sicherlich umfangreiche praktische Erfahrungen voraus (Schrader, 2017) und ist daher im Studium nicht zu erreichen. Die Arbeit mit diagnostischen Fällen, die nicht zwangsläufig aus der eigenen Berufspraxis stammen müssten, könnte bereits im Lehramtsstudium dazu beitragen, deklaratives wie prozedurales Wissen zu fördern (vgl. Hascher, 2008). Auf diesen Förderansatz wird im folgenden Abschnitt sowie im Abschnitt 4.3 noch einmal eingegangen. Für die hier vorgestellten Ansätze muss abschließend festgehalten werden, dass sie sich nach Angabe der Autorinnen zwar im Rahmen von regelmäßig durchgeführten Weiterbildungen für Lehrkräfte in der Berufspraxis bewährt haben (Hesse & Latzko, 2017) – jedoch stehen empirische Nachweise der Wirksamkeit dieser Maßnahmen bislang sowohl für praktisch tätige Lehrkräfte als auch Lehramtsstudierende noch aus.

4.2.2 Ansätze zur Förderung unterrichtsbezogener diagnostischer Tätigkeiten

Zwei Ansatzpunkte zur Verbesserung prozessbezogener diagnostischer Fähigkeiten, deren Wirksamkeit bereits empirisch bestätigt werden konnte, beziehen sich auf zwei spezifische Formen der Diagnostik im Rahmen unterrichtlichen Handelns von Lehrkräften. Beide Ansätze richten den Blick auf die

Verbesserung diagnostischer Prozeduren, die Lehrkräfte bzw. Referendar*innen in der Praxis an ihren realen Schüler*innen zur Anwendung bringen – mit dem Ziel, ihr Unterrichtshandeln zu optimieren.

Ein Ansatz, der gezielt die Verarbeitung diagnostischer Informationen verbessern will, ist der *Einsatz von Lerntagebüchern* (Klug, 2017). Durch den Einsatz der Lerntagebücher sollte die Reflexion von Lehrkräften in der Praxis bzw. im Referendariat in Bezug auf ihr diagnostisches Handeln gefördert werden, indem sie ihr eigenes diagnostisches Vorgehen dokumentieren, fortlaufend analysieren und schließlich verbessern sollen. Genutzt wird dabei der sogenannte „Reaktivitätseffekt“, d. h. dass sich das eigene Verhalten in eine gewünschte Zielrichtung entwickeln kann, indem es auf metakognitiver Ebene selbst beobachtet wird. Diesen Reaktivitätseffekt konnten Klug und Kollegen auch in einer aktuellen Trainingsstudie nachweisen, in der Lehrkräfte ihr eigenes diagnostisches Verhalten im Rahmen von Tagebüchern reflektierten und dokumentierten (vgl. Klug, 2017). In einem Schritt erhielten Referendar*innen sowie praktisch tätige Lehrkräfte in der Interventionsstudie über einen Zeitraum von mehreren Wochen Trainingssitzungen mit diagnostischen Themen, die aktiv mit verschiedenen Sozialformen und einschließlich der Bearbeitung eigener Fälle absolviert wurden. Ein Teil der Teilnehmer*innen führte zusätzlich ein strukturiertes Tagebuch, in dem anhand vorgegebener Leitfragen die Umsetzung der Trainingsinhalte und der Transfer in den eigenen Unterricht reflektiert werden sollten. Dabei zeigte sich, dass zwar gleichermaßen beide Versuchsgruppen vom Training hinsichtlich ihrer diagnostischen Kompetenz, ihres Wissens, ihrer Einstellungen, Selbstwirksamkeit und Gewissenhaftigkeit gegenüber Diagnostik profitierten. Eine genauere Analyse zeigte jedoch, dass diejenigen, die häufiger und tiefer reflektierten, in stärkerer Weise vom Training profitierten als diejenigen, die das Tagebuch nur selten und oberflächlich einsetzten. Die Ergebnisse dieser Studie belegen einerseits den förderlichen Effekt der Reflexion, andererseits wird deutlich, dass auch die Vermittlung diagnostischen Wissens einen entscheidenden Beitrag zur Förderung der diagnostischen Kompetenzen haben kann. Bevor die Reflexion als Methode eingeübt und eingesetzt wurde, absolvierten die Teilnehmer*innen ein zweigliedriges Wissenstraining. Dieses Wissenstraining zielte zunächst auf die Konsolidierung deklarativen Wissens ab (Trainingssitzungen), welches anschließend in Form von Falldarstellungen handlungsnah zur Anwendung kam. Erst in einem dritten Schritt wurde die reale Umsetzung diagnostischer Handlungen in der Praxis in den Tagebüchern reflektiert. Die zusätzliche Reflexion konnte hier über die Methoden zum Wissensaufbau hinaus zu einer Verbesserung diagnostischer Kompetenz beitragen – inwiefern sie allein zur Förderung diagnostischer Kompetenz geeignet ist, bleibt unklar.

Ein zweiter Ansatz zur Verbesserung diagnostischer Kompetenz setzt auf *formative Diagnostik*, die einerseits eine reliable Abbildung des Lernprozesses von Schüler*innen ermöglichen und Lehrkräfte dadurch bei der täglichen Unterrichtsplanung sowie der Individualisierung von Förderung unterstützen soll. Förster und Souvignier (2017) fokussieren dabei weniger die Verarbeitung diagnostischer Informationen durch die Lehrkraft, sondern die Nutzung formativer diagnostischer Informationen zur Ableitung pädagogischer Handlungen. Zentrales Erfolgskriterium ist hier die Leistungsentwicklung der Schüler*innen, die von den trainierten bzw. untrainierten Lehrkräften unterrichtet werden (Souvignier & Förster, 2011). In mehreren

groß angelegten Studien zur Wirkung formativer Diagnostik auf die Leistungsentwicklung von Viertklässler*innen arbeiteten die Lehrkräfte hierzu mit kurzen standardisierten Lesetests, die sie in unterschiedlicher Frequenz in ihren Klassen einsetzten (Förster & Souvignier, 2014; 2015). Die Lehrkräfte erhielten für alle Tests unmittelbare Rückmeldungen auf Ebene der Klasse und der einzelnen Schüler*innen. Neben diesen formativen diagnostischen Informationen erhielten die Lehrkräfte in den verschiedenen Studien zusätzlich verschiedene Interventionen. In einer ersten Studie (Förster & Souvignier, 2015) wurde untersucht, ob formative Diagnostik allein bereits ausreichend ist, um die Leseleistungen der Schüler*innen zu steigern, oder ob es einer zusätzlichen Intervention zur Unterstützung bei der Verarbeitung und Nutzung der diagnostischen Informationen bedarf. Hier zeigte sich erstens, dass formative Diagnostik an sich einen förderlichen Einfluss auf die untersuchten Schüler*innen hatte. Zweitens zeigte sich, dass kurze Schulungen mit nur wenigen Sitzungen zu Interventionsprinzipien noch nicht ausreichend waren, um darüber hinaus die Lesefähigkeiten der Schüler*innen zu verbessern. Anschlussstudien untersuchten daher explizit die Effekte einer diagnosebasierten Leseförderung, die durch die Lehrkräfte operationalisiert wurde (vgl. Studienüberblick bei Förster & Souvignier, 2017). Hier erhielten die Lehrkräfte in der Experimentalgruppe neben der bewährten Lernverlaufsdiagnostik auch ein ausführliches Manual, das neben Informationen zur Nutzung der formativen diagnostischen Daten auch konkrete Hinweise und Unterrichtsmaterial zur Förderung von Lesefähigkeiten enthielt. Es zeigten sich hier kurz- und längerfristige Verbesserungen der Lesefähigkeiten der Schüler*innen in der Experimentalgruppe, und diese waren deutlich größer als die Effekte in der ersten Studie mit lehrerseitiger Variation. Zum Erstaunen der Autor*innen gelang es jedoch trotz der umfangreichen konkreten Unterstützung der Lehrkräfte noch immer nicht in ausreichendem Maße, die Lernentwicklung der Schüler*innen optimal zu beeinflussen. Die Autor*innen selbst geben zu bedenken, dass die Lehrkräfte vor der Schulung möglicherweise nicht in ausreichendem Maße über die kognitiven Voraussetzungen verfügten, um die Inhalte der Schulung entsprechend umzusetzen. Die Ergebnisse dieser Studien zur Erforschung des Einflusses diagnostischer Handlungen auf den Lernverlauf von Grundschüler*innen zeigen damit vermutlich auch, dass die Güte der Anschlussmaßnahmen eine fundierte Wissensbasis in Form von deklarativem Wissen (Interpretation von Testwerten) sowie prozeduralem Wissen (Methoden der Förderung von Lesekompetenzen) voraussetzt. Die Untersuchung des Zusammenhangs der kognitiven Voraussetzungen der Lehrkräfte und der Wirksamkeit der Intervention sollten daher Gegenstand weiterer Studien sein.

4.3 Förderung kognitiver Voraussetzungen diagnostischer Kompetenz

In diesem Abschnitt sollen drei Förderansätze vorgestellt werden, die zur Förderung der kognitiven Voraussetzungen diagnostischer Kompetenz entwickelt wurden. Die Unterteilung von Förderansätzen dahingehend, ob sie an den *kognitiven Voraussetzungen* kompetenten Diagnostizierens oder am *Prozess des Diagnostizierens* selbst ansetzen, ist den Forscher*innen des NeDiKo-Netzwerks (vgl. Südkamp & Praetorius, 2017) zufolge zielführend für die Entwicklung weiterer Förderansätze. Zu den *kognitiven Voraussetzungen* zählen sie dabei insbesondere das zum Diagnostizieren relevante Wissen, welches sowohl

deklaratives als auch handlungsorientiertes, prozedurales Wissen umfasst. Die prozessbezogenen Kompetenzen auf der anderen Seite umfassen Fähigkeiten zur Auswahl und Verarbeitung diagnostischer Informationen im Entscheidungsprozess, die die Güte der konkreten diagnostischen Entscheidungen maßgeblich bestimmen (vgl. Kap. 4.2). Nachfolgend werden drei Ansätze vorgestellt, die auf die Förderung kognitiver Voraussetzungen diagnostischer Kompetenz fokussieren. Diesen ist gemein, dass sie den Stellenwert deklarativen und prozeduralen Wissens für die Entwicklung diagnostischer Kompetenzen betonen (Karing & Seidel, 2017). Sie beziehen sich allerdings teilweise auf sehr spezielle diagnostische Settings, sodass die empirischen Wirknachweise, sofern sie vorliegen, als vorläufig betrachtet werden müssen.

Stephanie Herppich und Kolleg*innen (2017a) versuchten, diagnostische Kompetenzen zu verbessern, indem sie *deklaratives Wissen über die Relevanz des Diagnostizierens* in der Situation des Tutorings vermittelten. Eins-zu-Eins-Tutoring, wie es beispielsweise in Nachhilfe-Situationen auftritt, ist eine Lernform, in der Lernende individuell unterstützt werden. Für diesen speziellen Diagnoseanlass zeigte sich, dass die Konsolidierung von deklarativem Wissen vor der Anwendung des Wissens wichtig ist. Hier erfolgte zunächst eine Wissensvermittlung über Diagnose- bzw. Instruktionsstrategien, welche anschließend durch die Vorgabe von videobasierten Lösungsbeispielen angewendet werden sollten. Die Wirkung und Strategien, die in den Videos für bestimmte Tutoring-Situationen dargestellt wurden, sollten sich die Tutor*innen selbst erklären. Es zeigte sich, dass die bloße Vermittlung von neuen Strategien, auch wenn diese selbst erarbeitet werden (aktivierende Lernform; vgl. Stern et al., 2016), zur Verbesserung diagnostischen Verhaltens noch nicht ausreichten (Herppich et al., 2013). In Folgestudien wurde daraufhin die Wissensvermittlung um eine Abfrage des erworbenen Wissens ergänzt, bei der die Studierenden direktes Feedback zu den Items erhielten, indem die richtige Lösung präsentiert und erklärt wurde (vgl. Herppich et al., 2017a). Dabei zeigte sich, dass die Konsolidierung von deklarativem Wissen durch die Methode des „Testings“ sowie durch Feedback in Form einer Musterlösung (Hattie & Timperley, 2007) dazu beitrug, die diagnostische Kompetenz in der Interventionsgruppe zu erhöhen, wobei offen bleibt, durch welche dieser Bausteine diese verbesserte Leistung im Tutoring-Prozess nun tatsächlich erreicht wurde. Die Studie ist dennoch ein Hinweis darauf, dass die Konsolidierung des deklarativen Wissens vor der Anwendung auf Fallbeispiele eine wirksame Strategie zur Verbesserung diagnostischer Kompetenz darstellt. Außerdem konnte gezeigt werden, dass das Training in beiden Varianten die Motivation zum Diagnostizieren erhöhen kann.

Ein weiterer Ansatz, der die Relevanz einer fundierten Wissensbasis zur Entwicklung diagnostischer Kompetenz herausstellt, stammt von Glogger-Frey und Renkl (2017). Die Autor*innen adressieren die *Verbesserung konzeptuellen Wissens im Kontext der Lernstrategien-Diagnostik*, indem sie instruktionale Methoden zum Aufbau von Vorwissen mit anschließenden beispielbasierten Kurztrainings kombinieren. Die Autor*innen machen deutlich, dass es bei angehenden Lehrkräften besonders wichtig ist, ihr unterrichtliches Handeln durch die Vermittlung von theoretisch fundiertem Grundlagenwissen vorzubereiten. Das beispielbasierte Lernen erfolgt im ersten Schritt durch die Analyse von

Lerntagebüchern von Schüler*innen, anhand derer verschiedene Lernstrategien erkannt und beurteilt werden sollen. Anschließend sollen die zuvor erlernten Prinzipien (Regeln, Definitionen) erklärt werden, um danach mehrere Aufgabenstellungen inkl. ausgearbeiteter Lösung bspw. mithilfe der Methode der Selbsterklärung zu bearbeiten. Die Studierenden wenden hierbei aktiv Wissen über Lernstrategien an, die sie zuvor anhand von Beispielen auf theoretischer Ebene kennengelernt haben. Konkrete empirische Studien zu den hier vorgestellten computerbasierten Lernumgebungen stehen zwar noch aus, aber das beispielbasierte Lernen konnte grundsätzlich als effektive Methode für Lehramtsstudierende in der Phase des beginnenden Kompetenzerwerbs herausgestellt werden (Renkl, 2011, 2014).

Abschließend soll eine Methode zur Förderung diagnostischen Anwendungswissens zur Durchführung des diagnostischen Prozesses vorgestellt werden, die der *diagnostischen Ausbildung von Mediziner*innen* entlehnt ist und aktuell für die Lehramtsausbildung erprobt wird (Hetmanek & Van Gog, 2017). Dieser Ansatz thematisiert das Wissen über die Anwendung diagnostischer Prozeduren, welche je nach diagnostischer Situation unterschiedlich ausfallen können (vgl. Abschnitt 1.2.4.1). Der Unterscheidung in schnelle vs. akkurate Diagnosen (Typ-1- vs. Typ-2-Diagnosen; vgl. Kap. 1.2..3.3) entsprechend leiten Hetmanek und van Gog (ebd.) zwei allgemeine Ansätze zur Förderung diagnostischer Kompetenz ab: die De-Automatisierung schnellen diagnostischen Vorgehens vs. die Automatisierung von akkuraten Diagnoseprozeduren. Zur Verbesserung der Qualität bei schnellen Diagnosen, wie sie im laufenden Unterrichtsgeschehen üblich sind (vgl. informelle Diagnose; Schrader, 2010; Hascher, 2008), schlagen die Autoren die Methode der strukturierten Reflexion vor. Diese Methode hat sich im Rahmen von Fortbildungen für niedergelassene Ärzt*innen als sehr erfolgreich erwiesen (Mamede et al., 2008) und wurde unlängst auch in der Ausbildung von Medizinstudierenden erfolgreich erprobt (Mamede et al., 2012, 2014). Entwickelt wurde die Methode der strukturierten Reflektion zur Reduzierung von Diagnosefehlern, die bei schnellen Diagnosen auftreten, da diese heuristisch und unbewusst ablaufen. Ziel des Ansatzes ist es daher, diese unbewusst entstehenden Spontandiagnosen systematisch zu reflektieren und auf einem analytischen Wege zu überprüfen. Die Reflexionskompetenz wird durch ein strukturiertes Vorgehen im Rahmen der Bearbeitung verschiedener Falldarstellungen trainiert, in der die Abwägung hypothesenkonformer und –widriger Befunde sowie das Mitbedenken erwartbarer, aber abwesender Befunde die zentralen Arbeitsaufträge sind. Diese Art der strukturierten Fallarbeit, in der explizit eine Erhöhung der Sorgfältigkeit durch die Reflektion der Diagnosen trainiert wird, könnte ein fruchtbarer Ansatz für die Erhöhung der Diagnoseakkuratheit für die Berufsgruppe der Lehrkräfte sein, auch wenn diskutiert werden kann, ob pädagogische Diagnosen ähnlich komplex sind wie medizinische Diagnosen.

Die drei vorgestellten Ansätze lassen die Vermutung zu, dass nicht nur der Erwerb deklarativen, sondern auch der Aufbau prozeduralen Wissens bereits im Lehramtsstudium möglich sein könnte. Zwar stehen konkrete empirische Nachweise zum Aufbau prozeduralen Wissens bei Lehramtsstudierenden noch aus, allerdings ergeben sich Hinweise, dass der Erwerb prozeduralen Wissens nicht nur im Rahmen tatsächlichen diagnostischen Geschehens erfolgen kann, sondern auch in *fallbasierten Lerngelegenheiten* trainiert werden kann. In vorgestellten Konzepten wurde deklaratives Wissen in Falldarstellungen

angewendet, die hier in Form von Videos (Herppich et al., 2017a) oder als schriftliche Fallbeispiele (beispielbasiertes Kurztraining: Glogger-Frey & Renkl, 2017; strukturierte Reflexion: Hetmanek & Van Gog, 2017) vorgegeben wurden. In Kombination mit dem Aufbau einer fundierten Wissensbasis könnten diese Anwendungen kompetente diagnostische Handlungen praxisnah vorbereiten. Der Einsatz fallbasierter Lerngelegenheiten im Zusammenhang mit der Konsolidierung deklarativen Wissens scheint zusammengefasst ein zielführender Ansatz zu sein, diagnostische Kompetenzen von angehenden Lehrkräften bereits im Lehramtsstudium zu fördern. Entsprechende Wirksamkeitsnachweise, auch im Rahmen verschiedener diagnostischer Situationen, sollten Gegenstand zukünftiger Forschung sein.

4.4 Zusammenfassung und Zwischenfazit

Die Förderung diagnostischer Kompetenz ist nach wie vor ein Desiderat der Forschung und so wird in diesem Kapitel eine sehr heterogene Sammlung von Konzepten vorgestellt, die vorläufige Hinweise zu den Prinzipien der Förderung diagnostischer Kompetenz geben. Erste Ansätze zur Förderung diagnostischer Kompetenz fokussieren das Ergebnis diagnostischen Handelns – die *Verbesserung der Diagnosegenauigkeit*. Dieser Ansatz berücksichtigt allerdings nicht, auf welchem Weg eine akkurate Diagnose erstellt wird. Konzepte, die an den situationsspezifischen Fähigkeiten zur *Durchführung des diagnostischen Prozesses* ansetzen, adressieren zumeist praktisch tätige Lehrkräfte und widmen sich praktischen diagnostischen Tätigkeiten. Für Studierende sind diese Übungen nur bedingt geeignet, weil sie oftmals keinen Zugang zu realen Schüler*innen haben. Für Lehramtsstudierende könnten daher insbesondere jene Methoden geeignet sein, die auf die *Verbesserung der kognitiven Voraussetzungen* für diagnostisch kompetentes Handeln abzielen. Zur Förderung kognitiver Voraussetzungen sind inzwischen einige Ansätze entwickelt und teilweise auch bereits erfolgreich evaluiert worden, die den Aufbau deklarativen und prozeduralen Wissen kombinieren, indem zunächst Grundlagenwissen vermittelt wird, welches anschließend in fallbasierten Formaten mehr oder weniger praxisnah zur Anwendung kommt. Jene praxisvorbereitenden Anwendungen könnten auch Interventionen ergänzen, die die Optimierung des diagnostischen Prozesses in der Praxis fokussieren. Studien zu prozessorientierten Interventionen zeigten eindrücklich, dass eine solide Wissensbasis hierfür eine essenzielle Voraussetzung ist. Sie zeigen ferner, dass das Ableiten pädagogischer Handlungen anhand von Diagnosen trainiert werden kann und dass es hierzu spezifischer Übungen zur Nutzung diagnostischer Informationen bedarf.

- Die Entwicklung und Evaluation von Ansätzen zur Förderung diagnostischer Kompetenz im Lehramtsstudium sind ein Desiderat. Erfolg versprechend scheinen insbesondere Ansätze zu sein, die die kognitiven Voraussetzungen adressieren. Die Festigung deklarativen Wissens in Kombination mit anschließenden praxisnahen fallbasierten Übungsmöglichkeiten könnte dazu beitragen, die Professionalisierung diagnostischer Prozesse in der Praxis vorbereiten. —

5 Hinführung zu den empirischen Studien

Ziel meiner Dissertation war die Entwicklung, Erprobung und Evaluation einer Lerngelegenheit, die geeignet ist, diagnostische Kompetenzen bereits im Lehramtsstudium zu fördern. Die Herausforderung bestand darin, eine praxisnahe Lerngelegenheit zu schaffen, die dabei ohne den Zugang zu Realdaten von Schüler*innen auskommt. Im vorausgehenden Kapitel (Kap. 4) wurde einerseits deutlich, dass für die Zielgruppe der Lehramtsstudierenden bislang kaum evaluierte Ansätze zur Förderung diagnostischer Kompetenz vorliegen. Andererseits konnten erste Ansätze aus dem deutschen Sprachraum zusammengestellt werden, die sowohl über das Konzept der Urteilsgenauigkeit hinausgehen als auch für die Förderung Lehramtsstudierender vielversprechend sind. Für Lerngelegenheiten in der Phase des beginnenden Kompetenzerwerbs sollte entsprechend aktueller Kompetenzmodelle (z. B. Blömeke et al., 2015; Herppich et al., 2017) die Förderung kognitiver Voraussetzungen im Mittelpunkt stehen, wobei auch affektiv-motivationale Voraussetzungen diagnostischer Kompetenz berücksichtigt werden sollten (vgl. Kap. 3). Im Fokus der entwickelten Lerngelegenheit stand der *Aufbau prozeduralen diagnostischen Wissens*, welches als essenzielle Voraussetzung für späteres professionelles diagnostisches Handeln betrachtet wird (diagnostische Expertise; vgl. Hascher, 2008; Hesse & Latzko, 2017). Da das konkrete diagnostische Vorgehen in unterschiedlichen diagnostischen Situationen unterschiedliche Gestalt annehmen kann, kann die Auswahl und Durchführung eines diagnostischen Prozesses, der für die Reichweite der diagnostischen Situation angemessen ist, ein Zeichen diagnostischer Kompetenz sein (vgl. Kap. 2). Die Feststellung von Lernbesonderheiten ist ein verbindlicher Diagnoseanlass, der entsprechend ein sorgfältiges, regelgeleitetes diagnostisches Vorgehen erfordert (z. B. Karst et al., 2017). Die hier entwickelte Lerngelegenheit sollte den Studierenden daher auch ermöglichen, Aspekte des regelgeleiteten diagnostischen Prozesses (Jäger, 2006), der für die Feststellung von Lernbesonderheiten anzustreben ist, im Rahmen von realistischen Falldarstellungen kennenzulernen. Die Studierenden sollten dadurch für die Herausforderungen bei der Feststellung von Lernbesonderheiten sensibilisiert werden sowie die Gelegenheit erhalten, ihr bereits erworbenes diagnostisches Wissen auf realistische diagnostische Aufgaben anzuwenden. Auf diese Weise sollte der Aufbau prozeduralen Wissens unterstützt werden.

Studie 1: Entwicklung des Fallinventars

In der ersten Studie dieser Dissertation wird die *Entwicklung und Pilotierung* einer fallbasierten Lerngelegenheit beschrieben, deren Herzstück ein diagnostisches Fallinventar ist. Hierin werden Schüler*innen mit und ohne Lernbesonderheiten mittels urteilsrelevanten und -irrelevanten Informationen beschrieben, auf deren Grundlage die Studierenden Diagnosevermutungen abgeben sowie, den Erfordernissen pädagogischer Diagnostik entsprechend (vgl. Kap. 1, Abb. 1.3), insbesondere pädagogische Anschlusshandlungen ableiten. Ziel der Pilotierung war es herauszufinden, inwiefern die entwickelten Falldarstellungen geeignet sind, um die Studierenden zu einer aktiven und intensiven Arbeit anzuregen, die wiederum Voraussetzungen für Veränderungen im kognitiven und motivationalen Bereich darstellen (vgl. Kap. 3). Dabei wurde auch untersucht, inwiefern die Studierenden in der Lage sind,

korrekte Diagnosen und Empfehlungen für die Falldarstellungen zu vergeben, da die diagnostischen Kompetenzen von Lehramtsstudierenden in Bezug auf die Feststellung von Lernbesonderheiten bislang kaum erforscht sind. Die Pilotierung diente damit der Vorbereitung der zweiten Studie.

Studie 2: Erprobung des Fallinventars

In der zweiten Studie wurde das *Fallinventar als szenariobasierte Testumgebung* eingesetzt, wodurch einerseits die Eignung des Fallinventars nach einer Überarbeitung und Erweiterung der Fallvignetten überprüft wurde. Andererseits stand die Analyse des konkreten Vorgehens der Studierenden bei der Bearbeitung der Fallvignetten im Zentrum. Dabei wurde erstens der Frage nachgegangen, wie schwer den Studierenden die Abgabe einer korrekten Diagnosevermutung im Vergleich zur Ableitung korrekter Anschlusshandlungen fällt. Ergänzend wurde die subjektive Sicherheit bei beiden Entscheidungen erfasst, um zu analysieren, inwiefern die Studierenden ihre Fähigkeiten in Bezug auf die Diagnostik von Lernbesonderheiten realistisch einschätzen. Zweitens wurde untersucht, ob die Studierenden bei ihren Entscheidungen von der normativen Reihenfolge im pädagogisch diagnostischen Prozess profitieren, bei der Empfehlungen auf der Grundlage von Diagnosen abgeleitet werden (vgl. Kap. 2). Drittens wurde der Frage nachgegangen, welchen Einfluss die Verwendung urteilsrelevanter, formeller Diagnoseinformationen im Gegensatz zu urteilsirrelevanten Diagnoseinformationen auf die Lösungsraten beim Vergeben von Diagnosen und Empfehlungen hat. Es wurde vermutet, dass die Studierenden für urteilsirrelevante, potentiell verzerrende diagnostische Informationen anfällig sind und ihre Trefferquote bei Diagnose- und Empfehlungsentscheidungen dadurch beeinträchtigt ist (vgl. z. B. Krolak-Schwerdt et al., 2009).

Studie 3: Evaluation des Fallinventars

In der dritten Studie dieser Dissertation wird die Evaluation des Fallinventars als *fallbasierte Lerngelegenheit für Lehramtsstudierende* beschrieben. Hierzu wurden zunächst die Reaktionen der Studierenden in Kontrast zu einer klassischen aufgabenbasierten Übung untersucht. Im Fokus der Evaluationsstudie stand die Frage, ob das Fallinventar zur Förderung prozeduralen diagnostischen Wissens beitragen kann. Die klassische aufgabenbasierte Übung wurde eingesetzt, um deklaratives Wissen zur Diagnostik von Lernbesonderheiten zu fördern. Beide Lerngelegenheiten sowie ein selbstentwickelter Szenariotest zur Erfassung von Wissen im Kontext der Diagnostik von Lernbesonderheiten wurden mithilfe eines Expert*innen-Ratings hinsichtlich kognitiver Prozessdimensionen (vgl. Glessmer & Lüth, 2016) validiert. Die differenziellen Effekte beider Übungen auf prozedurales wie deklaratives Wissen im Szenariotest wurden durch Implementierung einer dritten Bedingung analysiert, in der zunächst die aufgabenbasierte und anschließend die fallbasierte Übung bearbeitet wurde. Es wurde erwartet, dass sich insbesondere die Kombination beider Übungen besonders förderlich sowohl auf die kognitiven Voraussetzungen als auch auf affektiv-motivationale Voraussetzungen auswirkt. Die Veränderung auf beiden Dimensionen ist eine Voraussetzung dafür, dass das erworbene Wissen später auch in die Berufspraxis transferiert wird und war daher Ziel beider Lerngelegenheiten (vgl. Kap. 3).

II Empirischer Teil

6 Die Entwicklung und Pilotierung einer fallbasierten Lerngelegenheit (Studie 1)

6.1 Hintergrund und Überblick

Anhand der Ausführungen in Kapitel 4 wird deutlich, dass die Entwicklung von Interventionen zur Förderung diagnostischer Kompetenzen ein aktuelles Desiderat im deutschsprachigen Forschungsraum ist (vgl. Südkamp & Praetorius, 2017; Schrader, 2017). Ansätze zur Förderung diagnostischer Kompetenz sind zum einen rar und zum anderen häufig auf die Verbesserung der Urteilsgenauigkeit zugeschnitten (vgl. Artelt & Gräsel, 2009; Hetmanek & Van Gog, 2017; Schrader, 2009, 2013). Diese Methode setzt die Erfassung von realen Leistungsdaten voraus und ist daher eher für praktisch tätige Lehrkräfte gut geeignet, da diese im Rahmen ihres alltäglichen Handelns diagnostische Daten erheben und nutzen. In der ersten Phase der Lehrkräftebildung können angehende Lehrkräfte jedoch in der Regel nicht auf jene Realdaten zurückgreifen. Lehramtsstudierende benötigen daher alternative Lerngelegenheiten, die den Erwerb von Wissen und eine erste Anwendung dieses Wissens ermöglichen. Ein primäres Ziel der vorliegenden Arbeit war es daher, eine *Lerngelegenheit zu entwickeln und zu evaluieren*, die ohne diagnostische Realdaten auskommt.

In diesem Kapitel wird das selbstentwickelte *diagnostische Fallinventar* – eine fallbasierte Lerngelegenheit für Lehramtsstudierende – vorgestellt. Das Fallinventar wurde im Rahmen des Projektes K2teach³ (K2teach, 2019) für das Modul „Pädagogische Diagnostik“ an der Freien Universität Berlin entwickelt. Der Fokus dieser fallbasierten Lerngelegenheit ist ein besonderer Diagnoseanlass: die Feststellung von Lernbesonderheiten. Im folgenden Abschnitt 6.2 wird zunächst erläutert, was unter Lernbesonderheiten zu verstehen ist, welche einzelnen Phänomene diesem Begriff zugeordnet werden, und wie das diagnostische Vorgehen hier gestaltet sein muss. Darauf basierend wird erläutert, welche Wissensgrundlagen hierfür erforderlich sind und welche Ansätze es zur Förderung dieser Fähigkeiten bereits gibt (Abschn. 6.3). Im Zentrum des Kapitels stehen dann Ausführungen zur Konzeption (Abschn. 6.4) sowie zur Pilotierung (Abschn. 6.5) der Lerngelegenheit. Abschließend werden zentrale Überarbeitungsschritte des Fallinventars beschrieben und die finale Version vorgestellt, die in zwei weiteren Studien dieser Arbeit zum Einsatz kam (Studie 2: Kap. 7; Studie 3: Kap. 8).

³ K2teach ist ein Projekt, das im Rahmen der „Qualitätsoffensive Lehrerbildung“ im Zeitraum 2015-2018 an der FUB angesiedelt war. Projekt übergreifendes Ziel war es, Lehramtsstudierende bereits im Studium auf spätere Praxisanforderungen vorzubereiten und dementsprechend mehr Praxisanteile in das universitäre Curriculum zu integrieren. Im Teilprojekt 1 wurde das spezifische Ziel verfolgt, die Studierenden auf eine datenbasierte Unterrichtsentwicklung vorzubereiten. Um den Umgang mit diagnostischen Daten einzuüben, wurde das diagnostische Fallinventar entwickelt.

6.2 Diagnostik von Lernbesonderheiten als besondere diagnostische Aufgabe

Die Diagnostik von Lernbesonderheiten ist ein besonderer Diagnoseanlass, da er durch einen hohen Grad an Verbindlichkeit ausgezeichnet ist (vgl. Karst et al., 2017). Diese Verbindlichkeit entsteht dadurch, dass die Diagnostik von Lernbesonderheiten in der Regel mit weitreichenden Konsequenzen für die Diagnostikand*innen verbunden ist: Die Diagnose bildet die Grundlage für Entscheidungen, die Schüler*innen mit Lernbesonderheiten meist nachhaltig beeinflussen. Schüler*innen mit Lernbesonderheiten können in der Regelschulpraxis nicht automatisch ihren Fähigkeiten entsprechend lernen – vielmehr benötigen sie oftmals individuelle Unterstützung, um ihr kognitives Leistungspotential auszuschöpfen und damit ihre psychische Gesundheit aufrecht zu erhalten (für Lernstörungen: Berner, 2018; für Hochbegabung: Ziegler, 2017). Dass diese Schüler*innen korrekt und möglichst frühzeitig diagnostiziert werden, ist die Voraussetzung dafür, dass sie eine individuelle Förderung erhalten, die zu ihrem Bedarf sowie ihrem Entwicklungspotential passt. Die einzelnen Phänomene, die zusammen als „Lernbesonderheiten“ bezeichnet werden, werden nachfolgend beschrieben.

6.2.1 Definition und Vorkommen von Lernbesonderheiten

Als Lernbesonderheiten möchte ich *lern- und leistungsbezogene Auffälligkeiten* von Schülerinnen und Schülern bezeichnen, die ihre Fähigkeiten oder Fähigkeitspotentiale im regulären Unterricht nicht ausreichend entwickeln oder zeigen können, und dadurch Leidensdruck verspüren oder in ihrer gesellschaftlichen Teilhabe beeinträchtigt sind. Diese Auffälligkeiten können sich entweder in den Schulleistungen selbst oder bezüglich leistungsrelevanter Merkmale zeigen (vgl. Hosenfeld & Schrader, 2006).

Lernbesonderheiten können einerseits durch Verzögerungen von Kompetenzentwicklungen in einzelnen, umschriebenen schulischen Leistungsdomänen auftreten, wie etwa bei umschriebenen Entwicklungsstörungen schulischer Fertigkeiten (kurz „Lernstörungen“). Andererseits zähle ich auch Besonderheiten dazu, bei denen die kognitiven Lernvoraussetzungen im Extrembereich liegen. In diese Gruppe fallen Schüler*innen, deren kognitive Fähigkeiten entweder im weit unterdurchschnittlichen (allgemeine Lernschwäche oder „Lernbehinderung“) oder im weit überdurchschnittlichen Bereich („Hochbegabung“) liegen.

Umschriebene Entwicklungsstörungen schulischer Fertigkeiten werden in der ICD-10 (DIMDI, 2018, F81.0-F81.3) gelistet und umfassen die Phänomene der Lese-Rechtschreibstörung (F81.0), isolierten Rechtschreibstörung (F81.1), Rechenstörung (F81.2) oder der kombinierten Störung schulischer Fertigkeiten (F81.3). Charakteristisch für diese Lern- und Leistungsstörungen sind massive Verzögerungen in klar umschriebenen Fähigkeitsbereichen, wobei die allgemeinen kognitiven Fähigkeiten oder Leistungen in anderen Fächern nicht bzw. weniger beeinträchtigt sind. Ein zentrales Diagnosekriterium ist die Diskrepanz der schulischen Fähigkeiten zu den Leistungen, die aufgrund des Alters, der Klassenstufe oder der kognitiven Leistungsfähigkeit zu erwarten wären (DGKJP, 2015). Massive Auffälligkeiten im kognitiven Bereich sind Ausschlusskriterien für die Diagnose von Lernstörungen, da die damit verbundenen Lern- und Leistungsverzögerungen dann nicht umschrieben, sondern umfassend wären

(„Lernbehinderung“). In der Regel werden für die eindeutige Diagnostik von Lernstörungen daher durchschnittliche kognitive Fähigkeiten als ein Diagnosekriterium vorausgesetzt, auch wenn die Diskrepanz zur Intelligenz nach aktuellen medizinischen Leitlinien bei der Diagnostik der Lese-Rechtschreibschwäche oder –störung nur noch eines von drei möglichen Kriterien (vgl. DGKJP, 2015), und bei der Diagnostik der Rechenschwäche oder –störung gar kein relevantes Kriterium mehr darstellt (vgl. DGKJP, 2018). Da die diagnostischen Leitlinien und die damit verbundenen Diagnosekriterien sich in der Praxis jedoch erst langsam durchsetzen, wurde bei der Gestaltung der Fallvignetten in der vorliegenden Arbeit auf die klassischen Definitionen der ICD-10 zurückgegriffen. Diese werden in Tabelle 6.1 spezifiziert.

Tabelle 6.1 Definitionen und Prävalenzen von einzelnen Lernstörungen

Phänomen	Definition	Prävalenz
Lese-Rechtschreibstörung	„Hauptmerkmal ist eine bedeutsame Beeinträchtigung in der Entwicklung der Lesefertigkeiten, die nicht allein durch das Entwicklungsalter ¹ , Visusprobleme oder unangemessene Beschulung erklärbar ist“, zusätzlich können die Rechtschreibleistungen beeinträchtigt sein ²	ca. 3-8 % ³
Rechenstörung	„umschriebene Beeinträchtigung von Rechenfertigkeiten, die nicht allein durch eine Intelligenzminderung oder eine eindeutig unangemessene Beschulung erklärbar ist. Das Defizit betrifft die Beherrschung grundlegender Rechenfertigkeiten wie Addition, Subtraktion, Multiplikation und Division, weniger die höheren mathematischen Fertigkeiten“ ⁴	ca. 2-8 % ⁵
Kombinierte Störung schulischer Fertigkeiten	„schlecht definierte Restkategorie für Störungen mit deutlicher Beeinträchtigung der Rechen-, der Lese- und der Rechtschreibfähigkeiten“, wird vergeben, wenn Kriterien für eine Lese-Rechtschreibstörung oder isolierte Rechtschreibstörung und die Kriterien einer Rechenstörung erfüllt sind ⁶	ca. 2-5 % ⁵

Anmerkungen: Alle Definitionen sind der ICD-10 (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme), die angegebenen Prävalenzen sind den Leitlinien der DGJKP (Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie, Psychosomatik und Psychotherapie) entnommen. Verweise: ¹Das Entwicklungsalter umschreibt die allgemeinen kognitiven Fähigkeiten. ²DIMDI (2017) F81.0 Lese- und Rechtschreibstörung. ³DGJKP (2015) Leitlinie Lese-Rechtschreibstörung. ⁴DIMDI (2017) F81.2 Rechenstörung. ⁵DGJKP (2018) Leitlinie Rechenstörung. ⁶DIMDI (2017) F.81.3 kombinierte Störungen schulischer Fertigkeiten.

Anders als bei umschriebenen Lernstörungen zeichnen sich *Lernbesonderheiten im Bereich kognitiver Fähigkeiten* i. d. R. dadurch aus, dass die Fähigkeitsprofile der Schüler*innen konsistent im oberen oder unteren Fähigkeitsbereich liegen. Gemeint sind hier also keine umschriebenen, sondern umfassende Besonderheiten hinsichtlich der schulischen Fähigkeiten. Diese stehen in Zusammenhang mit weit über- oder unterdurchschnittlichen kognitiven Fähigkeiten. Hierzu zählt erstens eine „Lernbehinderung“, die sich allgemein durch geringe kognitive Fähigkeiten sowie entsprechend niedrige schulische Leistungen ausdrückt (Grünke & Grosche, 2014; für Berlin: SenBJW, 2012). Die schulischen Fähigkeiten entsprechen also den reduzierten kognitiven Fähigkeiten, die diese allgemeine Lernschwäche auszeichnen. Auch bei einer Hochbegabung, dem zweiten Phänomen, liegen zumeist relativ homogene Fähigkeitsprofile vor: Weit überdurchschnittliche kognitive Fähigkeiten gehen hier in der Regel auch mit hohen oder sehr hohen

schulischen Leistungen einher (Rost & Buch, 2010). Eine Ausnahme bildet das dritte Phänomen, das eine besondere Form der Hochbegabung ist. Hochbegabte Minderleister*innen zeigen zwar überdurchschnittlich hohe kognitive Fähigkeiten, aber ihre schulischen Leistungen bleiben dahinter zurück. Hier klaffen also kognitive Lernvoraussetzungen und die gemessenen schulischen Fähigkeiten weit auseinander, sodass diese Lernbesonderheit eine besondere Form von Lernstörungen bezeichnet werden können (Vohrmann, 2018). In Tabelle 6.2 sind sowohl gängige Definitionen als auch Prävalenzen der drei genannten Phänomene zusammengestellt. Die Angaben zu den Prävalenzen schwanken hier, wie in der vorigen Tabelle, da die Schätzungen der Prävalenz durch die jeweilige Untersuchungsstichprobe, die angesetzten Diagnosekriterien sowie die verwendeten Testverfahren mehr oder weniger stark variieren (Hasselhorn & Schuchardt, 2006).

Tabelle 6.2 Definition und Prävalenzen von Lernbesonderheiten im kognitiven Bereich

Phänomen	Definition	Prävalenz
Allgemeine Lernschwäche / „Lernbehinderung“	„schwerwiegende, anhaltende und umfängliche Schwierigkeiten bei der Bewältigung von intellektuellen Leistungsanforderungen, [...] eine besonders ausgeprägte Form einer Minderleistung bei der absichtsvollen und aktiven Verarbeitung sowie der Abspeicherung von Wissen. Diese Einschränkungen zeigen sich in erster Linie beim Erwerb kognitiv-verbaler und abstrakter Inhalte“ ¹	ca. 5 %
Hochbegabung	„ein wenig präzises Konzept [...], die Unschärfe des Hochbegabungsbegriffs liegt in der Unschärfe und Mehrdeutigkeit des Begriffs ‚Begabung‘ [...], i. d. R. meint Hochbegabung die besondere Ausprägung der Intelligenz im Vergleich zum Populationsdurchschnitt, die mindestens zwei Standardabweichungen über dem Mittelwert liegt (IQ > 129, PR > 97); Hochbegabung als „breit angelegte intellektuelle Potenz“ ²	ca. 2-3 %
Hochbegabung mit Minderleistung („Underachiever“)	„relative Diskrepanz zwischen Potenzial und Leistung“, d. h. beträchtliche Lücke zwischen sehr hohem Leistungspotenzial und tatsächlichen Leistungen nach individueller Bezugsnorm und damit eine besondere Form von Lernschwierigkeiten ³	ca. 10-15 % der Hochbegabten

Anmerkungen: Die Lernbesonderheiten, die sich auf besondere kognitive Fähigkeiten beziehen, werden nicht im Rahmen der ICD-10 definiert. Die angegebenen Prävalenzen sind den zur Definition verwendeten Quellen entnommen: ¹Grünke & Grosche (2014), S. 76; ²Rost & Buch (2010), S. 257f.; ³Vohrmann (2018), S. 37 f.

Anhand der Definitionen der verschiedenen Formen von Lernbesonderheiten wird deutlich, dass der Begriff einerseits Störungen im klinischen Sinne sowie pädagogische bzw. sonderpädagogische Auffälligkeiten beschreibt, für die unterschiedliche Diagnosekriterien, Förderansätze sowie Zugangsbedingungen für Unterstützungsmaßnahmen angelegt werden. Die Diagnostik von Lernbesonderheiten und die Ableitung und Umsetzung von Fördermaßnahmen sind damit eine besonders herausfordernde Aufgabe – die allerdings in der schulischen Praxis gar nicht selten ist. Wie die Angaben zu den Prävalenzen zeigen, kommen Lehrkräfte in der Schulpraxis sehr häufig mit den ausgewählten Lernbesonderheiten in Berührung. Bei einer relativen Häufigkeit von etwa 5 % im Fall der Lese-Rechtschreibstörung ist davon auszugehen, dass in jeder Schulklasse schätzungsweise ein bis drei Kinder mit einer solchen Lernbesonderheit beschult werden, wobei in Ballungsräumen von noch höheren

Fallzahlen auszugehen ist (Heine & Jacobs, 2011; May, 2012). Der weiten Definition von Inklusion folgend sind diese Schüler*innen – neben Schüler*innen mit Behinderungen im engeren Sinne – Teil einer *typisch heterogenen Schüler*innenschaft*, die allesamt innerhalb des unterrichtlichen Geschehens Berücksichtigung finden müssen (Werning, 2017).

Laut KMK-Beschluss (2004/2014) gehören die Feststellung von Lernschwierigkeiten oder – verzögerungen einerseits sowie die Diagnostik von besonderen Begabungen andererseits zum Kanon der üblichen *diagnostischen Aufgaben von Lehrkräften*. Die Entscheidungen, die auf diesen Diagnosen fußen, erfordern durch ihre Tragweite dabei stets ein explizites diagnostisches Vorgehen (Hesse & Latzko, 2009, 2017), das sich durch ein zielgerichtetes, systematisches und methodisch kontrolliertes diagnostisches Verhalten auszeichnet (formelle Diagnose, Schrader, 2010). Jene formellen Diagnosen können durch einen langsamen, analytischen, kognitiv aufwändigen diagnostischen Prozess sichergestellt werden (Typ2-Diagnose, Hetmanek & Van Gog, 2017). Ein derart zeit- und ressourcenintensiver Prozess findet daher in der Regel nicht während des laufenden Unterrichtsgeschehens statt, sondern erfordert eine außerunterrichtliche Diagnostik (vgl. Karst et al., 2017). Die Diagnostik und Förderung von Lernbesonderheiten sind damit nicht nur besonders verantwortungsvolle, sondern auch sehr anspruchsvolle Aufgaben, die Lehrkräften aufgetragen ist.

6.2.2 Prinzipien der Diagnostik und Förderung bei Lernbesonderheiten

Für das Vorliegen einer Lernstörung oder –behinderung sind in der Regel *Lernschwierigkeiten ein erster Indikator*. „Von Lernschwierigkeiten spricht man im Allgemeinen, wenn die Leistungen eines Schülers unterhalb der tolerierbaren Abweichung von verbindlichen institutionellen, sozialen und individuellen Bezugsnormen (Standards, Anforderungen, Erwartungen) liegen oder wenn das Erreichen (bzw. Verfehlen) von Standards mit Belastungen verbunden ist, die zu unerwünschten Nebenwirkungen im Verhalten, Erleben oder der Persönlichkeitsentwicklung des Lernenden führen.“ (Weinert & Zielinski, 1977, zitiert nach Hesse & Latzko, 2017, S. 291). Schüler*innen mit Lernschwierigkeiten weichen mit ihren Leistungen demnach entweder von institutionellen Bezugsnormen, d. h. von Schulform abhängigen Lehrplanzielen, oder von sozialen Normen, d. h. von der Durchschnittsleistung der jeweiligen Klasse, ab (Hesse & Latzko, 2017). Diese Nutzung des klasseninternen Bezugsrahmens als Bezugsnorm ist allerdings nicht objektiv und daher begrenzt aussagekräftig, denn in einer leistungsschwachen Klasse würde ein*e lernschwache*r Schüler*in vermutlich gar nicht auffallen, und in einer leistungsstarken Klasse hingegen würden u. U. auch Schüler*innen mit ausreichenden Leistungen weit vom Klassenmittel entfernt liegen (vgl. Hesse & Latzko, 2017; Ingenkamp & Lissmann, 2008).

Die objektive Abklärung des Vorliegens von Lernstörungen oder auch Hochbegabungen erfordert daher immer den Vergleich mit einer repräsentativen Normstichprobe, der nur mithilfe *standardisierter Leistungs- oder Fähigkeitstests* durchgeführt werden kann. Die Durchführung von Schulleistungstests, die die umschriebenen Fähigkeiten im Bereich des Lesens, Schreibens oder Rechnens umfassen, obliegt in der schulischen Praxis Berlins beispielsweise den jeweiligen Fachlehrkräften, klassenübergreifend arbeitenden

Sonderpädagog*innen oder einzelnen explizit benannten und fortgebildeten Lehrkräften, die diese Aufgabe für die ganze Schule wahrnehmen (GsVO, 2005/2019). Die Diagnostik der Rechtschreibfähigkeiten mithilfe der Hamburger Schreibprobe (HSP; May, 2012) ist in Berliner Grundschulen beispielsweise inzwischen bereits flächendeckend zu beobachten und trägt zur Absicherung informell beobachteter Lernschwierigkeiten im schriftsprachlichen Bereich bei. Die Diagnostik „gravierender und langanhaltende Schwierigkeiten“ im Bereich des Lesens, Schreibens und Rechnens gehören damit klar in den diagnostischen Aufgabenbereich der Schule (GsVO, 2005/2019).

Die Diagnostik von Lernstörungen im klinischen Sinne (vgl. ICD-10: DIMDI, 2017) erfordert allerdings i. d. R. auch eine *Überprüfung des kognitiven Potentials*. Sollen eine Lese- oder Rechtschreibstörung diagnostiziert werden, können neben Diskrepanzen der spezifischen Testleistung zur Vergleichsgruppe auch beträchtliche Unterschiede zwischen einer umschriebenen Fachleistung und den allgemeinen kognitiven Fähigkeiten bestehen. In beiden Fällen ist die Diagnose einer Lese- oder Rechtschreibstörung angemessen (DGKJP, 2015). Bei der Diagnostik der Rechenstörung spielt die Diskrepanz zu den kognitiven Fähigkeiten zwar keine zentrale Rolle für die Vergabe der Diagnose, dennoch ist die Diagnostik des allgemeinen kognitiven Potentials einschließlich der Arbeitsgedächtnis- und Exekutivfunktionen hier empfohlen (vgl. DGKJP, 2018). Die Abklärung der kognitiven Fähigkeiten dient einerseits der diagnostischen Abgrenzung umschriebener Lernstörungen von umfassenden Beeinträchtigungen wie einer allgemeinen Lernschwäche oder „Lernbehinderung“. Andererseits werden die zur Verfügung gestellten Informationen genutzt, um die erforderliche Breite von Fördermaßnahmen festzustellen. Die Diagnostik von kognitiven Fähigkeiten fällt dabei nicht in den Aufgabenbereich von Fachlehrkräften, sondern kommt Sonderpädagog*innen sowie (Schul-)Psycholog*innen zu – die Einleitung diagnostischer Schritte sowie die Planung und Umsetzung von Fördermaßnahmen obliegt hingegen in erster Linie den Fachlehrkräften (Zuständigkeiten und Prozeduren für Berlin: siehe SenBJF, 2019).

Tabelle 6.3 Kategorisierung von Lernschwierigkeiten hinsichtlich zeitlicher und bereichsbezogener Stabilität

	Bereichsspezifische Schwächen (partiell)	Bereichsübergreifende Beeinträchtigung (generell)
Vorübergehende Beeinträchtigung	<ul style="list-style-type: none"> ▪ Lernrückstände in einzelnen Fächern ▪ Lese-Rechtschreib-/ Rechenschwierigkeiten 	<ul style="list-style-type: none"> ▪ Schulschwierigkeiten z. B. durch psychische Erkrankungen / traumatische Ereignisse
Dauerhafte Beeinträchtigung	<ul style="list-style-type: none"> ▪ Lese-Rechtschreibschwäche /-störung (Legasthenie) ▪ Rechenschwäche/-störung (Dyskalkulie) 	<ul style="list-style-type: none"> ▪ Lernbehinderung¹

Anmerkungen: Darstellung in Anlehnung an Leitfaden zur Feststellung sonderpädagogischen Förderbedarfs an Berliner Schulen (SenBJW, 2012, S. 45).

¹Für den Begriff der „Lernbehinderung“ werden auch die Begriffe „geistige Behinderung“ bzw. der „Förderschwerpunkt Lernen“ teilweise synonym verwendet.

Bei der *Ableitung und Umsetzung von pädagogischen Maßnahmen*, die in der Regel Aufgabe des gesamten pädagogischen Teams einer Schule sind, sind einige Grundprinzipien zu beachten, die hier kurz skizziert werden sollen. Erstens ist die „Breite“ der Beeinträchtigungen zu berücksichtigen, da diese auch für die Breite der Maßnahmenplanung ausschlaggebend ist. Hierbei müssen zwei Dimensionen betrachtet werden (vgl. Tab. 6.3): Zum einen muss überprüft werden, inwiefern die beobachteten Auffälligkeiten zeitlich überdauernd sind. Vorübergehende Rückstände und Schwierigkeiten erfordern in der Regel weniger intensive und langwierige Fördermaßnahmen, wie etwa vorübergehend vermehrtes Üben, welche bei dauerhaften Beeinträchtigungen nicht ausreichen. Und zweitens muss analysiert werden, ob nur einzelne schulische Leistungsbereiche betroffen sind oder ob bereichsübergreifende Beeinträchtigungen vorliegen (SenBJW, 2012). In diesem Zusammenhang gilt zweitens der Grundsatz der symptomorientierten Förderung, demzufolge stets diejenigen Fähigkeitsbereiche gefördert werden sollten, in denen die Beeinträchtigung vorliegt (für Entwicklungsstörungen schulischer Fertigkeiten vgl. Heine et al., 2012). Im Falle einer umschriebenen Leseschwäche etwa sind daher insbesondere die Lesefähigkeiten zu fördern. Im Falle einer bereichsübergreifenden Beeinträchtigung, wie etwa einer Lernbehinderung, müssten diesem Prinzip folgend die schwachen schulischen Fähigkeiten in verschiedenen Bereichen sowie die allgemeinen kognitiven Fähigkeiten gefördert werden. In Berlin können Lehrkräfte hierfür auf eine Handreichung mit konkreten Fördermaßnahmen zurückgreifen (SenBJF, 2018), in der allgemeingültige Vorschläge zur Unterstützung von Schüler*innen mit sonderpädagogischem Förderbedarf zusammengestellt sind. Diese Vorschläge erfüllen ein drittes Prinzip der Förderung: Die vorgestellten Fördermaßnahmen sind entweder wissenschaftlich überprüft oder basieren auf umfangreicher praktischer Evidenz von Expert*innen. Für die Förderung hochbegabter Schüler*innen stand den Lehrkräften in Berlin zum Zeitpunkt der Entwicklung der Lerngelegenheit bedauerlicherweise bislang kein derartiges Manual zur Verfügung.

6.3 Fähigkeiten zur Diagnostik von Lernbesonderheiten fördern

Die vorausgehenden Ausführungen zeigen, dass die Diagnostik von Lernbesonderheiten eine wichtige Aufgabe ist, die spezifische Anforderungen an die Diagnostiker*innen stellt. Anders als bei Diagnoseanlässen mit weniger weitreichenden Konsequenzen ist hier die Korrektheit der Diagnose besonders wichtig, da sie die Grundlage dafür ist, dass entsprechende Anschlusshandlungen abgeleitet werden. Derartige Diagnosen sind mit hoher Sorgfalt zu erstellen und erfordern daher einen formellen, analytischen diagnostischen Prozess (Schrader, 2010; Hesse & Latzko, 2017; Hetmanek & Van Gog, 2017). Um diesen Prozess kompetent durchführen zu können, sind zwei Voraussetzungen erforderlich (Herppich et al., 2017; Hascher, 2008): Zum einen bedarf es kognitiver Voraussetzungen, d. h. einer fundierten Wissensbasis, und zum anderen sind praktische Erfahrungen oder Übungsmöglichkeiten erforderlich (vgl. auch Helmke, 2009; Hesse & Latzko, 2017). In der universitären Ausbildung sollten beide Aspekte berücksichtigt werden, um angehende Lehrkräfte auf die anspruchsvolle Aufgabe der Diagnostik von Lernbesonderheiten vorzubereiten. Die zum Aufbau von diagnostischer Expertise nötige Wissensbasis sollte neben deklarativen auch prozedurale Wissenskomponenten beinhalten (vgl. Hascher, 2008, S. 77). Im Rahmen aktueller Modellierungen diagnostischer Kompetenz werden beide Wissensarten

den Voraussetzungen zugerechnet, die (angehende) Lehrkräfte in die Lage versetzen, situationsspezifische Fähigkeiten zu entwickeln, die dabei helfen, in konkreten diagnostischen Situationen diagnostisch kompetent zu handeln (Herppich et al., 2017, 2018; vgl. auch Blömeke et al., 2015).

6.3.1 Förderung von Grundlagenwissen zur Diagnostik von Lernbesonderheiten

Das *deklarative Wissen*, das für die Diagnostik von Lernbesonderheiten ausschlaggebend ist, sollte insbesondere zur Durchführung eines regelgeleiteten diagnostischen Prozesses befähigen (vgl. Jäger, 2006). Das hierfür grundlegende Fakten- und Konzeptwissen sollte sich dementsprechend auf folgende Bereiche methodischen und pädagogisch-psychologischen Wissens beziehen, die von Jäger und anderen (vgl. Kap. 3.1.5 Inhaltsbereiche des Wissens) als zentrale Wissensbereiche im Zusammenhang mit diagnostischen Kompetenzen genannt werden:

- *Bedingungswissen*: Wissen, wie sich Lernbesonderheiten in Bezug auf Schulleistungen zeigen und wie sie zustande kommen können, Wissen über Diagnosekriterien und differentialdiagnostische Aspekte bei der Unterscheidung von Lernbesonderheiten;
- *Methodisch-technologisches Wissen*: Wissen über geeignete Erhebungs- und Auswertungsmethoden zur Feststellung von Lernbesonderheiten, Wissen über die Auswahl und Anwendung von spezifischen Verfahren;
- *Vergleichswissen*: Wissen über die Interpretation von Ergebnissen aus standardisierten Schulleistungstests, Wissen über die Interpretation von Standardnormwerten;
- *Änderungswissen*: Wissen über Strategien zur Veränderung der auffälligen Leistungsbereiche und über passende Fördermöglichkeiten für einzelne Phänomene.

Dieses Grundlagenwissen zur Diagnostik von Lernbesonderheiten erwerben Lehramtsstudierende an der Freien Universität Berlin regulär im ersten Mastersemester im Rahmen des *Moduls „Pädagogische Diagnostik“*. Dieses Modul setzt sich aus zwei Veranstaltungen zusammen: Eine Vorlesung dient der Vermittlung grundlegenden Faktenwissens, das neben testtheoretischen und statistischen Grundlagen auch Kenntnisse über standardisierte Leistungstests sowie die Diagnosekriterien und Förderansätzen für Lernstörungen und kognitive Lernvoraussetzungen beinhaltet. Im begleitenden Seminar werden die Inhalte aus der Vorlesung vertieft und durch ausgewählte Übungen auch erstmals angewendet. Um Fähigkeiten im Umgang mit standardisierten Schulleistungstests zu schulen, erhalten die Studierenden bspw. die Gelegenheit, einen standardisierten Leseverständnistest selbst durchzuführen, auswerten und anschließend ihre Erfahrungen zu diskutieren. Um den Studierenden ein konzeptuelles Verständnis im Bereich der Diagnostik von Lernbesonderheiten zu vermitteln, wird eine klassische aufgabenbasierte Übung angeboten. Die Studierenden erhalten hier Aufgaben verschiedener Anforderungsniveaus, die insbesondere die Interpretation von in Leistungstests gängigen Standardnormwerten, die Definitionen und Diagnosekriterien von Lernstörungen und kognitiven Lernvoraussetzungen sowie allgemeine Förderansätze thematisieren (ausführliche Darstellung s. Kap. 8).

Dass beim Aufbau deklarativen Wissens auch ein tiefergehendes Verständnis von Lernbesonderheiten und den jeweiligen Diagnosekriterien erreicht wird, ist wichtig, damit dieses Wissen überhaupt auf diagnostische Probleme angewendet werden kann. Die *Vermittlung konzeptuellen Wissens* bildet insbesondere in der Phase des beginnenden Kompetenzerwerbs die Grundlage dafür, dass prozedurales Wissen gefördert werden kann (Herppich et al., 2017a; Glogger-Frey & Renkl, 2017; Stern et al., 2016). Im Kontext der Diagnostik von Lernbesonderheiten bezieht sich prozedurales Wissen auf die regelgerechte Durchführung eines kriteriengeleiteten, methodisch kontrollierten diagnostischen Prozesses (Jäger, 2006; Schrader, 2010). Dieses prozedurale Wissen entsteht durch Anwendung des deklarativen Grundlagenwissens und befähigt Lehrkräfte, diagnostisch kompetent zu handeln.

6.3.2 Förderung prozeduralen Wissens zur Diagnostik von Lernbesonderheiten

Die o. g. Übungsangebote im Modul bieten zwar die Gelegenheit, das erworbene Grundlagenwissen in Form erster Anwendungen zu vertiefen, diese Anwendungen sind allerdings im didaktischen Sinne zu verstehen (vgl. z. B. Glessmer & Lüth, 2016). Lerngelegenheiten, die eine „echte“ praktische Anwendung dieses Grundlagenwissens ermöglichen, waren also ein Desiderat für das Modul. Diese „echten“ Anwendungen sind essenziell, um prozedurales Wissen aufzubauen, das für die Entwicklung diagnostischer Expertise unerlässlich ist (Hascher, 2008). Diagnostische Expertise entsteht vor allem durch „die fortlaufende Konfrontation mit den typischen Aufgabenstellungen eines Arbeitsfeldes“ (Schrader, 2017, S. 253), d. h. durch Praxiserfahrungen.

Vor dem Hintergrund, dass während des Lehramtsstudiums wenige praktische Erfahrungsmöglichkeiten bestehen, können diese Konfrontationen im Modul „Pädagogische Diagnostik“ allerdings nicht in der realen Praxis erfolgen. Abhilfe können hier Lerngelegenheiten schaffen, die eine *Annäherung an die realen Praxisbedingungen* ermöglichen. Die Praxisannäherung kann dabei als ein Stufenprozess verstanden werden, bei dem die Annäherung schrittweise erfolgt (Approximations-of-Practice Framework: Grossman et al., 2009). Die einzelnen Schritte führen Lehramtsstudierende stufenweise an die Komplexität und Authentizität der realen Praxisprobleme heran und bereiten die Studierenden auf diese Weise optimal auf die Praxisanforderungen vor. Auf der ersten Stufe in dieser Praxisannäherung ist die Arbeit mit schriftlichen Fallvignetten vorgesehen (vgl. Abb. 6.1).

Mithilfe von *schriftlichen Fallvignetten* können realitätsnahe Szenarien geschaffen werden, in denen diagnostische Handlungen darstellbar sind (z. B. Kaiser et al., 2015). Sie können damit einen entscheidenden Beitrag zur Entwicklung diagnostischer Expertise bereits vor der eigentlichen Berufspraxis leisten, indem sie auch ohne selbsterhobene Realdaten erlauben, diagnostische Fähigkeiten von Lehramtsstudierenden zu trainieren (z. B. Glogger-Frey & Renkl, 2017; Herppich et al., 2017a; vgl. Abschn. 4.3). Selbst in der dritten Phase der Lehrkräftebildung können Fallvignetten noch gewinnbringend eingesetzt werden, um diagnostische Expertise zu fördern (vgl. Hesse & Latzko, 2017; Klug, 2017; Abschn. 4.2). Fallbasierte Lerngelegenheiten können ganz allgemein in verschiedenen Phasen

der Praxisannäherung eingesetzt werden (s. a. Abb. 6.1), d. h. auch in späteren, in denen mehr Authentizität und Komplexität hergestellt werden.

Zur optimalen Vorbereitung diagnostischer Expertise sollte die Arbeit mit Fallvignetten aus meiner Sicht dabei unbedingt die regelgeleitete *Durchführung des diagnostischen Prozesses* thematisieren. In Kapitel 2 dieser Arbeit wurde herausgestellt, dass Aspekte des diagnostischen Prozesses in Zusammenhang mit korrekten Diagnosen stehen, die bei der Diagnostik von Lernbesonderheiten unbedingt anzustreben sind. Eine professionelle Informationssammlung und –verarbeitung, die im diagnostischen Prozess erfolgen, sind Voraussetzung für die Güte der diagnostischen Daten und Urteile (Behrmann & Van Ophuysen, 2017). Da Noviz*innen ggü. Expert*innen bei der Informationsverarbeitung eine deutlich höhere Anfälligkeit für Verzerrungen zeigen, sind ihre Urteile besonders fehleranfällig (vgl. Abschnitt 2.2.1.2; Klug et al., 2016; Krolak-Schwerdt et al., 2009, 2013). Die Verbesserung prozessbezogener Fähigkeiten sollte also zur Vermeidung von Fehldiagnosen beitragen. Dementsprechend sind Lerngelegenheiten, in denen diagnostische Prozeduren eingeübt werden können, im Lehramtsstudium erfolgversprechend (Hetmanek & Van Gog, 2017).

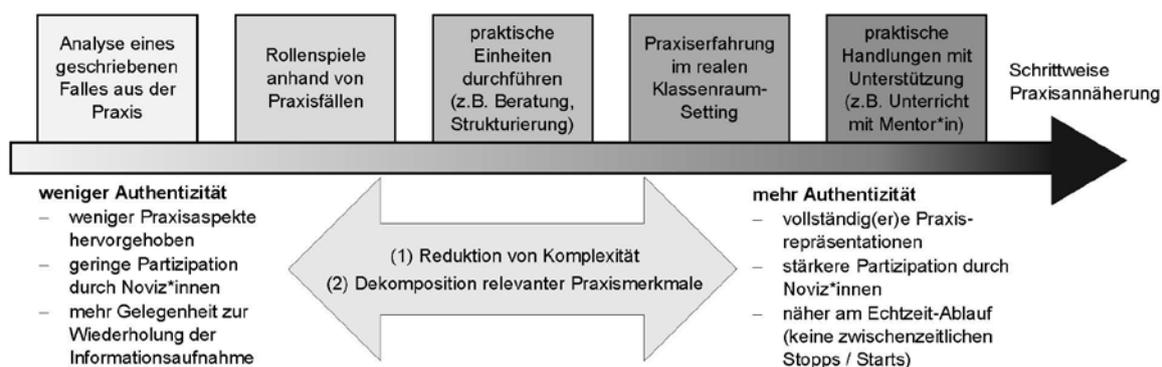


Abbildung 6.1 Rahmenmodell schrittweiser Praxisannäherung in der Lehrkräftebildung (Grossman et al., 2009)

6.4 Konzeption und Gestaltung eines diagnostischen Fallinventars

Die Entwicklung fallbasierter Lerngelegenheiten für die erste Phase der Lehrkräftebildung sowie dazugehörige empirische Wirksamkeitsnachweise sind ein Desiderat in der deutschsprachigen Forschung. Im Rahmen der Qualitätsoffensive Lehrerbildung werden daher aktuell fallbasierte Lerngelegenheiten für verschiedene Anwendungen entwickelt (K2teach, 2018; Wedel & Pfetsch, 2017). Ziel der vorliegenden Arbeit war es, eine fallbasierte Lerngelegenheit für die Feststellung von Lernbesonderheiten zu entwickeln, die im Masterstudium einsetzbar ist. Im deutschsprachigen Raum liegt meines Wissens keine Lerngelegenheit vor, die angehende Lehrkräfte bereits im Studium auf die Diagnostik von Lernbesonderheiten vorbereiten will. Die fallbasierte Lerngelegenheit, die ich im Rahmen des K2teach-Projektes entwickelt habe, soll insbesondere *prozedurales diagnostisches Wissen fördern*, das ein wichtiger

Baustein diagnostischer Expertise ist (Hascher, 2008; Helmke, 2009), aber m. E. in der universitären Bildung noch eine zu geringe Rolle spielt.

Mit der Entwicklung des sog. „diagnostischen Fallinventars“ werden drei Kernziele verfolgt, die mit der Entwicklung prozeduralen Wissens verknüpft sind: (1) Die Studierenden sollen im Rahmen dieser neu entwickelten fallbasierten Übung erstens die Gelegenheit erhalten, das im Modul „Pädagogische Diagnostik“ erworbene deklarative Wissens zur Diagnostik von Lernbesonderheiten auf realitätsnahe diagnostische Probleme anzuwenden. (2) Mithilfe realitätsnaher Fallvignetten von Schülerinnen und Schülern mit und ohne Lernbesonderheiten sollen die Studierenden zweitens das Erstellen von formellen Diagnosen sowie das Ableiten entsprechender Anschlusshandlungen einüben. (3) Die eigene Durchführung eines komplexitätsreduzierten diagnostischen Prozesses soll sie drittens anregen, über ihr eigenes diagnostisches Vorgehen zu reflektieren (vgl. Hesse & Latzko, 2017) und so für die damit verbundene Reichweite dieser Diagnosen und eigene Urteilsfehler sensibilisieren (vgl. Helmke, 2009). Die nachfolgenden Ausführungen stellen die zentralen konzeptionellen Überlegungen vor, die bei der Gestaltung des Fallinventars maßgeblich waren.

6.4.1 Komplexitätsreduzierte Fallvignetten

Vorbild des vorliegenden diagnostischen Fallinventars war eine fallbasierte Umgebung zur Untersuchung diagnostischer Urteilsprozesse – das sogenannte „Schülerinventar“ (Kaiser et al., 2015). Diese szenariobasierte Umgebung ist eine Weiterentwicklung des simulierten Klassenzimmers, das seit etwa zehn Jahren zur Simulation und Untersuchung diagnostischer Handlungen genutzt wird (z. B. Südkamp et al., 2008; vgl. Abschnitt 2.1.1 Urteilstgenauigkeit). Das Schülerinventar ermöglicht die Untersuchung von Urteilsprozessen und Einflüssen auf die Urteilsbildung, indem Schüler*innenmerkmale im Rahmen von Fallvignetten systematisch variiert werden. Durch die systematische Variation sind experimentelle Untersuchungen der Urteilstgenauigkeit möglich, ohne dass reale diagnostische Daten erhoben werden müssen. Beim Einsatz des Schülerinventars zeigte sich eindrücklich, dass angehende Lehrkräfte in der zweiten Phase ihrer Ausbildung (Referendariat) bei der Beurteilung der Mathematikleistung nicht nur durch relevante Merkmale – wie etwa mündliche und schriftliche Matheleistungen – beeinflusst waren, sondern auch durch irrelevante Merkmale wie die Deutschnote oder die allgemeine Intelligenz. Bei der Entwicklung der vorliegenden fallbasierten Lernumgebung wurde die Idee der kompakten Zusammenstellung verschiedener relevanter und irrelevanter Informationen in Form schematischer Fallvignetten über verschiedene Schüler*innen übernommen. Anders als dort sollten die Studierenden im Fallinventar jedoch keine Leistungsbeurteilungen vornehmen, sondern Lernbesonderheiten feststellen.

Die Entwicklung der Kurzvignetten⁴ erfolgte auf der *Basis realer Fallbeschreibungen*, die ich im Rahmen meiner lerntherapeutischen Arbeit vor dem Projekt erhalten hatte. Diese Falldarstellungen enthielten eine

⁴ Ursprünglich war der Einsatz längerer Fallvignetten angedacht, wie sie bspw. auch im Rahmen von Fortbildungen für Lehrkräfte eingesetzt werden (z. B. Hesse & Latzko, 2017). Im Wintersemester 2015/16 wurden probenhalber vier Fallbeschreibungen eingesetzt, die einen Umfang von insgesamt 5-7 Seiten hatten. Aufgabe der Studierenden war es,

Anamnese, Berichte von Eltern und Lehrkräften, Einschätzungen der Schüler*innen selbst, Arbeitsproben sowie Ergebnisse aus standardisierten Schulleistungs- sowie kognitiven Fähigkeitstests. Anhand dieser Informationen wurden schematische Fallvignetten erarbeitet, die eine Auswahl der ursprünglichen Informationen zu schulischen und kognitiven Fähigkeiten enthielten. Um sicherzustellen, dass die betreffenden Schüler*innen nicht wiedererkannt werden können, und weil nicht für alle Schüler*innen vollständige Informationen vorlagen, wurden die jeweiligen Informationen verändert oder vervollständigt.

Die realitätsnahen Fallvignetten unterscheiden sich von den realen Praxisfällen insbesondere dadurch, dass sie nur ausgewählte Aspekte der realen Fälle schildern. Diese Vereinfachung der Realität entspricht einer *Komplexitätsreduktion*, die für Noviz*innen besonders wichtig ist, wenn sie ihre ersten praxisvorbereitenden Erfahrungen machen: Praxisannäherung sollte in Stufen gedacht werden, bei denen die tatsächliche Komplexität von Praxissituationen allmählich erhöht wird (Grossman et al., 2009). Innerhalb des Modells der schrittweisen Praxisannäherung (vgl. Abb. 6.1) ist das diagnostische Fallinventar mit seinen schriftlichen Fallvignetten der ersten von fünf Stufen der Annäherung an die Berufspraxis einzuordnen, auf der die Komplexität am geringsten ist. Die Komplexitätsreduktion wurde hier durch zwei Aspekte erreicht: Erstens wurde die Auswahl an Informationen begrenzt, indem nur einzelne Testergebnisse sowie ergänzende Beschreibungen von Gesprächen und Beobachtungen vorgegeben wurden. Zweitens wird die Fallkomplexität reduziert, indem die Informationen sich ausschließlich auf schulische und schulrelevante Fähigkeiten bezogen und somit die Beachtung von weiteren Einflussvariablen auf die Schulleistung wie bspw. Motivation (für eine Übersicht über mögliche Einflussfaktoren auf Schüler*innenleistungen vgl. z. B. Angebot-Nutzungs-Modell: Kunter, 2011), außen vor bleiben kann. Diese Reduktion von Komplexität durch die Reduktion der Informationsmenge ermöglicht den Studierenden, die vorhandenen Informationen mit wenig Zeitdruck und daher gründlich zu bearbeiten, bei Bedarf mehrmals zu lesen und so Gefühle der Überforderung zu vermeiden (zum Kompetenzerleben von Lehramtsstudierenden in komplexitätsreduzierten Lerngelegenheiten s. z. B. Klempin, 2019).

6.4.2 Darstellung prototypischer Fähigkeitsprofile

Für die Falldarstellungen der Schüler*innen aus der Grund- und weiterführenden Schule wurden im Zuge der Reduktion der Informationsmenge *drei Fähigkeitsbereiche* ausgewählt, die für jede Diagnostik von Lernbesonderheiten relevant sind: Lesefähigkeiten, Rechenfähigkeiten und kognitive Fähigkeiten. Die Diagnostik dieser zwei Teilleistungen (Lesen, Rechnen) sowie die Durchführung von mind. einem Intelligenztest⁵ stellen einerseits die Mindestkriterien zur Diagnostik von Lernstörungen dar (vgl. auch

eine Diagnosevermutung abzugeben sowie auf Basis einer qualitativen Fehleranalyse Förderschwerpunkte zu benennen. In einer regulären Seminarsitzung von 90 Minuten gelang dem Großteil der Studierenden dabei nur knapp, ihre Falldarstellung gründlich zu bearbeiten und mit der*dem Sitznachbar*in auszutauschen. Auch war eine Nachbesprechung in derselben Sitzung nicht mehr möglich. Zahlreiche Rückfragen und Rückmeldungen über Schwierigkeiten beim Bearbeiten der Fälle zeigten zudem, dass diese Übungsform für die Studierenden noch zu komplex für den zeitlichen Rahmen war, der zur Verfügung stand. Vor diesem Hintergrund wurden schematische Kurzvignetten erstellt, die auch die Bearbeitung mehrerer Fälle in einer Sitzung erlaubten.

⁵ Wenn zwei verschiedene Intelligenztests erforderlich sind, wie bspw. bei der Feststellung einer Lernbehinderung, müssen diese in ihren Ergebnissen im Wesentlichen übereinstimmen (vgl. Rost & Buch, 2010; SenBJW, 2012).

Diagnosekriterien ICD-10). Diese drei Fähigkeitsbereiche sind außerdem wichtig zur diagnostischen Abgrenzung einzelner Phänomene. In den erstellten Fallvignetten können mithilfe dieser drei Fähigkeitsbereiche umschriebene bzw. umfassende Schwächen und Stärken der Schüler*innen dargestellt werden. Hierzu wurden *prototypische Fähigkeitsprofile* für die einzelnen Lernbesonderheiten erarbeitet, die die notwendigen Bedingungen für die Vergabe der jeweiligen Diagnosen berücksichtigen (vgl. Tab. 6.4).

Tabelle 6.4 Prototypische Fähigkeitsprofile der Lernbesonderheiten

	Lesefähigkeiten					Rechenfähigkeiten					Kognitive Fähigkeiten ⁵				
	--	-	○	+	++	--	-	○	+	++	--	-	○	+	++
Lese-Rechtschreibstörung	x	x						x	x	x		(x)	x	x	x
	unterdurchschnittlich bis weit unterdurchschnittlich					Mindestens durchschnittlich					i. d. R. mindestens durchschnittlich ¹				
Rechenstörung			x	x	x	x	(x)					(x)	x	x	x
	mindestens durchschnittlich					unterdurchschnittlich bis weit unterdurchschnittlich					i. d. R. mindestens durchschnittlich ²				
Kombinierte Störung schul. Fertigkeiten	x	x				x	x						x	x	x
	unterdurchschnittlich bis weit unterdurchschnittlich					unterdurchschnittlich bis weit unterdurchschnittlich					mindestens durchschnittlich ³				
Allg. Lernschwäche („Lernbehinderung“)	x	x				x	x				x	x			
	unterdurchschnittlich bis weit unterdurchschnittlich					unterdurchschnittlich bis weit unterdurchschnittlich					unterdurchschnittlich bis weit unterdurchschnittlich				
Hochbegabung				x	x				x	x					x
	überdurchschnittlich bis weit überdurchschnittlich					überdurchschnittlich bis weit überdurchschnittlich					weit überdurchschnittlich				
Hochbegabung mit Minderleistung			x	x				x	x						x
	i. d. R. durchschnittlich ⁴					i. d. R. durchschnittlich ⁴					weit überdurchschnittlich				

Anmerkungen: In der Tabelle sind Diagnosekriterien der jeweiligen Phänomene zusammengefasst. Die prototypische Erscheinung ist dabei jeweils durch fett gedruckte Kreuze dargestellt. Kreuze, die in Klammern stehen, deuten an, dass die betreffende Diagnose bei dieser Merkmalsausprägung nur unter bestimmten Voraussetzungen getroffen werden kann. Die jeweiligen Bedingungen sind hier zusammengetragen: ¹ bei weit unterdurchschnittlichen Leseleistungen auch IQ im unterdurchschnittlichen Bereich möglich; ² bei weit unterdurchschnittlichen Rechenleistungen auch IQ im unterdurchschnittlichen Bereich möglich; ³ zur Abgrenzung der allgemeinen Lernschwäche darf der IQ hier nicht unterdurchschnittlich sein; ⁴ bei überdurchschnittlichen Lese- oder Rechenleistungen muss die Diskrepanz zum IQ mind. 2 SD betragen; ⁵ Im Fallinventar wird für jeden Fall i. d. R. entweder nur ein Intelligenztest berichtet; wenn zwei Werte berichtet werden, stimmen diese überein. Die jeweiligen Fähigkeitsausprägungen sind folgendermaßen dargestellt: (--) weit unterdurchschnittlich, (-) unterdurchschnittlich, (○) durchschnittlich, (+) überdurchschnittlich, (++) weit überdurchschnittlich.

Die Diagnosekriterien⁶, die für die Fallgestaltung im Fallinventar maßgeblich waren, können folgendermaßen zusammengefasst werden:

a. *Lese-Rechtschreibstörung:*

- mindestens unterdurchschnittliche Leseleistungen in standardisierten Tests im Vergleich zur Bezugsgruppe, zusätzlich kann Rechtschreibleistung beeinträchtigt sein (vgl. ICD-10);

⁶ Die Formulierung von eindeutigen Diagnosekriterien ist bei klinischen Diagnosen (Lesestörung, Rechenstörung, kombinierte Störung schulischer Fertigkeiten) sehr gut möglich, wohingegen sich dies bei (sonder)pädagogischen Diagnosen (Lernbehinderung, Hochbegabung, Hochbegabung mit Minderleistung) weniger eindeutig gestaltet.

- Diskrepanz von mind. eineinhalb Standardabweichungen zur Vergleichsgruppe oder beträchtlicher Unterschied zu eigenen kognitiven Fähigkeiten (vgl. DGKJP, 2015);
- b. *Rechenstörung*:
 - Rechenleistungen liegen weit unterhalb der Vergleichsgruppe (DGKJP, 2018);
 - Schwierigkeiten betreffen insbesondere den umschriebenen Bereich der Grundrechenarten (ICD-10);
 - zugunsten diagnostischer Eindeutigkeit wurde bei der Fallgestaltung eine Diskrepanz von zwei Standardabweichungen zwischen den Werten im standardisierten Rechentest sowie dem Intelligenztest angesetzt;
- c. *kombinierte Störung schulischer Fertigkeiten*:
 - gleichzeitiges Vorliegen einer Lese-Rechtschreib- und Rechenstörung (ICD-10);
 - zugunsten diagnostischer Eindeutigkeit wurde bei der Fallgestaltung eine Diskrepanz von zwei Standardabweichungen zwischen den Werten in standardisierten Schulleistungstests sowie dem Intelligenztest angesetzt;
- d. *umfassende Lernschwäche (Lernbehinderung)*:
 - weit unterdurchschnittliche kognitive Fähigkeiten gemessen in standardisierten Intelligenztests;
 - den kognitiven Fähigkeiten weitestgehend entsprechende Schulleistungen, die beträchtliche und umfassende Lernentwicklungsverzögerungen anzeigen (Grünke & Grosche, 2014; SenBJW, 2012);
- e. *Hochbegabung*:
 - weit überdurchschnittliche kognitive Fähigkeiten gemessen in standardisierten Intelligenztests;
 - den kognitiven Fähigkeiten entsprechende Schulleistungen, die im oberen Leistungsbereich anzusiedeln sind (Rost & Buch, 2010);
- f. *Hochbegabung mit Minderleistung*:
 - weit überdurchschnittliche kognitive Fähigkeiten gemessen in standardisierten Intelligenztests, die sich in den Schulleistungen allerdings nicht widerspiegeln (Vohrmann, 2018);
 - zugunsten diagnostischer Eindeutigkeit wurde bei der Fallgestaltung eine Diskrepanz von zwei Standardabweichungen zwischen den Werten im Intelligenztest und standardisierten Schulleistungstests angesetzt.

Die Fähigkeitsinformationen, die in den Fallvignetten gegeben werden, sind entsprechend diejenigen, die zur Diagnostik von Lernbesonderheiten essentiell vorhanden sein müssen. Sie beziehen sich daher

sowohl auf Schulleistungen als auch auf kognitive Fähigkeiten. Kognitive Fähigkeiten werden als schulleistungsrelevante Fähigkeiten bezeichnet und sind entsprechend wichtige Informationen für Lehrkräfte (vgl. Hosenfeld & Schrader, 2006). Die Schulleistungen werden in den Fallvignetten über zwei Fähigkeitsbereiche operationalisiert, die im Rahmen der Bildungsstandards genannt werden. Der Fähigkeitsbereich Lesen wird über Informationen zu Lesegeschwindigkeit und Leseverständnis abgedeckt. Für die Rechenfähigkeiten werden Hinweise zu Leistungen im Kompetenzbereich „Zahlen und Operationen“ gegeben. Zu jedem Fähigkeitsbereich wurde dabei nicht nur eine, sondern es wurden jeweils drei verschiedene Informationen vorgegeben. Die Vorgabe unterschiedlicher diagnostischer Informationen soll die Fähigkeit der Studierenden trainieren, *relevante Informationen oder Daten* für die Diagnostik von Lernbesonderheiten zu identifizieren und sie von irrelevanten Informationen zu unterscheiden (vgl. Kaiser et al., 2015; Klug et al., 2013; 2016).

6.4.3 Relevante und irrelevante Informationen zu Fähigkeiten der Schüler*innen

Im diagnostischen Fallinventar werden zu jedem Leistungsbereich Informationen unterschiedlicher Güte vorgegeben, die unterschiedlich relevant zur diagnostischen Urteilsbildung sind. Zu jedem Fähigkeitsbereich wird zum einen eine *formelle diagnostische Information* gegeben, die systematisch, zielorientiert und methodisch kontrolliert erhoben wurde (formelle Diagnose; Schrader, 2010). Vor dem Hintergrund, dass bei der Feststellung von Lernbesonderheiten eine sorgfältige diagnostische Prozedur angezeigt ist, bei der diagnostische Informationen mittels standardisierter Testverfahren erhoben werden (Hesse & Latzko, 2017), sind hier die formellen diagnostischen Informationen als relevante Informationen einzustufen. Diese formellen Diagnoseinformationen stammen überwiegend aus:

- standardisierten Lernstandserhebungen (ILeA: Individuelle Lern-Ausgangslage) und Vergleichsarbeiten (VERA-3, VERA-8)
- standardisierten Schulleistungstests, die von diagnostisch geschulten Pädagog*innen der Schule durchgeführt werden (Sonderpädagog*in, Fachmultiplikator*in)
- schulexternen testpsychologischen Untersuchungen (z. B. Schulpsychologie, Sonderpädagogik)

In einem Fall wird ein selbstkonstruierter Test einer Lehrkraft, der für die betreffende Klassenstufe an der gesamten Schule zum Einsatz gekommen ist, als formelle diagnostische Information vorgegeben. Bei dieser und allen anderen formellen diagnostischen Informationen innerhalb der Lerngelegenheit wird davon ausgegangen, dass diese ohne schwerwiegende Verzerrungen oder Fehler durch den Test selbst (Testgüte), die Testperson oder durch Anwendungsfehler (vgl. Bortz & Döring, 2016, S. 41) entstanden und daher objektiv, reliabel und valide sind.

Neben der formellen diagnostischen Information werden in jeder Fallvignette zu jedem Leistungsbereich zusätzlich je eine *informelle und semiformelle diagnostische Information* (vgl. Hascher, 2008; Schrader, 2010; s. Abschnitt 1.2.3.2) vorgegeben. Diese Informationen können zusätzliche diagnostische Hinweise geben, sind aber nicht als ausschlaggebende Informationen für die zu treffenden diagnostischen Entscheidungen anzusehen und können daher als irrelevant angesehen werden. Diese Informationen haben eine geringere

diagnostische Güte, da sie unsystematisch und/oder basierend auf subjektiven Eindrücken gewonnen wurden. Typische nichtformelle diagnostische Informationen werden gespeist aus:

- beiläufig beobachteten oder zufällig erhaltenen Informationen, z. B. unsystematische Beobachtungen aus dem laufenden Unterrichtsgeschehen oder Gesprächen unter Schüler*innen oder Eltern;
- Selbsteinschätzungen der Schüler*innen;
- Eindrücken und Überzeugungen der Eltern der Schüler*innen;
- subjektiven Eindrücken von anderen Lehrkräften.

Während informelle Informationen als vollkommen beiläufig erhobene Informationen gelten, die in keiner Weise zielgerichtet oder methodisch kontrolliert zustanden gekommen sind, sind semiformelle Informationen als etwas höherwertig einzustufen (vgl. Hascher, 2008). Sie werden im Fallinventar mit standardisierten Instrumenten, aber unsystematisch erhoben oder umgekehrt mit klarer Zielorientierung, aber methodisch nicht kontrolliert (vgl. Tab. 1.1 in Kapitel 1).

Beide Formen der nichtformellen Informationen können, wie es in der Praxis auch üblich ist, *inkongruent zu den formellen Informationen* und daher nicht zur Lösung eines Falles geeignet sein (Übersicht der inkongruenten Informationen für die Fälle in der finalen Fassung des Fallinventars siehe Tab. 6.6). Von den insgesamt neun diagnostischen Informationen pro Fall, die sich aus drei Informationsarten (formell, semiformell, informell) für je drei Fähigkeitsbereiche (Lesen, Rechnen, Intelligenz) ergeben, sind zur Falllösung jeweils nur drei Informationen notwendig und daher ausschlaggebend - nämlich die formellen Informationen (für eine schematische Übersicht über die Gestaltung der Fallvignetten vgl. Abb. 6.2). Durch die Vorgabe verschiedener Arten diagnostischer Informationen, die eine unterschiedliche Relevanz für die Lösung des Falles haben, soll das Bewusstsein angehender Lehrkräfte geschult werden, dass eigene Urteile und die Eindrücke anderer Lehrkräfte stets vorläufig und revisionsbedürftig sind und durch formelle Diagnosen abgesichert werden müssen (vgl. Schrader, 2009, 2010). Dieses Bewusstsein über eigene Urteilsfehler ist ein wichtiger Entwicklungsschritt für die Ausbildung diagnostischer Expertise (vgl. Helmke, 2009). Die Vorgabe inkongruenter Informationen kann darüber hinaus konfliktreiche Erfahrungen hervorrufen, die ein besonders geeignetes Mittel für die Veränderungen konzeptuellen Wissens sind (Helmke & Lenske, 2013).

6.4.4 Irrelevante und potentiell urteilsverzerrende Zusatzinformationen

Die Vorgabe verschiedener Informationen soll dazu beitragen, dass die Studierenden ein Bewusstsein für die eigene Informationsnutzung entwickeln. Das Erkennen und Unterdrücken irrelevanter Informationen fällt Noviz*innen ggü. erfahrenen Lehrkräften aus der Praxis noch schwer: Ihre Urteile sind besonders anfällig für Informationen, die für die aktuelle diagnostische Aufgabe eigentlich irrelevant sind oder sogar verzerrend wirken (z. B. Krolak-Schwerdt et al., 2009, 2013). Um die Studierenden auf die Wirkkraft verzerrender Einflüsse aufmerksam zu machen, enthalten die Fallvignetten im Fallinventar neben den nichtformellen Diagnoseinformationen noch weitere urteilsirrelevante Informationen. In der vorliegenden Arbeit wird das Merkmal des *Geschlechts der virtuellen Schüler*innen* in den Vordergrund gestellt, da es im Rahmen von Fallvignetten experimentell sehr gut variiert werden kann (vgl. Kaiser et al., 2015) und sich in verschiedenen Feldstudien als eine robuste Einflussgröße auf die Urteile von Lehrkräften erwiesen hat. So werden Mädchen gemeinhin positiver beurteilt als Jungen (z. B. Ritts et al., 1992), und es zeigen sich außerdem domänenspezifische Fehlurteile, die durch geschlechterstereotype Vorurteile bedingt sind (vgl. Kaiser et al., 2015): Während Mädchen stereotyp im Bereich des Lesens kompetenter eingeschätzt werden als im Bereich des Rechnens, verhält es sich bei Jungen i. d. R. genau anders herum (aber auch widersprüchliche Befunde, vgl. Kaiser et al., 2015). Denkbar wären andere Merkmale von Schüler*innen, von denen bekannt ist, dass diese die Urteile von Lehrkräften verzerren können (vgl. heuristisches Modell der Urteilsgenauigkeit: Südkamp et al., 2012). Die Einführung ausgewählter urteilsverzerrender Variablen zielt auf die Optimierung des diagnostischen Prozesses der Studierenden. Auf diese Weise können die Informationsauswahl und –verarbeitung trainiert werden, die wiederum einen Einfluss auf die Güte der diagnostischen Produkte am Ende des diagnostischen Prozesses haben (z. B. Behrmann & Van Ophuysen, 2017; vgl. Kap. 2.2.1.1). Durch die Vorbereitung der Studierenden auf die optimale Gestaltung des diagnostischen Prozesses soll dementsprechend auch die Qualität der diagnostischen Produkte in der späteren Berufspraxis erhöht werden.

6.4.5 Diagnosen und pädagogische Entscheidungen als diagnostische Produkte

Zu den diagnostischen Produkten am Ende des diagnostischen Prozesses zählen in erster Linie die Diagnosen selbst (vgl. z. B. Behrmann & Van Ophuysen, 2017; Herppich et al., 2017, 2018). Da Diagnosen im schulischen Kontext in der Regel „nur“ als Grundlage für daraus abgeleitete pädagogische Maßnahmen dienen (Hesse & Latzko, 2017; Weinert, 2000/2001), ist im Fallinventar die Ableitung von pädagogischen Anschlusshandlungen genauso im Vordergrund wie das diagnostische Urteil selbst. Mithilfe der vorliegenden Lerngelegenheit soll also die Güte von zwei Produkten des pädagogisch-diagnostischen Prozesses verbessert werden, was durch die gezielte Gestaltung des diagnostischen Prozesses erreicht werden soll (vgl. Abschnitte 6.4.3 und 6.5.4). Hierzu erhalten die Studierenden zwei Aufgabenstellungen:

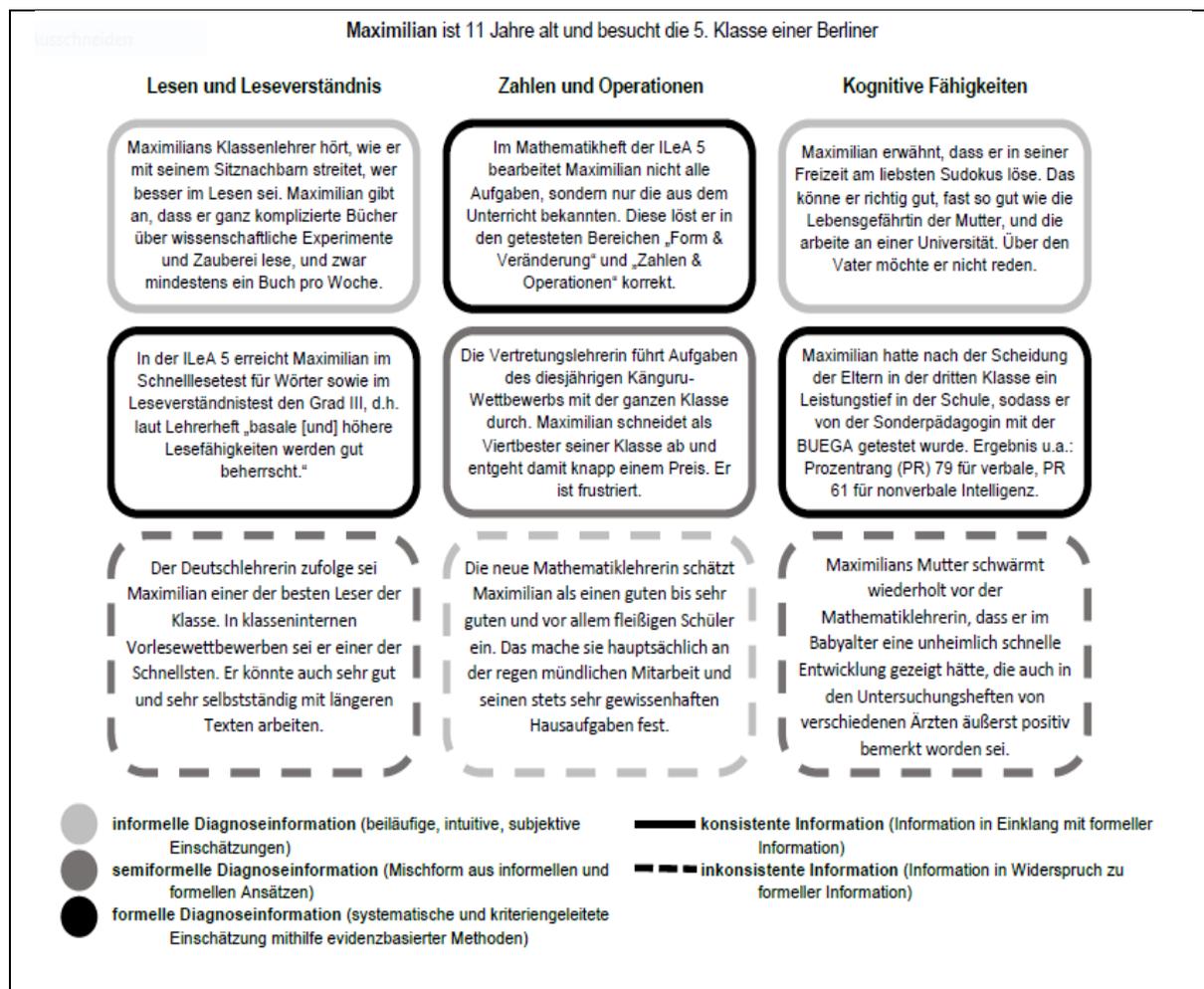


Abbildung 6.2 Schematischer Aufbau eines Fallblattes im diagnostischen Fallinventar

Anmerkungen: Abgebildet ist hier eine Fallvignette der finalen Version des Fallinventars. Diese Fallvignette gibt einen beispielhaften Einblick über den Aufbau der Falldarstellungen: Jede Fallvignette enthält verschiedene diagnostische Informationen, die zur Lösung der diagnostischen Aufgaben mehr oder weniger relevant sein können. Zu drei Fähigkeitsbereichen (Lesen und Leseverständnis, Zahlen und Operationen, Kognitive Fähigkeiten) werden jeweils drei Arten von Diagnoseinformationen vorgegeben. Formelle Diagnoseinformationen (schwarz) sind bevorzugt zu nutzen und ausschlaggebend für die Lösung. Informelle (hellgrau) sowie semiformelle (mittelgrau) Diagnoseinformationen können ergänzend zur formellen Information genutzt werden, sind jedoch nicht relevant, um korrekte diagnostische Entscheidungen abzuleiten. Sie können kongruent (durchgezogene Linie) oder inkongruent (gestrichelte Linie) zur formellen Diagnoseinformation innerhalb eines Fähigkeitsbereichs sein.

(1) Abgabe einer Diagnosevermutung

Erstes Ziel bei der Bearbeitung des Fallinventars ist die Erstellung korrekter Diagnosen, die die Grundlage für die pädagogischen Entscheidungen bilden. Da allerdings die Menge an verfügbaren diagnostischen Daten sehr begrenzt ist, werden die Studierenden darauf hingewiesen, dass Diagnosen hier nicht mit Sicherheit vergeben werden können, sondern Vermutungen darstellen. Um die Vergabe einer korrekten Diagnose im Fallinventar zu erleichtern, sind die diagnostischen Informationen so gestaltet, dass sie nicht integriert werden müssen, wie es in der Praxis i. d. R. erforderlich ist (vgl. Jäger, 2006), sondern selektiert werden können. Die Selektion der korrekten Diagnose wird durch die Vorgabe von sechs

Antwortoptionen unterstützt, von denen eine Antwort auszuwählen ist (Multiple-Choice-Aufgabe). Für die Abgabe einer Diagnosevermutung stehen folgende Optionen zur Verfügung:

- a. Normal begabte/r Schüler*in mit evtl. Stärken/Schwächen in einem/mehreren Fächern,
- b. Schüler*in mit einer spezifischen Lernstörung im Bereich des Lesens oder Rechnens ,
- c. Kombinierte Störung schulischer Fertigkeiten,
- d. Schüler*in mit einer allgemeinen Lernschwäche bzw. „Lernbehinderung“,
- e. Hochbegabte/r Schüler*in mit guten bis hervorragenden Schulleistungen,
- f. Hochbegabte/r Minderleister/in.

(2) Abgabe einer Empfehlung für eine pädagogische Anschlusshandlung

Das zweite und in der pädagogischen Praxis wichtigere Ziel ist die Ableitung pädagogischer Anschlusshandlungen für die Schüler*innen in den Fallvignetten. Die Studierenden können hierzu Empfehlungen abgeben, die sich auf zusätzlich zum regulären Unterricht stattfindende Fördermaßnahmen beziehen. Bei diesen Selektionsentscheidungen (vgl. Jürgens & Lissmann, 2015) wird die*der Schüler*in einer Bildungsmaßnahme, -gruppe oder -institution zugewiesen, die ihren Lernprozess optimal unterstützen kann. Die Zusammenstellung der konkreten Fördermöglichkeiten orientiert sich an Maßnahmen, die im Rahmen von Schulverordnungen oder Leitfäden für die inner- sowie außerschulische Förderung in Berlin regulär angedacht sind (GsVO, 2005; SenBJF, 2018). Auf innerschulischer Ebene kommen bspw. der Förderunterricht in Kleingruppen, auf außerschulischer Ebene kommen lerntherapeutische Angebote oder spezielle Maßnahmen zur Hochbegabtenförderung in Betracht. Um die Studierenden bei der Vergabe von Empfehlungen zu unterstützen, werden hier ebenfalls sechs Antwortoptionen zur Verfügung gestellt, wobei je nach Phänomen mehrere der nachfolgenden Antworten richtig sein können (Multiple-Answer-Format):

- a. sonderpädagogische Förderung der kognitiven Fähigkeiten,
- b. schulische Kleingruppenförderung oder lerntherapeutische Förderung im Lesen,
- c. schulische Kleingruppenförderung oder lerntherapeutische Förderung im Rechnen,
- d. Teilnahme an der Arbeitsgruppe „Club Einstein“ für mathematisch sehr begabte Schüler*innen,
- e. Teilnahme an der Arbeitsgruppe „Club Goethe“ für schriftsprachlich sehr begabte Schüler*innen,
- f. keine besonderen Maßnahmen sind erforderlich, die reguläre Förderung im Rahmen des Unterrichts ist ausreichend (diese Antwortoption schließt die vorigen aus).

Bei der Auswahl ihrer Antwortoptionen werden die Studierenden darauf hingewiesen, dass die Plätze in den jeweiligen Angeboten begrenzt sind und nur Schüler*innen mit dringendem Handlungsbedarf zur Verfügung stehen. Sie werden daher gebeten, nur diejenigen Fördermaßnahmen für die*den entsprechende*n Schüler*in auszuwählen, die aus ihrer Sicht dringend notwendig sind (Prinzip der Sparsamkeit). Dringend notwendig sind solche Maßnahmen, die der Schülerin oder dem Schüler in der Falldarstellung ein Lernen und Leisten ermöglichen, die ihren allgemeinen kognitiven Fähigkeiten entsprechen. Bei der Auswahl der korrekten Antworten müssen die Studierenden demnach erstens auf die

spezifischen Symptome (Prinzip der *symptomorientierten Förderung*, vgl. Abschnitt 6.2.2) sowie zweitens auf eventuelle Diskrepanzen zwischen Teilleistungen und allgemeinen kognitiven Fähigkeiten achten.

6.5 Pilotierung des Fallinventars

Die erste Version des diagnostischen Fallinventars enthielt sechs Fallvignetten, die bereits im Wesentlichen den beschriebenen konzeptionellen Prinzipien folgten. Die Fallgestaltung sowie die Formulierung der Aufgabenstellungen wurden nach den Erfahrungen mit dem Prototypen optimiert und in die finale Fassung gebracht (vgl. Abschnitt 6.6). Die prototypischen Fallvignetten wurden im Sommersemester 2016 pilotiert. Dabei sollte geprüft werden, wieviel Zeit die Studierenden bei der Bearbeitung der Fallvignetten benötigen und wie schwierig sich die Fallbearbeitung für sie darstellt. Von besonderem Interesse waren die Lösungsraten für die einzelnen Fälle, d. h. eine Abschätzung darüber, ob die Fälle ein angemessenes Schwierigkeitsniveau aufweisen.

6.5.1 Stichprobe

Die Pilotierung erfolgte an 90 Lehramtsstudierenden, die das Seminar „Pädagogische Diagnostik“ im ersten Mastersemester an der Freien Universität besuchten. Die Teilnehmer*innen an der Studie ($N = 58$) stammten aus drei Seminaren einer Seminarleiterin (Seminar 1: $N = 26$ von 40 Teilnehmer*innen, Seminar 2: $N = 17$ von 25 Teilnehmer*innen, Seminar 3: $N = 15$ von 25 Teilnehmer*innen). Die Rücklaufquoten lagen in allen Seminaren vergleichbar zwischen 60 % (Seminar 3) und 68 % (Seminar 2). Die Teilnehmer*innen waren überwiegend Studierende mit Lehramtsoption für weiterführende Schulen ($N = 55$) und zu einem geringen Teil Studierende der Grundschulpädagogik ($N = 2$). Über die Hälfte der Studierenden waren weiblich (60,7 %) und die große Mehrheit ($N = 54$) hatten zum Zeitpunkt der Erhebung keine (66,7 %) oder sehr wenig (28 %) Erfahrung mit standardisierten Tests.

6.5.2 Instrument

Die sechs prototypischen Fallvignetten wurden in drei Fallheften mit jeweils zwei Fällen zusammengestellt:

- Fallheft 1: normalbegabter Schüler – Grundschule (Jakob), hochbegabte Hochleisterin – Gymnasium (Johanna);
- Fallheft 2: normalbegabter Schüle mit Mathestärke – Gymnasium (Julius), Junge mit kombinierter Störung schulischer Fertigkeiten – Grundschule (Louis);
- Fallheft 3: Mädchen mit einer Lernbehinderung – Grundschule (Louisa), hochbegabter Minderleister – Integrierte Sekundarstufe (Jonathan).

Jedes Fallheft enthielt eine Fallvignette aus dem Grundschulbereich sowie eine aus der weiterführenden Schule (Gymnasium oder Integrierte Sekundarstufe). Der Aufbau der Fallblätter war für beide Schulstufen vergleichbar: Für drei Fähigkeitsbereiche (Lesen, Rechnen, Intelligenz) wurden jeweils drei diagnostische Informationen zur Verfügung gestellt, von denen eine formell und zwei nichtformell waren. Die

nichtformellen Informationen waren teilweise inkongruent zu den formellen Informationen. Die konsequente Abgrenzung informeller und semiformeller Informationen wurde erst in der Überarbeitung umgesetzt (vgl. Abschnitt 6.6.1). Die Namen und die Fallgeschlechter der dargestellten Schüler*innen unterschieden sich hier ebenfalls von der finalen Version (vgl. Abschnitt 6.6.2).

6.5.3 Durchführung

Jedes der prototypischen Fallhefte wurde in je einem der drei Seminare vorgegeben, sodass innerhalb jedes Seminars jede/r Studierende die gleichen zwei Fallvignetten erhielt. Für diese zwei Fälle sollten die Studierenden drei Aufgabenstellungen bearbeiten: Zunächst wurden sie gebeten, eine *Empfehlung für eine pädagogische Maßnahme* auf Basis der vorliegenden Daten abzugeben. Da die Korrektheit der Empfehlung in der Pilotierung als zentrale Zielvariable angenommen wurde, wurde diese Entscheidung hier vor der Vergabe der Diagnose getroffen.

Die *Angabe der Diagnosevermutung* erfolgte hier erst im zweiten Schritt und sollte absichern, dass die Studierenden nicht deshalb zu falschen Empfehlungen gelangten, weil sie implizit von einer falschen Diagnose ausgingen. Zur Angabe der Empfehlung und der Diagnose wurden je sechs Antwortoptionen vorgegeben, für die die Studierenden für jede einzelne Antwortoption durch Ankreuzen (ja/nein) entscheiden sollten, ob sie diese Option wählen würden („Würden Sie diese Maßnahme empfehlen?“ / „Würde Sie diese Diagnose vergeben?“).

Drittens wurden die Studierenden um die Angabe der *subjektiv empfundenen Schwierigkeit* des Falles gebeten. Die Angabe der subjektiv empfundenen Fallschwierigkeit erfolgte auf einer sechsstufigen Likert-Skala (-3 = sehr leicht, +3 = sehr schwer). Für die finale Version sollten nur Fälle aufgenommen werden, die weder als sehr leicht noch als sehr schwer bewertet wurden. Abschließend wurden die Studierenden um eine Einschätzung gebeten, wie hilfreich sie die Übung für die Modul abschließende Klausur sowie die spätere Praxis empfinden (-3 = gar nicht hilfreich, +3 = sehr hilfreich). Außerdem wurden Bemerkungen und Verbesserungsvorschläge in offenem Antwortformat von den Studierenden erbeten. Die Studierenden vermerkten ihre gesamten Angaben auf Antwortbögen, die sie nach Bearbeitung der Übung anonym und freiwillig abgeben konnten.

6.5.4 Auswertung

Die Angabe der Studierenden dazu, wie hilfreich sie die Übung für die Abschlussprüfung bzw. die spätere Berufspraxis empfinden, wurde in den positiven Wertebereich transkodiert (1 = gar nicht hilfreich, 6 = sehr hilfreich).

Zur Berechnung der *Lösungsraten der Fälle* wurden die Empfehlungen und Diagnosen der Teilnehmer*innen jeweils mit einer Musterlösung abgeglichen. Ein Item, d. h. eine fallbezogene Entscheidung, galt als gelöst, wenn die ausgewählten Antwortoptionen mit der jeweiligen Musterlösung übereinstimmten. Während für die Diagnosen stets nur eine Antwortoption pro Fall korrekt war, gab es bei den Empfehlungen teilweise mehrere Muster korrekter Antwortoptionen. Die Musterlösung zu den

pädagogischen Anschlusshandlungen orientierte sich sowohl an den spezifischen Schwächen bzw. Stärken hinsichtlich der schulischen Teilleistungen (Lesen, Rechnen) als auch anhand der Diskrepanz dieser Teilleistungen zu den kognitiven Fähigkeiten (vgl. Abschnitt 6.4.5). Grundlage der Musterlösungen bildete also das jeweilige Fähigkeitsprofil, das jedem Fall zugrunde liegt (eine Übersicht über die Musterlösungen für Diagnosen und Empfehlungen findet sich in Tab. 2.1 im folgenden Kapitel). Für jede korrekte Antwort (Diagnose) bzw. jedes korrekte Antwortmuster (Empfehlung) erhielten die Studierenden je einen Punkt, für falsche oder teilweise richtige Muster erhielten sie null Punkte. Die über alle Studierenden gemittelte Punktzahl pro Aufgabe und Fall entspricht der Lösungsrate bzw. der Aufgabenschwierigkeit.

Für die Berechnung der mittleren *subjektiv empfundenen Schwierigkeit* der Fälle wurden die Ratings der Studierenden zunächst in den positiven Wertebereich transkodiert (Skala 1 - 6). Um die subjektive Schwierigkeit der Fälle mit den tatsächlichen Schwierigkeitsindizes vergleichen zu können, wurde die Skala anschließend rekodiert (1 = sehr schwer, 6 = sehr leicht). Durch Subtraktion des Wertes 1 von jeder Zahl ergaben sich Skalenwerte im Bereich 0 – 5. Durch anschließende Division jedes Wertes durch den Zahlenwert 5 entstand eine Skala, deren Wertebereich 0 – 1 betrug (0 = sehr schwer, 1 = sehr leicht) und deren Interpretation mit der Aufgabenschwierigkeit übereinstimmte. Für jeden Fall wurde abschließend die mittlere subjektive Schwierigkeit über alle Studierenden berechnet.

6.5.5 Ergebnisse

Die Bearbeitung der Fälle selbst gelang den Studierenden augenscheinlich recht gut. Die durchschnittliche Bearbeitungsdauer für einen Fall betrug 10 - 15 Minuten. Die Dozentin der Seminare berichtete, dass die Sitzung mit dem Fallinventar einen positiven Eindruck bei ihr hinterlassen habe und dass die Studierenden interessiert und angeregt an der Sitzung teilgenommen hätten. Dieser Eindruck bestätigte sich bei der Betrachtung der Nutzenempfindungen der Studierenden: In Bezug auf die Klausur empfanden die Teilnehmer*innen die Sitzung im Mittel eher hilfreich ($M = 3.85$, $SD = 1.37$, $N = 53$), und in Bezug auf die Praxis eindeutig hilfreich ($M = 4.73$, $SD = 1.14$, $N = 56$).

Die *Analyse der Lösungsraten* (vgl. Tabelle 6.5) zeigte, dass die Studierenden die unterschiedlichen Fälle unterschiedlich gut lösten. Insbesondere hinsichtlich der Diagnosen fielen die Lösungsraten weit auseinander, sodass sich eine Spannweite von sehr schwer bis sehr leicht ergab ($p = .13$ bis 1.00). Bei den Empfehlungen lagen fünf von sechs Items im mittleren Schwierigkeitsbereich ($p = .47$ bis $-.54$) und eines im leichten Bereich ($p = .80$). Neben der Varianz zwischen den Fällen waren hinsichtlich der Lösungsraten für Diagnosen vs. Empfehlungen teilweise auch deutliche Unterschiede innerhalb der Fälle zu beobachten. Beim Fall „Louisa“ (Schülerin mit Lernbehinderung) war die Diskrepanz zwischen den Lösungsraten für Empfehlungen und Diagnosen am größten: Dort gelang es den Studierenden i. d. R. gut, Empfehlungen abzuleiten, wohingegen die Vergabe der Diagnose im Durchschnitt sehr schwer fiel. In vier weiteren Fällen erreichten die Studierenden bei den Diagnosen jedoch höhere Lösungsraten als bei den Empfehlungen. In diesen Fällen waren die Diskrepanzen geringer. Lediglich im Fall Louis (kombinierte

Störung schulischer Fertigkeiten) stimmten die Lösungsraten für Diagnose und Empfehlung weitestgehend überein.

Tabelle 6.5 Ergebnisse der Pilotierung des Fallinventars für drei Seminare

	Prototypische Fallvignette					
	Jakob (<i>N</i> = 24)	Johanna (<i>N</i> = 23)	Julius (<i>N</i> = 16)	Louis (<i>N</i> = 15)	Louisa (<i>N</i> = 15)	Jonathan (<i>N</i> = 14)
Empfehlung mittlere Lösungsrate (<i>SD</i>)	.54 (.50)	0.54 (.50)	.50 (.50)	.47 (.50)	.80 (.40)	.53 (.50)
Diagnose mittlere Lösungsrate (<i>SD</i>)	.85 (.36)	1.00 (.00)	.69 (.46)	.41 (.49)	.13 (.34)	.80 (.40)
mittlere geratete Schwierigkeit (<i>SD</i>)	3.38 (1.25)	4.52 (.93)	3.94 (1.20)	2.93 (1.12)	3.07 (1.00)	4.86 (1.30)
subjektive Schwierigkeit	.48	.70	.55	.39	.41	.77

Anmerkungen: Die Fallvignetten „Jakob“ (normalbegabter Schüler der Grundschule) und „Johanna“ (hochbegabte Hochleisterin am Gymnasium) kamen in Seminar 1 zum Einsatz. Die Fallvignetten „Julius“ (normalbegabter Schüler mit Mathestärke am Gymnasium) sowie „Louis“ (Grundschüler mit kombinierter Störung schulischer Fertigkeiten) wurden in Seminar 2 eingesetzt. Im dritten Seminar wurden die Fallvignetten „Louisa“ (Grundschülerin mit Lernbehinderung) und „Jonathan“ (hochbegabter Minderleister an einer Integrierten Sekundarschule) vorgegeben.

Die *subjektiv empfundene Schwierigkeit*, die die Studierenden bei der Fallbearbeitung empfanden, wich von der tatsächlichen Itemschwierigkeit mehr oder weniger weit ab (vgl. Tab. 6.5). Die Studierenden gaben für fünf Fälle ein mittelgroßes Schwierigkeitsempfinden an, das im Bereich 2.93 – 4.42 („eher leicht“ bis „eher schwer“) angesiedelt war. Bei der Bearbeitung des Falles „Jonathan“ (hochbegabter Minderleister) gaben die Studierenden an, dass sie diesen Fall als „leicht“ ($M = 4.86$) empfunden hätten. Die Umrechnung der Ratings in subjektive Itemschwierigkeiten ergaben, dass die Fälle im Bereich mittlerer bis relativ geringer Itemschwierigkeit ($p_{\text{subj}} = .39 - .77$) einzuordnen waren. Der auf Grundlage dieser Umrechnung mögliche Vergleich zwischen tatsächlichen Itemschwierigkeiten und subjektiven Fallschwierigkeiten ergab folgendes Bild: In zwei Fällen (Louis, Jonathan) waren die subjektiven Schwierigkeiten vergleichbar mit den Diagnoseschwierigkeiten, bei zwei Fällen (Jakob, Julius) war sie der Schwierigkeit der Empfehlung ähnlich, und in den verbleibenden zwei Fällen (Johanna, Louisa) ordnete sie sich zwischen die Schwierigkeiten von Diagnose und Empfehlung ein.

6.5.6 Diskussion und Schlussfolgerungen

Dieser erste Einsatz der Fallvignetten zeigte, dass sie für die Bearbeitung durch Studierende im intendierten Seminar geeignet sind. Die Rezeption der Übung fiel durchaus positiv aus – und das, obwohl der Großteil der Teilnehmer*innen das Lehramt an weiterführenden Schulen anstrebte und daher zur Diagnostik von Lernbesonderheiten i. d. R. weniger Bezug hatte als es bspw. angehende Grundschullehrkräfte hätten. Eine mögliche Erklärung für diese Bewertung könnte die relativ breite Streuung der Schwierigkeitsindices sein, die als zweites Ergebnis festgehalten werden konnte: Während einige fallbezogene Entscheidungen von nur wenigen Studierenden gelöst werden konnten, waren andere

Items wiederum von einem Großteil der Studierenden gut lösbar. In einem Fall konnten sogar ausnahmslos alle Studierenden die richtige Diagnose vergeben. Im Gegensatz dazu konnte in diesem Fall aber nur die Hälfte der Studierenden auch korrekte Empfehlungen ableiten. Für die finale Version wurden sämtliche Fallvignetten, die hier erprobt wurde, als grundsätzlich geeignet betrachtet, auch wenn sie extreme Schwierigkeitsindices aufwiesen. Für die Gestaltung des Fallinventars waren Items mit mittlerer Schwierigkeit angestrebt (Atkinson, 1957), aber zusätzlich Items extremer Schwierigkeit erwünscht: Während leichte Items dem Empfinden von Lernfreude zuträglich sein könnten, können schwere Items potentiell zu einer intensiven kognitiven Auseinandersetzung und damit zum Aufbau konzeptuellen Wissens beitragen (vgl. Schnotz, 2006). Die subjektive Schwierigkeit lag für alle Fälle im mittleren Bereich, sodass auch davon auszugehen war, dass sich die Studierenden nicht überfordert fühlten.

Unerwartet war der Befund der *Diskrepanzen zwischen den Lösungsraten* für Diagnosen und Empfehlungen. Mögliche Ursachen für das Zustandekommen dieser Unterschiede konnten im Rahmen dieser Studie nicht geprüft werden, jedoch sind grundsätzlich drei Aspekte denkbar, die die höheren Lösungsraten bei Diagnoseentscheidungen erklären könnten. Erstens könnte das Ergebnis dafür sprechen, dass Empfehlungen im Rahmen dieser fallbasierten Lerngelegenheit grundsätzlich schwerer zu vergeben sind als Diagnosen. Während bei Diagnoseentscheidungen auf Grundlage der hier sehr eindeutigen Umsetzung der Diagnosekriterien nur eine Diagnose richtig sein konnte, war bei den Empfehlungen teilweise die Angabe mehrerer Antwortoptionen korrekt. Bei den Empfehlungen war dadurch die Fehlerwahrscheinlichkeit rein statistisch eine höhere als bei Diagnosen. Ein Teil der Fehler der Studierenden bei den Empfehlungen stand auch mit der Formulierung der Antwortoptionen im Zusammenhang: So hatte ein Teil der Studierenden bei der Beantwortung der Empfehlungsaufgabe nicht erkannt, dass das Ankreuzen der Antwortoption „Regelunterricht“ die anderen Optionen ausschloss. Die Instruktion enthielt ebenfalls keinen expliziten Hinweis, dass bei den Empfehlungen zunächst eine Entweder-Oder-Entscheidung hinsichtlich der Zusatzmaßnahmen zum Regelunterricht getroffen werden musste. Die Formulierung der Aufgabenstellung wurde entsprechend vor dem nächsten Einsatz des Fallinventars überarbeitet (vgl. Abschnitt 6.6). In der zweiten Studie dieser Arbeit wurden die Musterlösungen außerdem durch Expert*innen im Feld validiert (vgl. Tab. 2.1).

Zweitens könnte die Bearbeitungsreihenfolge bei der Pilotierung unvoreilhaft gewesen sein, die Formulieren einer Empfehlung vor der Diagnosevermutung vorsah. Diese Reihenfolge war gewählt worden, da in der Pilotierung die Abgabe der Empfehlung zentral war und die Erfassung der Diagnoseentscheidung zunächst von nachgeordnetem Interesse war. Die Erforschung eines Effekts der Reihenfolge diagnostischer Entscheidungen wurde vor dem Hintergrund der Pilotergebnisse zum Desiderat für weitere Studien. Denkbar wäre, dass die Vergabe der Diagnose vor der Empfehlung, wie es im normativen diagnostischen Prozess vorgesehen ist (vgl. Jäger, 2006; Hesse & Latzko, 2017), sich günstiger auf die Lösungsraten bei den Empfehlungen auswirken könnte. Um herauszufinden, ob die Lösungsrate der beiden Entscheidungen durch die Reihenfolge ihrer Vorgabe beeinflusst werden kann, müsste diese systematisch variiert werden. Durch diese Variationen könnte sich zeigen, ob die

Lösungsraten bei den Empfehlungen dadurch zu steigern wären oder ob sie trotzdem niedriger bleiben als bei Diagnosen. Wenn letzteres zuträfe, wäre es ein Hinweis darauf, dass das Vergeben von Empfehlungen grundsätzlich schwerer sein könnte als das von Diagnosen. Die Effekte dieser Variation wurden in der zweiten Studie dieser Arbeit untersucht (vgl. Kap. 7).

Drittens könnten diese Unterschiede auch ein Hinweis auf die diagnostische Informationsverarbeitung sein: Während bei Diagnosen möglicherweise die formellen Informationen – wie intendiert – tatsächlich stärker genutzt wurden, könnten sich die Bezugsinformationen bei Empfehlungen möglicherweise hiervon unterscheiden haben. Durch vermehrte Nutzung nichtformeller Informationen, die in vielen Fällen in Widerspruch zur formellen Information standen, würden die Studierenden zu falschen Schlüssen kommen, die sich möglicherweise in den höheren Fehlerzahlen bei den Empfehlungen niederschlugen. Dieser Vermutung konnte anhand der vorliegenden Daten allerdings nicht geprüft werden. Die Analyse der Nutzung diagnostischer Informationen für die jeweiligen Entscheidungen ist ein weiteres Forschungsdesiderat dieser Arbeit. In der zweiten Studie (vgl. Kap. 7) wird geprüft, inwieweit sich die Nutzung der unterschiedlichen Informationen für Diagnosen und Empfehlungen unterscheidet und ob diese auch in Zusammenhang mit den Lösungsraten steht.

Anhand dieser ersten Daten bleibt ferner unklar, wodurch die Abweichungen der *subjektiv empfundenen Schwierigkeiten* der Fälle von den tatsächlichen Schwierigkeitsindices zustande kamen. Diese Abweichung könnte zum einen ein Hinweis darauf sein, dass die Studierenden ihre eigenen Fähigkeiten überschätzten, was bei Noviz*innen nicht ungewöhnlich wäre (Müller, 2010). Es wäre andererseits auch denkbar, dass dies ein Hinweis für die Verarbeitung der diagnostischen Informationen ist. Die vorliegenden Ergebnisse lassen vermuten, dass sich das Schwierigkeitsempfinden in den verschiedenen Fällen an verschiedenen Bezugsinformationen bzw. wahlweise an der Schwierigkeit der Diagnose bzw. Empfehlung orientieren könnte. Da das subjektive Schwierigkeitsempfinden fallübergreifend und nicht für die einzelnen Entscheidungen abgefragt wurde, kann hier keine Aussage darüber gemacht werden, ob sich diese eher an der Vergabe der Diagnose oder der Ableitung der Empfehlung orientiert hat oder ob ggf. auch eine Integration beider Entscheidungsschwierigkeiten stattgefunden hat. Eine nach der Entscheidung getrennte Abfrage könnte möglicherweise erklären, woran sich das Schwierigkeitsempfinden orientiert. In der zweiten Studie dieser Arbeit erfolgte daher eine getrennte Betrachtung der Sicherheit beim Treffen bei Empfehlungen vs. Diagnosen.

6.6 Entwicklung der finalen Version des Fallinventars

Um einem Teil der durch die Pilotierung aufgeworfenen Fragen nachgehen zu können, wurden vor dem weiteren Einsatz des Fallinventars im Wintersemester 2016/17 zentrale Änderungen vorgenommen. Diese bezogen sich sowohl auf die Vignetten selbst als auch auf die Aufgabenstellungen, die mithilfe der Vignetten bearbeitet werden.

6.6.1 Überarbeitung der Aufgabenstellungen

Hinsichtlich der *Aufgabenstellungen* wurden folgende Änderungen vorgenommen:

- *Formulierung der Empfehlungs-Antworten als besondere Form von Multiple-Answer-Items*: Um die Aufgabenstellung hier eindeutiger zu formulieren, wurde die Option „Regelunterricht ausreichend“ ans Ende der Liste der Antwortoptionen gesetzt und durch ein „oder“ klar von den vorausgehenden Optionen getrennt. Dadurch sollte den Studierenden bei der Bearbeitung klar werden, dass diese Option die anderen ausschließt, und diesen möglicherweise durch ein Missverständnis bedingten Fehler verhindern. Die Antwortoptionen sollten nun außerdem durch einfaches Ankreuzen der bevorzugten Option möglich sein (keine Nein-Antworten mehr möglich).
- *Formulierung der Diagnose-Antworten als Multiple-Choice-Items*: Auch bei der Diagnoseentscheidung sollten die Studierenden nicht mehr für jede Option einzeln entscheiden, sondern sie sollten nun aufgefordert werden, sich für die Angabe einer Diagnose zu entscheiden. Dadurch könnte sich die Ratewahrscheinlichkeit zwar geringfügig erhöhen, aber gleichzeitig das Verständnis der Aufgabe erleichtern und zudem eine Zeitersparnis bei der Bearbeitung erreichen.
- *Erweiterung zur Angabe der subjektiven Schwierigkeit*: Anstelle der Abfrage der subjektiv empfundenen Fallschwierigkeit sollte die subjektive Sicherheit für beide Entscheidungen erfolgen. Dieses Maß wurde in der Folgestudie eingesetzt (vgl. Kap. 7), um Hinweise auf die entscheidungsbezogenen Sicherheiten oder Unsicherheiten für die Fallvignetten zu erhalten.

6.6.2 Überarbeitung der Fallvignetten

Zunächst wurden zwei weitere Fallszenarien erarbeitet, die das Fallheft komplettieren sollten. Die bestehenden Vignetten und die neukonstruierten Szenarien wurden anschließend dahingehend überarbeitet, dass sie auch zur Untersuchung von Fragestellungen zum diagnostischen Prozess und der Verarbeitung diagnostischer Informationen geeignet sind. Diese Überarbeitung enthielt zwei zentrale Aspekte, die nachfolgend erläutert werden. Das Fallinventar in der finalen Version enthält insgesamt acht Fälle, von denen drei Fälle Schüler*innen ohne Lernbesonderheiten und fünf Fälle Schüler*innen mit einer Lernbesonderheit darstellen. Jede der ausgewählten Lernbesonderheiten wird in je einem Fall dargestellt. Eine Kurzzusammenfassung der Szenarien zu den einzelnen Phänomenen findet sich in Tabelle 6.6.

6.6.2.1 Formalisierungsgrad, Positionierung und Kongruenz der diagnostischen Informationen

Die sechs Fälle in der Pilotversion gaben pro Szenario drei formelle (für jeden Fähigkeitsbereich eine) und sechs nichtformelle (für jeden Fähigkeitsbereich zwei) Informationen vor. Die unterschiedliche Nutzung

Tabelle 6.6 Kurzbeschreibung der Fallvignetten in der finalen Version des Fallinventar

Fallname	Fallbeschreibung (Diagnose)	Inkongruente Informationen ²		
		Lesen und Leseverstehen	Zahlen und Operationen	Kognitive Fähigkeiten
Julia / Maximilian	insgesamt homogenes Leistungsbild mit guten Lese- und Rechenleistungen, guten Mathematikleistungen und gut durchschnittlichen kognitiven Leistungen („normal begabte/r“ Schüler/in)	semiformell	informell	semiformell
Lena / Christian	bei weit überdurchschnittlichen kognitiven Fähigkeiten liegen die Lese- und Rechenleistungen erwartungswidrig nur im durchschnittlichen Bereich (Hochbegabte/r Minderleister/in)	informell	semiformell	informell
Katrin / Felix ¹	insgesamt homogenes Leistungsbild mit Schulleistungen wie auch allgemeinen kognitiven Fähigkeiten im unteren Durchschnittsbereich („normal begabte/r“ Schüler/in)	semiformell	keine	semiformell
Sophie / Lukas ¹	Leseleistungen im schwachen Leistungsbereich, aber gute mathematische Kompetenzen und allgemeine kognitive Fähigkeiten (Schüler/in mit umschriebener Lesestörung)	keine	keine	keine
Hanna / Jonas	homogenes Leistungsbild mit schwachen Lese- und Rechenleistungen, die aufgrund der allgemeinen Intelligenz auch zu erwarten wären (Schüler/in mit „Lernbehinderung“ / umfassender Lernschwäche)	informell	keine	informell
Jana / Luca	schwache Lese- und Rechenleistungen, die in deutlichem Widerspruch zur allgemeinen Intelligenz stehen (kombinierte Störung schulischer Fertigkeiten)	informell	informell	keine
Leonie / Jan	bei gut durchschnittlichen kognitiven Fähigkeiten und Leseleistungen liegen die rechnerischen Fähigkeiten im deutlich überdurchschnittlichen Bereich („normal begabte/r“ Schüler/in)	informell	informell	informell
Laura / Thomas	homogenes Leistungsbild, bei dem die Lese-, Rechen- sowie die kognitiven Fähigkeiten allgemein als weit überdurchschnittlich zu bezeichnen sind (Hochbegabte/r Hochleister/in)	keine	informell	informell

Anmerkungen. ¹ Fallvignette in der Pilotversion noch nicht vorhanden; ² Informationen, die innerhalb des genannten Fähigkeitsbereiches im Widerspruch zur formellen Information in dieser Fallvignette stehen. Die Sortierung der Fälle entspricht der Reihenfolge im Fallheft, wie sie in Studie 3 (vgl. Kap. 8) vorgegeben wurde und in nachfolgenden Semestern im Seminar zum Einsatz kam. Dabei wechseln sich Fälle mit niedriger, mittlerer und hoher Aufgabenschwierigkeit ab, damit alle Studierenden möglichst jedes Schwierigkeitsniveau mindestens einmal bearbeiten können. Die Begriffe „Begabung“ sowie „Behinderung“ werden verwendet, da sie gängige Bezeichnungen in Materialien sind, die Lehrkräften in Berlin zur Verfügung stehen (z. B. Senatsverwaltung für Bildung, Jugend und Wissenschaft, 2012) – auch wenn die Verwendung dieser Begriffe nicht uneingeschränkt gutgeheißen wird.

dieser Informationen durch die Studierenden könnte einen Teil der vorliegenden Ergebnisse erklären: Bei Nutzung nichtformeller Informationen ist eine höhere Fehlerrate gegenüber der Nutzung formeller Informationen wahrscheinlich, da erstere teilweise inkongruent zur formellen Information waren. Um die gezielte Untersuchung der Informationsverarbeitung in zukünftigen Studien zu ermöglichen, wurden die Profile der drei Informationsarbeiten im Zuge der Überarbeitung der Fallvignetten präzisiert. Somit entstanden drei Diagnoseinformationen pro Fähigkeitsbereich, die sich noch deutlicher voneinander

abgrenzten. In der finalen Fassung des Fallinventars konnten je drei formelle, informelle und semiformelle Informationen pro Vignette vorgegeben werden. Dabei wurde insbesondere darauf geachtet, dass die zu den formellen Informationen inkongruenten Informationen kein regelmäßiges oder einseitiges Muster ergaben, sodass die Studierenden diese zur Fallbearbeitung irrelevanten Informationen nicht offensichtlich erkennen konnten. In der überarbeiteten Version des Fallinventars waren informelle Informationen insgesamt elfmal inkongruent zu den formellen Informationen und die semiformellen Informationen fünfmal (Verteilung der inkongruenten Informationen für einzelne Fälle vgl. Tab. 6.6). Zudem wurden die Informationen innerhalb der Fallvignetten unterschiedlich positioniert. Informelle und semiformelle Informationen erscheinen in der finalen Version insgesamt jeweils neunmal an erster, siebenmal an zweiter und achtmal an dritter Stelle (insgesamt 24 Platzierungen pro Informationsart, die sich bei acht Fällen mit je drei Fähigkeitsangaben ergeben). Auch die Positionen der formellen Informationen, die als einzige entscheidungsrelevante diagnostische Informationen gelten, wurden variiert. Im gesamten Fallinventar werden sie nun sechsmal an erster, zehnmal an zweiter und achtmal an dritter Stelle präsentiert. Diese Maßnahme diente der Verhinderung eines Lerneffekts hinsichtlich der Position der (nicht) zu nutzenden Informationen innerhalb der 3x3-Matrix der Fallvignetten.

6.6.2.2 Kontrolle des Fallgeschlechts über Fallnamen

In der Pilotversion des Fallinventars wurde das Fallgeschlecht der realen Fälle in den Fallvignetten zunächst beibehalten. In der endgültigen Fassung des Fallinventars sollte jede Fallvignette jedoch in zwei parallelen Varianten, d. h. in einer männlichen sowie in einer weiblichen Version, vorliegen. Durch die Variation des Fallgeschlechts in den Fallvignetten können Dozierende, die mit dem Fallinventar arbeiten, die Studierenden für den potentiell Urteil verzerrenden Einfluss des Schüler*innengeschlechts sensibilisieren. Außerdem ist diese Variation im Rahmen von Erhebungen, in denen das Fallinventar als Untersuchungsinstrument eingesetzt wird, eine Möglichkeit der Kontrolle des Einflusses des Fallgeschlechts auf die Urteilsgenauigkeit (s. a. Kaiser et al., 2015).

Die Variation des Fallgeschlechts erfolgt hier über die *Namen der dargestellten Schüler*innen* sowie die im Beschreibungstext verwendeten Personalpronomen. Bei der Zusammenstellung der Namen der Schüler*innen in wurde berücksichtigt, dass die Namen selbst ebenfalls ein Merkmal sind, das zu Urteilsverzerrungen führen kann: Spezifische Namen werden von Lehrkräften mit Annahmen über die Attraktivität, die Intelligenz oder teilweise sogar mit Verhaltensauffälligkeiten assoziiert (z. B. Rudolph et al., 2007; Kube, 2009). Bei der Auswahl der Namen wurde daher auf Rating-Daten deutscher Vornamen (Rudolph et al., 2007) zurückgegriffen, wobei die beiden einer Fallvignette zugeordneten parallelen Namen hinsichtlich des Attraktivitäts- und Intelligenzratings in etwa vergleichbar sein sollten (vgl. Tab. 6.7). Die über alle Fälle gemittelten Intelligenzratings zu den Namen waren für Jungen- sowie Mädchennamen identisch ($M_{\text{Jungen}} = M_{\text{Mädchen}} = 4.07$, sechsstufige Likertskala). Um diese Vergleichbarkeit herzustellen, wurden die Jungennamen mit den höchsten mittleren Intelligenzratings verwendet, während bei den Mädchennamen die Namen mit den höchsten Werten nicht verwendet wurden (die Intelligenz- und Attraktivitätsratings für Jungennamen waren in der Studie von Rudolph und Kolleg*innen niedriger als für

Mädchennamen). Auf Namen mit einer vergleichsweise hohen Streuung in den Ratings wurde ebenfalls verzichtet.

Tabelle 6.7 Kennwerte für die Auswahl und Zusammenstellung der weiblichen und männlichen Fallnamen

Fallname bei Pilotierung	Fallname in finaler Version		Mittl. geratete Attraktivität		Mittl. geratete Intelligenz		Summe Intelligenz-ratings ²	z-Wert der Intelligenz in Fallvignette ³
	w	m	w	m	w	m		
Louis	Jana	Luca	4.05	3.91	4.16	3.89	7.80	1.07
(N. N.) ¹	Katrin	Felix	3.72	3.93	3.92	3.94	7.87	-.60
Julius	Leonie	Jan	4.22	3.98	3.76	3.99	7.97	.97
Johanna	Laura	Thomas	4.21	4.03	3.68	4.03	8.06	2.33
Jakob	Julia	Maximilian	4.43	4.14	4.05	4.13	8.27	.54
Jonathan	Lena	Christian	4.26	4.15	3.95	4.16	8.31	2.63
(N. N.) ¹	Sophie	Lukas	4.40	4.21	4.16	4.20	8.41	1.30
Louisa	Hanna	Jonas	3.85	4.24	4.03	4.22	8.46	-1.34

Anmerkungen: ¹ Fallvignette war in Pilotversion noch nicht enthalten; ² mittlere Intelligenzratings von weiblichem und männlichem Fallnamen addiert; ³ z-transformierter Wert des Intelligenztests im Fall.

Die Auswahl der Namen für die finale Version des Fallinventars sowie die entsprechenden Kennwerte wurden einer Rating-Studie von Rudolph und Kolleg*innen (2007) entnommen, in der deutsche Vornamen hinsichtlich verschiedener Parameter bewertet werden sollten. In der vorliegenden Studie wurden die Attraktivität sowie die Intelligenz, die im Mittel Personen mit den entsprechenden Namen zugeschrieben werden, herangezogen. Die verwendeten Namen sind der Kategorie „modern/zeitlos“ zugeordnet und weisen assoziierte Attraktivitäts- und Intelligenzratings auf, die im oberen positiven Bereich (1 = gar nicht attraktiv/intelligent, 6 = sehr attraktiv/intelligent) anzusiedeln sind.

Die Zuordnung der Namen zu den Fallvignetten wurde schließlich so vorgenommen, dass eine Konfundierung der verwendeten Namen mit dem jeweiligen Phänomen vermieden wird. Die Intelligenzratings der Namen sollten daher nicht systematisch mit dem im Fall dargestellten Intelligenztestergebnis zusammenhängen. Dies wurde erreicht, indem die in den Fällen simulierten Intelligenztestergebnisse zunächst in z-Werte umgerechnet wurden, um diese in einem zweiten Schritt den mittleren Intelligenzratings der Namen zuzuordnen – und zwar so, dass die Rangfolgen beider Werte unsystematisch kovariierten ($r = -.06$).

6.6.3 Zwischenfazit zu Studie 1

In diesem Kapitel wurde die Konzeption und Entwicklung einer fallbasierten Lerngelegenheit beschrieben, die dazu dient, Lehramtsstudierende auf eine besondere diagnostische Aufgabe vorzubereiten: die Diagnostik von Lernbesonderheiten. Durch das Fallinventar erhalten Studierende im ersten Mastersemester die Gelegenheit, ihr zuvor erworbenes deklaratives Wissen auf realitätsnahe Fälle

anzuwenden (vgl. auch Glogger-Frey & Renkl, 2017). Die Fälle werden als Kurzvignetten dargestellt, in denen Schülerinnen und Schüler mit und ohne Lernbesonderheiten beschrieben werden. Wie in der Praxis üblich erhalten die Studierenden für ihre diagnostischen Aufgaben nicht nur relevante und diagnostisch hochwertige Daten, sondern auch potentiell urteilsverzerrende sowie widersprüchliche Informationen. Ziel der fallbasierten Übung mit dem Fallinventar ist, dass die Studierenden in relativ kurzer Zeit drei Aspekte der professionellen Diagnostik von Lernbesonderheiten kennen- und anwenden lernen:

1. die *Auswahl* von denjenigen Informationen, die im Rahmen eines sorgfältigen und methodisch kontrollierten diagnostischen Prozesses gewonnen wurden (formelle Diagnose; Schrader, 2010);
2. die Abgabe treffsicherer *Diagnosen*, deren Wahrscheinlichkeit durch die Bevorzugung formeller Diagnoseinformationen positiv beeinflusst werden könnte (Behrmann & Van Ophuysen, 2017);
3. die Ableitung passender pädagogischer *Anschlussbehandlungen*, deren Grundlage zutreffende Diagnosen sein sollten (Hesse & Latzko, 2017; Jäger, 2006).

Die Erfahrungen mit den prototypischen Fallvignetten zeigen, dass die Lerngelegenheit grundsätzlich für die Zielgruppe geeignet ist und zu einer Auseinandersetzung mit dem wichtigen Thema der Diagnostik von Lernbesonderheiten anregt. Die Pilotierung ergab, dass die Arbeit mit dem Fallinventar bei der großen Mehrheit der Studierenden auf positive Resonanz stieß. Die einzelnen Fallvignetten waren allesamt verwendbar und – wie intendiert – unterschiedlich schwer zu lösen: Während einige Fälle von einem Großteil der Studierenden korrekt gelöst werden konnten, gab es einen Fall, der nur eine sehr geringe Lösungsrate aufwies. Hier bleibt unklar, ob diese Lösungsrate durch die Fallgestaltung oder das zu diagnostizierende Phänomen in diesem Fall zustande gekommen ist. Ein weiterer Befund, der Erklärung bedarf, war die relative Schwierigkeit der Studierenden, Empfehlungen abzuleiten. Mögliche Erklärungen dafür, dass die Vergabe von Diagnosevermutungen in der Regel leichter fiel, könnten mit der Informationsauswahl oder der Bearbeitungsreihenfolge zusammenhängen. Diese Vermutungen waren jedoch mit den hier vorliegenden Daten nicht überprüfbar.

Nach einer Überarbeitung der pilotierten Fallvignetten sowie der entsprechenden Arbeitsaufträge und Ergänzung um zwei weitere Fallvignetten stehen mit der finalen Version des Fallinventars acht Fallvignetten zur Verfügung, die in weiterführenden Studien eingesetzt werden können. In der zweiten Studie dieser Arbeit (Kap. 7) werden die Fallvignetten als Szenario basierte Umgebung verwendet, um damit Hinweise auf die Bearbeitung des Fallinventars zu erhalten. Dabei stehen die Informationsauswahl sowie die Reihenfolge diagnostischer Schritte im Mittelpunkt des Interesses. In der dritten Studie (Kap. 8) wird evaluiert, inwiefern das diagnostische Fallinventar zur Entwicklung diagnostischer Expertise bei den Studierenden beitragen kann.

7 Diagnostische Produkt- und Prozessvariablen: Untersuchung der Bearbeitung des Fallinventars durch Studierende (Studie 2)

7.1 Hintergrund und Überblick

Im vorausgehenden Kapitel 6 (Studie 1) wurde die Entwicklung des Fallinventars – einer fallbasierten Lerngelegenheit – vorgestellt, die Studierende darin unterstützen soll, pädagogisch-diagnostische Handlungen praxisnah einzuüben. Die *Pilotierung der ersten Version* dieses Fallinventars zeigte, dass die entwickelten Fallvignetten für den Einsatz im Masterstudium geeignet sind und weitere Vignetten nach den Konstruktionsprinzipien (vgl. Abschnitt 6.2) konzipiert werden können. Die Ergebnisse der Pilotstudie werfen zudem Forschungsfragen auf, die sich auf die Vorgehensweise der Studierenden bei der Bearbeitung des Fallinventars beziehen. In dieser ersten Folgestudie werden Zusammenhänge zwischen korrekten Lösungen bei pädagogisch-diagnostischen Entscheidungen (Diagnosen und Empfehlungen) sowie Aspekten der Durchführung des diagnostischen Prozesses untersucht.

Im Rahmen der vorliegenden Studie dient das Fallinventar daher als experimentelle Umgebung, mit deren Hilfe die *Erfassung diagnostischer Produkte und Prozessaspekte* realisiert wird (vgl. Ansatz Schülerinventar: Kaiser et al., 2015). Die Untersuchung der Prozessaspekte liefert erstens Hinweise auf die Frage, inwiefern die Abfolge von Diagnose- und Empfehlungsentscheidungen sich auf die Trefferquote bei beiden Entscheidungen auswirkt. Die Ergebnisse dieser Studie zeigen zweitens, inwiefern Lehramtsstudierende bereits in diesem Ausbildungsabschnitt formelle diagnostische Informationen nutzen und inwiefern die Nutzung die Trefferquoten bei den beiden Entscheidungen begünstigen kann. Bevor das Forschungsvorhaben konkret beschrieben wird, wird zunächst der konkrete theoretische Hintergrund für die Entwicklung der beiden Forschungsfragen kurz zusammengefasst.

7.2 Der diagnostische Prozess im Forschungsinteresse

7.2.1 Betrachtung des diagnostischen Prozesses als Desiderat

Dass Lehrkräfte zutreffende Diagnosen vergeben können, ist eine zentrale Facette diagnostischer Kompetenz. Die Fähigkeit, zutreffende Diagnosen zu erstellen, wird als Urteilsgenauigkeit bezeichnet und häufig mit diagnostischer Kompetenz gleichgesetzt (Schrader, 2010; vgl. Kap. 2.1). Studien zur Urteilsgenauigkeit belegen, dass die Güte diagnostischer Entscheidungen je nach Urteilsgegenstand und Beurteilungsmethode starken inter- wie intraindividuellen Schwankungen unterworfen ist (z. B. Spinath, 2005). Darüber hinaus konnte eine Vielzahl von Variablen identifiziert werden, die das Ausmaß der Urteilsgenauigkeit beeinflussen (für eine Übersicht siehe Südkamp et al., 2008). Um diese Varianz hinsichtlich der Urteilsgenauigkeit aufzuklären und Hinweise darauf zu erhalten, auf welche Weise diagnostische Urteile hoher Güte gefördert werden können, muss der Prozess der Urteilsbildung in den Fokus der Betrachtungen rücken (vgl. Kap. 2.2). Im diagnostischen Prozess werden diagnostische Daten gewonnen, deren Qualität maßgeblich für die Qualität der Diagnosen ist – und der Prozess bestimmt darüber, welche Qualität die diagnostischen Daten aufweisen (Behrmann & Van Ophuysen, 2017). Die

Gestaltung des diagnostischen Prozesses selbst kann demnach als ein zum Qualitätsmerkmal von Diagnosen und die Fähigkeit zur professionellen Gestaltung des diagnostischen Prozesses damit als eine weitere Facette diagnostischer Kompetenz betrachtet werden (vgl. Kap. 2.2.2).

Aktuelle Ansätze zur Modellierung diagnostischer Kompetenz versuchen entsprechend, Aspekte des diagnostischen Prozesses bei der Analyse diagnostischer Entscheidungen zu integrieren (Südkamp & Praetorius, 2017). Nachfolgend werden zwei Aspekte der Gestaltung des diagnostischen Prozesses betrachtet, die sich auf die Qualität der diagnostischen Entscheidungen auswirken können: erstens die Abfolge der operativen Schritte des diagnostischen Prozesses, zweitens Aspekte der Verarbeitung diagnostischer Informationen. Zuerst wird die *Abfolge diagnostischer Schritte im diagnostischen Prozess* als Faktor für die Güte diagnostischer Entscheidungen beleuchtet (vgl. Kap. 1.2.4).

7.2.1.1 Normative Vorstellungen vom diagnostischen Prozess und Gegenentwürfe

Die Annahme, dass die Güte der diagnostischen Entscheidungen von der Ausführung des Prozesses abhängt, ist normativen Modellen des diagnostischen Prozesses immanent: Sie geben eine feste Reihenfolge operativer Schritte innerhalb des diagnostischen Prozesses vor, deren Zielpunkt die diagnostische Urteilsbildung (z. B. Jäger, 2006; Klug et al., 2013, 2016) bzw. darüber hinausgehend auch die Ableitung und Durchführung pädagogischer Anschlusshandlungen bildet (Hesse & Latzko, 2017; Mandinach & Gummer, 2016; vgl. Kap. 2.2.3). Insbesondere schwerwiegende diagnostische Entscheidungen, die durch eine hohe Verbindlichkeit ausgezeichnet sind, erfordern dieses regelgeleitete Vorgehen (Karst et al., 2017). Die explizite Einhaltung einer festen Schrittfolge sowie bestimmter methodischer Regeln mündet in einer formellen Diagnose (Schrader, 2009, 2010). Formelle Diagnosen sind mit höchster Sorgfalt erstellt und sollten daher auch zu den bestmöglichen diagnostischen Entscheidungen führen: Die formelle Prozedur unter Hinzuziehung objektiver Verfahren stellt der diagnostizierenden Person qualitativ hochwertige diagnostische Daten zur Verfügung, auf deren Grundlage qualitativ hochwertige diagnostische Entscheidungen getroffen werden können (vgl. Behrmann & Van Ophuysen, 2017). Auch um individuelle Fördermaßnahmen oder Maßnahmen der Unterrichtsentwicklung abzuleiten, ist die Nutzung diagnostischer Daten hoher Güte anzustreben (z. B. Helmke & Hosenfeld, 2005; Mandinach & Gummer, 2016).

Dass Lehrkräfte in der Lage sind, eine formelle diagnostische Prozedur auszuführen, kann demnach als eine grundlegende diagnostische Fähigkeit angesehen werden. Aktuelle Ansätze zur Modellierung des diagnostischen Prozesses im deutschsprachigen Raum betonen allerdings – im Gegensatz zu den normativen Prozessmodellen und an frühere Diskussionen zur normativen Gestaltung diagnostischer Prozesse erinnernd (vgl. Ingenkamp & Lissmann, 2008) –, dass die Reihenfolge und Umfänglichkeit der einzelnen Schritte des diagnostischen Prozess in Abhängigkeit von den jeweiligen diagnostischen Situationen variieren können und keinesfalls immer gleich gestaltet sein müssen (z. B. Herppich et al., 2017, 2018; vgl. Kap. 1.2.4.1). In Anlehnung an Zwei-Wege-Modelle sozialer Urteilsbildung können Überlegungen zur Auswahl diagnostischer Prozeduren angestellt werden, die zur diagnostischen Situation passen (Hetmanek & Van Gog, 2017): Bei nicht planbaren Situationen im laufenden Unterrichtsgeschehen

ist Schnelligkeit oft wichtiger und nützlicher als Akkuratheit, sodass hier Typ-1-Diagnosen (heuristisch, automatisiert, unbewusst) angemessen sind. In planbaren Situationen, wie z. B. Lernstandserhebungen, ist andererseits Akkuratheit gegenüber Schnelligkeit zu bevorzugen, sodass hier Typ-2-Diagnosen (analytisch, systematisch, bewusst kontrolliert) anzustreben sind. Ob die diagnostizierende Person einen für die Situation und das Diagnoseziel angemessenen Diagnoseweg auswählt, ist im Rahmen dieses Zwei-Wege-Ansatzes ein Zeichen diagnostischer Kompetenz (vgl. auch Böhmer et al., 2017). Erste empirische Untersuchungen zur diagnostischen Kompetenz bei Medizinstudierenden geben Hinweise darauf, dass die beiden grundlegenden Diagnosewege differenzielle Beiträge zu den diagnostischen Entscheidungen leisten (vgl. Kap. 4.3). Die systematische Untersuchung der Güte diagnostischer Urteile in Abhängigkeit von der jeweiligen diagnostischen Prozedur – im Kontext verschiedener diagnostischer Anlässe – steht derzeit allerdings noch aus.

7.2.1.2 Die Informationsverarbeitung als Aspekt des Prozesses

Neben der Passung zum diagnostischen Anlass und Ziel spielt bei der Auswahl der jeweiligen diagnostischen Prozedur auch eine Rolle, in welchem Ausmaß der diagnostizierenden Person kognitive Ressourcen zur Verfügung stehen – denn die beiden skizzierten Diagnosewege bedeuten ein unterschiedliches Ausmaß an kognitivem Aufwand. Die kognitive Fähigkeit, diagnostische Informationen zu verarbeiten und für diagnostische Entscheidungen zu nutzen, ist ein zweiter Aspekt für die Gestaltung des diagnostischen Prozesses. Hierbei ist weniger die Schrittfolge innerhalb des Prozesses im Fokus, sondern die *Auswahl und Verarbeitung diagnostischer Informationen* innerhalb des Beurteilungsprozesses (vgl. Kap. 2.2.1). Die Art der Informationserfassung und –verarbeitung kann die Güte des Urteils entscheidend beeinflussen, weil sie bestimmend ist für die Qualität der diagnostischen Daten (Vier-Komponenten-Modell der Diagnosequalität: Behrmann & Van Ophuysen, 2017).

Für diese Auswahl- und Verarbeitungsprozesse diagnostischer Informationen haben Lehrkräfte jedoch nicht in gleichem Maße die *notwendigen kognitiven Voraussetzungen* (Linninger et al., 2015). Vielmehr unterscheiden sich Lehrkräfte in der Praxis teilweise deutlich hinsichtlich der Beachtung und Nutzung diagnostischer Informationen, wenn sie diagnostische Urteile fällen (für Leistungsbeurteilung: Kaiser et al., 2015; für Lernprozessdiagnostik: Klug et al., 2013). Ein Faktor, der die unterschiedlichen Urteilsfähigkeiten von Lehrkräften erklären kann, ist die Berufserfahrung: Erfahrene Lehrkräfte sind gegenüber Noviz*innen weniger anfällig für urteilsverzerrende Faktoren. So sind sie besser in der Lage, relevante von irrelevanten Informationen zu unterscheiden und irrelevante Informationen in ihrer Entscheidungsfindung auszublenden, sofern der diagnostische Anlass es gebietet (z. B. Übertrittsempfehlung vs. Eindrucksbildung, vgl. Krolak-Schwerdt et al., 2009, 2013).

Allgemein scheint die Professionalisierungsphase, in der sich Lehrkräfte befinden, ein Prädiktor für die Fähigkeit zur Nutzung diagnostischer Informationen zu sein (Klug et al., 2016): Lehramtsstudierende, Referendar*innen und praktisch tätige Lehrkräfte weisen sowohl ein unterschiedliches Profil kognitiver Voraussetzungen (gemessen in einem Wissenstest) als auch unterschiedliche Fähigkeiten bei der Bewältigung der verschiedenen Phasen des diagnostischen Prozesses (gemessen in einem Szenariotest, s. a.

7.2.2.2) auf. Dabei zeigen insbesondere Lehramtsstudierende in allen Phasen des diagnostischen Prozesses geringere Leistungen als Referendar*innen und erfahrene Lehrkräfte. Die gefundenen Unterschiede sind besonders eklatant bei der methodisch begründeten Auswahl der Instrumente sowie bei der Interpretation und Nutzung der diagnostischen Informationen zur Entscheidungsfindung. Offenbar haben Studierende, die bislang wenig praktische Erfahrungen mit der Durchführung des diagnostischen Prozesses gemacht haben, genau in dieser Phase einen Bedarf an gezielter Unterstützung. Studien, die die Auswirkungen von Interventionen zur Verbesserung des diagnostischen Prozesses auf die Güte diagnostischer Urteile untersuchen, stehen allerdings noch aus.

7.2.2 Möglichkeiten der Erfassung diagnostischer Produkt- und Prozessvariablen

Die vorausgehenden Ausführungen verdeutlichen, dass die Analyse diagnostischer Produkte (Diagnosen, Entscheidungen über Fördermaßnahmen) davon profitieren kann, wenn sie in Zusammenhang mit Aspekten des diagnostischen Prozesses (Schrittfolge, Informationsverarbeitung) betrachtet werden. Systematische Untersuchungen dazu, inwiefern Prozessmerkmale sich auf die diagnostischen Urteile und Entscheidungen auswirken, sind jedoch ein Desiderat (vgl. Behrmann & Van Ophuysen, 2017). Eine Voraussetzung, um diese Forschungsfrage zu beantworten, ist das *Vorhandensein von Erhebungsinstrumenten* – und das ist insbesondere für die Erfassung von Prozessaspekten keine triviale Voraussetzung. Nachfolgend werden daher zum einen Verfahren vorgestellt, die sich zur Erfassung diagnostischer Produkte eignen, und die sich bereits in empirischen Studien bewährt haben. Diese Ausführungen werden ergänzt um Ansätze, die zur Erfassung von diagnostischen Prozessaspekten vielversprechend sind. Die Überlegungen orientieren sich dabei an der Leitidee, dass sie für die Diagnostik von Lernbesonderheiten geeignet sein sollten, da dieser Diagnoseanlass im Fokus der vorliegenden Arbeit steht (vgl. Kap. 6).

7.2.2.1 Erfassung diagnostischer Produkte

Zur Erfassung der Güte diagnostischer Produkte wird in der Regel die *Urteilsgenauigkeit* herangezogen, für die drei etablierte Maße zur Verfügung stehen (vgl. Kap. 2.1). Dabei werden diagnostische Einschätzungen von Lehrkräften mit den tatsächlichen Merkmalsausprägungen der diagnostizierten Schüler*innen verglichen. Diese Operationalisierungen diagnostischer Kompetenz haben aus messtheoretischer Sicht zwei große Vorteile: Erstens handelt es sich dabei um eng umgrenzte und gut messbare Ergebnismaße, die zweitens außerhalb des unterrichtlichen Handelns operationalisiert werden können. Auf diese Weise ist es möglich, das diagnostische Urteil einer Lehrkraft hinsichtlich seiner Güte einzuschätzen, ohne dass dies mit pädagogischen Handlungen vermischt wird – was gleichzeitig einen Nachteil für die ökologische Validität darstellen kann (Praetorius et al., 2017). Meines Erachtens trifft diese Kritik jedoch nur für innerunterrichtliche Diagnoseanlässe zu, bei denen diagnostische und pädagogische Situation eng miteinander verwoben sind. Bei außerunterrichtlichen Diagnoseanlässen wiederum ist die diagnostische Situation hingegen per definitionem von der pädagogischen Situation getrennt, sodass dieses Argument hier weniger greift.

Die *Feststellung von Lernbesonderheiten* ist ein solcher außerunterrichtlicher Diagnoseanlass, der daher von dieser Kritik unberührt bleibt. Allerdings muss für diesen Diagnoseanlass die Operationalisierung von Urteilsgenauigkeit überdacht werden: Hier ist ein bloßer Abgleich von tatsächlichen und prognostizierten Ergebnissen von Schulleistungstests oder kognitiven Fähigkeitstests nicht ausreichend, da die Diagnostik von Lernbesonderheiten erfordert, dass verschiedene Fähigkeitsbereiche von Schüler*innen gemeinsam betrachtet werden (vgl. 6.4 prototypische Fähigkeitsprofile). Die verschiedenen testdiagnostischen Informationen aus mindestens drei Fähigkeitsbereichen müssen, zusammen mit weiteren diagnostischen Informationen, zu einer Diagnose integriert werden – dies ist der entscheidende Schritt im Rahmen psychologischer Diagnostik (vgl. Jäger, 2006). Diese Diagnose ist im Ergebnis kategorial und kann nur zwei Ausprägungen annehmen (Diagnose positiv / negativ). Dementsprechend ist hier – ähnlich wie bei Tests mit dichotomen Ergebnissen – ausschlaggebend, dass positiv diagnostizierte Personen auch tatsächlich das Phänomen aufweisen und umgekehrt Personen, die das Phänomen nicht aufweisen, auch einen negativen Befund haben (Faller, 2005).

7.2.2.2 Erfassung diagnostischer Prozessaspekte

Zur Analyse des diagnostischen Prozesses stehen bislang keine etablierten Methoden zur Verfügung, die konkrete diagnostische Prozeduren und Entscheidungsprozesse im Rahmen des tatsächlichen Unterrichtsgeschehens erfassen. Eine Herausforderung hierbei ist, dass es einerseits nahezu unmöglich ist, schnelle Verarbeitungsprozesse objektiv und valide abzubilden. Andererseits vermischen sich in Felduntersuchungen innerunterrichtlicher Urteilsprozesse typischerweise diagnostische und pädagogische Handlungen, was zulasten der internen Validität geht (vgl. Praetorius et al., 2017). Um Prozessaspekte erfassen zu können, scheint daher erstens eine *Trennung von diagnostischer und pädagogischer Situation* unumgänglich, und zweitens müssen insbesondere für innerunterrichtliche Diagnoseanlässe Einschränkungen der ökologischen Validität zugunsten interner Validität akzeptiert werden. Diese Einschränkungen ökologischer Validität sind in der Form für außerunterrichtliche Diagnoseanlässe indes nicht anzunehmen, wie bereits erläutert wurde.

Eine Möglichkeit, um die zur Messung notwendige Trennung diagnostischer und pädagogischer Situationen zu operationalisieren, können sog. *Szenario basierte Testverfahren* sein. Diese sind eine Sonderform pädagogisch-psychologischer Wissenstests, welche i. d. R. deklaratives Wissen über theoretische und methodische Grundlagen pädagogischer Diagnostik oder von Beurteilungsprozessen abfragen (Voss et al., 2011) und auch für Studierende im Lehramt eingesetzt werden können (z. B. Linninger et al., 2015). Die Wissensfragen, die oftmals in Form von Multiple-Choice-Items oder Kurz-Antwort-Formaten gestellt werden, werden hierbei in bestimmte Unterrichtssituationen eingebettet und können somit zu einer relativ handlungsnahen Gestaltung von Wissenstests beitragen, die damit auch prozedurales oder konditionales Wissen abfragen können (Lenke et al., 2015). Die diagnostischen Situationen können in Form schriftlicher oder videografiertter Fallvignetten beschrieben werden, und dadurch Prozesse der diagnostischen Informationsverarbeitung und Entscheidung abbilden. Gegenüber klassischen Wissenstests bieten diese künstlichen Umgebungen somit die Möglichkeit, die ökologische

Validität der Messung diagnostischer Kompetenz zu erhöhen, ohne die innere Validität des Designs zu gefährden (vgl. Praetorius et al., 2017).

Zwei Szenario basierte Umgebungen, die erfolgreich zur Untersuchung von diagnostischen Prozessvariablen eingesetzt wurden, sollen nachfolgend kurz skizziert werden. Ein Beispiel für eine *stark komplexitätsreduzierte Umgebung* ist das Schülerinventar (Kaiser et al., 2015; vgl. Kap. 6.4), das zur Untersuchung der Beeinflussbarkeit von Leistungsurteilen durch Schüler*innenmerkmale eingesetzt wurde. In dieser Umgebung wurden relevante (z. B. Punkte in einer Mathematikklassenarbeit) und irrelevante Informationen (z. B. Ergebnis in einem Intelligenztest) in kurzen auf Fallvignetten zusammengestellt, auf deren Basis Referendar*innen eine Beurteilung der mathematischen Leistungen vorgenommen haben. Über die experimentelle Variation potentiell urteilsverzerrender Faktoren konnten hier einzelne Merkmale, die die Urteilsgenauigkeit reduzierten, identifiziert werden (z. B. Intelligenz, Deutschnote). Unklar bleibt dabei, wie die Beurteilung auf Basis der vorgegebenen Informationen konkret zustande gekommen ist und an welcher Stelle des diagnostischen Prozesses die urteilsverzerrenden Einflüsse gewirkt haben.

Die Beantwortung prozessbezogener Fragestellungen ist im Rahmen von Szenario basierten Testansätzen möglich, die *komplexere diagnostische Situationen* sowie verschiedene Phasen des diagnostischen Prozesses abbilden. Für die Diagnostik von Lernprozessen (statt Leistungen) entwickelten Julia Klug und Kolleg*innen (Klug et al., 2013) einen solchen Szenariotest, in dem der Fall eines Schülers mit Lernschwierigkeiten geschildert wird, deren Ursachen die Lehrkraft finden soll. Konkrete Arbeitsaufträge orientieren sich dabei an verschiedenen Phasen des diagnostischen Prozesses, der die methodische Planung („präaktionale Phase“), Sammlung und Auswertung diagnostischer Daten („aktionale Phase“) sowie deren Interpretation und Ableitung einer pädagogischen Anschlusshandlung („postaktionale Phase“) beinhaltet. Statt der tatsächlichen Handlungen wurden die Handlungsabsichten der Lehrkraft erfasst und mit Punkten bewertet. Die Punktesumme der einzelnen Arbeitsaufträge konnte im Rahmen einer Studie zur Untersuchung der Verbesserung diagnostischer Kompetenzen durch Tagebücher als Evaluationsmaß genutzt werden. Mithilfe dieses Tests war es darüber hinaus möglich, Unterschiede zwischen Lehrkräften in verschiedenen Ausbildungsphasen (Studium, Referendariat, Berufspraxis) hinsichtlich verschiedener Prozessphasen aufzuzeigen (Klug et al., 2016).

7.3 Das diagnostische Fallinventar als Szenario basierte Umgebung

Die Befunde zu beiden vorgestellten Instrumenten zeigen, dass Szenario basierte Umgebungen vielversprechende Ansätze sind, um diagnostische Vorgehensweisen und Fähigkeiten zu erfassen. In der vorliegenden Studie kam daher das im vorausgehenden Kapitel vorgestellte diagnostische Fallinventar als Szenario basierte Umgebung zum Einsatz, um diagnostische Fähigkeiten von Lehramtsstudierenden im *Kontext der Diagnostik von Lernbesonderheiten* zu untersuchen. Da für diesen Diagnoseanlass m. W. bislang kein spezifisches Instrument oder empirische Befunde zu Urteilsprozessen und –fähigkeiten von Lehramtsstudierenden vorliegen, widmet sich diese Studie der explorativen Untersuchung von

diagnostischen Prozessaspekten im Zusammenhang mit Produkten des diagnostischen Prozesses. Angesichts der engen Verwobenheit von Diagnostik und Intervention im Kontext der pädagogischen Diagnostik (vgl. Hesse & Latzko, 2017; Jürgens & Lissmann, 2015), werden neben Diagnosen auch Empfehlungen für Fördermaßnahmen als zentrale diagnostische Produkte betrachtet (vgl. Abb. 1.3 Modell des diagnostischen Prozesses).

7.3.1 Fragestellungen und Hypothesen

Der Einsatz des diagnostischen Fallinventars als Szenario basierte Umgebung dient der Untersuchung zweier Kernfragen: Erstens sollte auf diese Weise überprüft werden, inwiefern die nach der Pilotierung überarbeiteten Fallvignetten (vgl. Kap. 6.6) für die Lerngelegenheit geeignet sind (Abschnitt 7.3.1.1). Zweitens sollte die Untersuchung mit dem Fallinventar Hinweise darauf geben, inwiefern die Studierenden in dieser Umgebung in der Lage sind, korrekte pädagogisch-diagnostische Entscheidungen bei der Diagnostik von Lernbesonderheiten zu treffen (Abschnitt 7.3.1.2). Dabei war nicht nur die Güte dieser diagnostischen Produkte (Diagnosen und Empfehlungen) im Interesse, sondern insbesondere auch die Frage, inwieweit sich das Vorgehen der Studierenden bei der Nutzung und Verarbeitung diagnostischer Informationen im Fallinventar auf die Güte ihrer Entscheidungen auswirkt. Die Untersuchung sollte somit Hinweise darauf geben, welche Schwierigkeiten Lehramtsstudierende bei der Diagnostik von Lernbesonderheiten zeigen und ob sich ihr diagnostisches Vorgehen im Rahmen der Lerngelegenheit optimieren lässt.

7.3.1.1 Eignung der Fallvignetten für die Lerngelegenheit

Fragestellung 1a. Rezeption der Lerngelegenheit

Die Untersuchung der Eignung der Fallvignetten für die Lerngelegenheit erfolgte auf zwei Ebenen. Zunächst sollte erfasst werden, ob die *Lerngelegenheit als Ganzes* auf positive Resonanz bei den Studierenden stieß. Dass Rezipient*innen eine Lerngelegenheit als relevant und nützlich empfinden, ist Angebots-Nutzungs-Modellen zur Kompetenzentwicklung von Lehrkräften zufolge eine zentrale Voraussetzung dafür, dass die Inhalte aufgenommen und Lernprozesse damit überhaupt erst ermöglicht werden (z. B. Lipowsky, 2014; Kittel & Rollett, 2018). Ziel war es, dass die Studierenden das Fallinventar als Lerngelegenheit wahrnehmen, die für die Bewältigung der Abschlussprüfung im Modul „Pädagogische Diagnostik“ sowie für die spätere Berufspraxis hilfreich ist.

Fragestellung 1b. Eignung der Fallvignetten

Zweitens sollten die *einzelnen Fallvignetten* hinsichtlich ihrer Eignung für die Lerngelegenheit untersucht werden. Hierzu wurde einerseits der Anteil der Studierenden, die korrekte Diagnose- und Empfehlungsentscheidungen trafen, zur Bemessung des Schwierigkeitsgrades der einzelnen Fallvignetten herangezogen. In der Pilotierung konnten die Studierenden im Mittel je nach Fall 13 – 100 % der Diagnosen sowie 47 – 80 % der Empfehlungen korrekt auswählen. Für die finale Version des Fallinventars waren einige schwierige Items erwünscht, um die Studierenden kognitiv anzuregen und

somit konzeptuelle Umstrukturierungen grundsätzlich zu ermöglichen (Stern et al., 2016). Ein Großteil der Items sollte jedoch von der Mehrheit der Studierenden zu lösen sein, um Erfolgserlebnisse und damit das eigenständige Lernen anzuregen (vgl. Schwarzer & Warner, 2014). Zu diesem Zwecke wurde außerdem angestrebt, dass die Studierenden sich beim Treffen ihrer pädagogisch-diagnostischen Entscheidungen im Mittel eher sicher fühlten.

7.3.1.2 Untersuchung des diagnostischen Prozesses der Studierenden

Fragestellung 2a. Güte der diagnostischen Produkte

Die Untersuchung des diagnostischen Prozesses der Studierenden erfolgte entlang dreier Leitfragen. Erstens stand die *Korrektheit der diagnostischen Entscheidungen* im Fokus, deren Ergebnisse (Produkte) Diagnosen und Empfehlungen waren. Für die Feststellung von Lernbesonderheiten ist die Vergabe von korrekten Diagnosen ein zentrales Gütekriterium, da korrekte Diagnosen die Grundlage dafür sind, dass passende Fördermaßnahmen eingeleitet werden. Ausgehend von Ergebnissen einer Szenario basierten Studie zur Lernprozessdiagnostik, die Herausforderungen von Studierenden ohne Praxiserfahrungen bei der Ableitung von Anschlusshandlungen auf Grundlage der diagnostischen Daten illustriert (Klug et al., 2016), wurden auch beim Diagnostizieren von Lernbesonderheiten im Rahmen des Fallinventars besondere Schwierigkeiten bei der Empfehlungsentscheidung erwartet. Bereits bei der Pilotierung des Fallinventars ergaben sich Hinweise darauf, dass die Abgabe von Empfehlungen der Mehrheit der Studierenden schwerer fällt als das Vergeben von Diagnosen. Ergänzend wurde in der vorliegenden Studie die *subjektive Sicherheit* der Studierenden beim Treffen der Diagnose- und Empfehlungsentscheidungen analysiert. Diese sollte tendenziell mit der tatsächlichen Korrektheit der Entscheidungen übereinstimmen, da eine realistische Selbsteinschätzung – neben tatsächlichen Fähigkeiten – als Zeichen von Kompetenz von Lehramtsstudierenden interpretiert werden kann (z. B. Schladitz et al., 2015).

Fragestellung 2b. Abfolge der diagnostischen Entscheidungen

Die verbleibenden zwei Leitfragen widmeten sich dem *diagnostischen Prozess im engeren Sinne*. Ziel der Untersuchungen war es zu explorieren, inwiefern Studierende ohne nennenswerte Praxiserfahrungen bereits ein systematisches diagnostisches Vorgehen zeigen, das für die Diagnostik von Lernbesonderheiten als verbindlicher Diagnoseanlass unbedingt wünschenswert ist. Dabei wurde einerseits untersucht, ob die Studierenden von einer bestimmten *Abfolge der diagnostischen Entscheidungen* im diagnostischen Prozess profitieren konnten. Hierzu wurden die Studierenden zwei Experimentalgruppen zugeteilt, die ihre Entscheidungen entweder in einer normativen (Diagnose vor Empfehlung) oder aber in einer nicht-normativen (Empfehlung vor Diagnose) Abfolge trafen. Von praktisch tätigen Lehrkräften wird ein professionelles Vorgehen derart erwartet, dass Fördermaßnahmen auf Grundlage von Diagnosen (diagnostische Prozessmodelle: z. B. Hesse & Latzko, 2017; Herppich et al., 2017) bzw. der Interpretation und Integration diagnostischer Daten (datengestützte Unterrichtsentwicklung: Helmke, 2004; Mandinach & Gummer, 2016) abgeleitet werden. Unklar ist jedoch, inwiefern bereits Lehramtsstudierende von diesem professionellem Vorgehen, das von Lehrkräften in der Praxis erwartet wird, profitieren. Im

Rahmen der Pilotierung des Fallinventars war zu beobachten, dass die Studierenden bei der Vergabe von Diagnosen treffsicherer waren als bei der Abgabe von Empfehlungen – letztere wurden dort allerdings stets vor der Diagnoseentscheidung abgegeben, was sich möglicherweise nachteilig auf die Trefferquoten bei Empfehlungen ausgewirkt hat. In der vorliegenden Studie wurde daher noch einmal explizit untersucht, ob die normative Reihenfolge von Diagnose- und Empfehlungsentscheidungen sich auf die Trefferquoten bei den diagnostischen Entscheidungen bei Lehramtsstudierenden als vorteilhaft erweist.

Fragestellung 2c. Nutzung formeller Diagnoseinformationen

Die zweite Untersuchungsfrage hinsichtlich des diagnostischen Prozesses bezieht sich auf die Informationsverarbeitung beim Treffen pädagogisch-diagnostischer Entscheidungen. Erste Studien, die Aspekte der Informationsverarbeitung in Zusammenhang mit der Güte diagnostischer Entscheidungen untersuchen, zeigen einerseits, dass die Genauigkeit von Urteilen durch Vorgabe irrelevanter Informationen reduziert werden kann (Kaiser et al., 2015). Andererseits deuten sie darauf hin, dass die Verarbeitung diagnostischer Informationen von Noviz*innen anfälliger für urteilsverzerrende Informationen ist als die von Expert*innen (Krolak-Schwerdt et al., 2009, 2013). Im Fallinventar wurden den Studierenden formelle und nicht-formelle Diagnoseinformationen vorgegeben. Dabei wurde exploriert, inwiefern die Studierenden formelle Informationen als relevante Informationen nutzen und inwieweit diese Nutzung zur korrekten Vergabe von Diagnosen und Empfehlungen beitragen kann. Es wurde erwartet, dass eine vermehrte Nutzung nicht-formeller Informationen umgekehrt mit geringeren Trefferquoten bei den jeweiligen Entscheidungen einhergeht.

7.3.2 Methoden

7.3.2.1 Stichprobe

Die Stichprobe der Studierenden wurde im Wintersemester 2016/2017 aus *Seminaren zur Pädagogischen Diagnostik* rekrutiert, die regulär im ersten Master-Semester des Lehramtsstudiums an der Freien Universität Berlin stattfinden. Das Seminar ergänzt eine Vorlesung, die mit einer Klausur als Modulabschlussprüfung abgeschlossen wird. Zum Zeitpunkt der Erhebung (Januar 2017) hatten die Seminarteilnehmer*innen bereits ca. 2/3 des Seminars absolviert und sollten über grundlegende Kenntnisse zu standardisierten Tests und testtheoretischen Grundlagen, zu Ausprägungsgraden kognitiver Fähigkeiten sowie über Lernstörungen verfügen. Dieses Grundlagenwissen ist zur Bearbeitung des Fallinventars notwendig, da es dort im Rahmen der fallbasierten Arbeit zur Anwendung kommt (vgl. Kap. 6.3).

Zur Teilnahme gebeten wurden zu diesem Zeitpunkt neben Studierenden aus dem Modul „Pädagogische Diagnostik“ auch Teilnehmer*innen von Begleitseminaren zu diagnostischen Lernforschungsprojekten im Praxissemester. Für 121 Studierende war durch fehlende oder fehlerhafte Angaben pseudonymisierter Codes im Rahmen der Online-Umfrage-Plattform (vgl. Abschnitt 7.2.3.4) keine eindeutige Zuordnung zum Modul möglich, sodass diese in der vorliegenden Studie nicht berücksichtigt werden konnten. Die relevante Untersuchungsstichprobe von $N = 128$ Studierenden (entspricht 48.5 % der Gesamtstichprobe)

wurden seminarübergreifend zufällig zwei Experimentalbedingungen zugewiesen, die sich in der Bearbeitungsreihenfolge unterschieden (Fragestellung 2a): Etwa die Hälfte der Studierenden vergab für eine Auswahl von Fallvignetten zuerst Diagnosen und anschließend Empfehlungen (normative Abfolge, $N = 55$), und die andere Hälfte vergab Empfehlungen vor Diagnosen (nicht-normative Abfolge, $N = 62$).

Von denjenigen, die an der Befragung teilnahmen, machten etwa ein Fünftel der Studierenden keine Angaben zu demografischen und studienbezogenen Fragen. Diese Personen wurden trotz teilweise unvollständiger Datensätze nicht von der Untersuchung ausgeschlossen. Anhand der zur Verfügung stehenden Angaben kann die Stichprobe folgendermaßen beschrieben werden: Die Studienteilnehmer*innen waren zu zwei Dritteln weiblich ($N = 67$) und zu einem Drittel männlich ($N = 29$; 3 ohne Zuordnung, 29 ohne Angabe). Das durchschnittliche Alter der Teilnehmer*innen lag bei $M = 27.6$ Jahren ($SD = 5.6$) und die durchschnittliche Semesterzahl, die bisher absolviert wurde, lag bei $M = 7.7$ ($SD = 2.7$) Studiensemestern. Der Großteil der Teilnehmer*innen studierte mit Lehramtsoption für weiterführende Schulen ($N = 81$) und knapp ein Fünftel der Teilnehmer*innen studierte Grundschulpädagogik ($N = 17$; 30 ohne Angabe). Der Anteil der Nicht-Deutsch-Muttersprachler*innen lag bei 21.9 %. Der Großteil der Teilnehmer*innen hatte bisher bislang noch gar keine ($N = 66$) bzw. nur wenig Erfahrungen mit standardisierten Tests ($N = 24$; 29 ohne Angabe).

7.3.2.2 Instrument

Zum Einsatz kam das diagnostische Fallinventar in der *überarbeiteten Fassung* (vgl. Kap. 6.6). Diese Version enthielt acht Fallvignetten, die realitätsnahe Szenarien von Schüler*innen verschiedener Altersstufen mit (5) und ohne (3) Lernbesonderheiten schildern. Die Präsentation der Fallvignetten erfolgte hier wieder in Form von ausgedruckten Fallheften, bei der je ein Fall auf einem Fallblatt dargestellt wurde. Die einzelnen Fallvignetten enthielten allgemeine Informationen zum jeweiligen Schüler oder zur jeweiligen Schülerin (Name, Klassenstufe, Schulform) sowie fähigkeitsbezogene Informationen in drei schulisch relevanten Bereichen (Lesefähigkeiten, Zahlen und Operationen, kognitive Fähigkeiten). Die fähigkeitsbezogenen Informationen wurden in jeweils drei Varianten (formell, informell, semiformell) angegeben, die eine unterschiedliche Güte diagnostischer Informationen widerspiegeln. Auf diese Weise entstand eine 3x3-Matrix, innerhalb derer die Position der diagnostischen Informationsgüte variiert wurde (vgl. Abb. 6.2).

Um etwaige verzerrende Einflüsse durch die Fälle selbst zu kontrollieren, wurde die *Reihenfolge der Fallvignetten* bei der Zusammenstellung der Fallhefte mithilfe der Methode des lateinischen Quadrates variiert, sodass sich acht Reihenfolge-Versionen ergaben (vgl. Anhang 2A). Auf diese Weise wurde sichergestellt, dass jede Fallvignette auf jeder Position (Fall 1 – Fall 8) gleich häufig vorgegeben wird. Für jede dieser Fallheftversionen wurden wiederum zwei Varianten erstellt, die hinsichtlich der Zusammensetzung weiblicher und männlicher Fälle ausbalanciert wurden. Die *Häufigkeit der Fallgeschlechter* wurde durch die Nutzung paralleler Fallnamen (vgl. Tab. 6.7) für die Fallhefte kontrolliert (je vier weibliche und vier männliche Fallvignetten pro Fallheft). Jeder Fall konnte somit über alle Versuchspersonen hinweg etwa gleich häufig in weiblicher und männlicher Variante präsentiert werden.

Für jeden Fall wurde außerdem geprüft, ob sich die Lösungsraten für die männliche bzw. weibliche Form unterscheiden (t-Tests für unabhängige Stichproben). Da in keinem der Fälle ein signifikanter Unterschied zu verzeichnen war (alle $p \geq .109$), wurden die über beide Fallvarianten gemittelten Lösungsraten für alle Berechnungen verwendet.

7.3.2.3 Variablen

Unabhängige Variablen

Diagnostischer Fall (Innersubjekt-Variable). Die Fallvignetten, die den Studierenden vorgelegt wurden, bildeten eine Variable, die den Schwierigkeitsgrad modulieren sollte. Diese Modulation erfolgte einerseits durch die unterschiedlichen Phänomene, die in individuellen Fähigkeitsprofilen dargestellt waren (vgl. Tab. 6.4).

Andererseits setzten sich die fallspezifischen diagnostischen Informationen aus einer variierenden Zahl an kongruenten und inkongruenten Diagnoseinformationen zusammen (für eine Übersicht der Phänomene vgl. Tab. 6.6).

Diagnostische Entscheidung (Innersubjekt-Variable). Für jede Fallvignette leiteten die Studierenden zwei pädagogisch-diagnostische Entscheidungen ab: Eine Diagnose- und eine Empfehlungsentscheidung. Für jede dieser Entscheidungen wurden den Studierenden je sechs Antwortoptionen zur Verfügung gestellt. Bei Diagnosen konnte eine aus sechs Antwortoptionen gewählt werden, bei den Empfehlungen waren teilweise mehrere Antwortoptionen auszuwählen (vgl. Tab. 7.1).

Reihenfolge der Entscheidungen (Zwischensubjekt-Variable). Um die Schrittfolge innerhalb der Evaluationsphase des diagnostischen Prozesses zu untersuchen, wurde die Reihenfolge der zu treffenden pädagogisch-diagnostischen Entscheidungen experimentell variiert. In der Bedingung „nicht-normativ“ (Empfehlung vor Diagnose) gaben die Studierenden wie in der Pilotierung zunächst ihre Empfehlung ab und erst danach äußerten sie ihren Diagnoseverdacht. In der Bedingung „normativ“ (Diagnose vor Empfehlung) hingegen folgten sie dem normativen Prozess, indem sie zunächst ihre Diagnosevermutung abgaben und anschließend die Empfehlung ableiteten.

Abhängige Variablen

Subjektiver Nutzen der Lerngelegenheit. Um zu evaluieren, inwiefern die Studierenden das Fallinventar als nützliche Lerngelegenheit empfinden, wurden sie gebeten anzugeben, wie hilfreich sie diese zum einen für die Klausur (Modulabschlussprüfung) sowie andererseits für die spätere Berufspraxis empfanden. Hierzu stand eine sechsstufige Likert-Skala (1 = überhaupt nicht hilfreich, 6 = sehr hilfreich) zur Verfügung.

Lösungsrate der Entscheidungen. Die Lösungsrate für jede Entscheidung in jedem Fall spiegelt den Anteil der Studierenden wider, die eine korrekte Lösung erzeugen konnten (Minimum = 0, Maximum = 1). Die Korrektheit der jeweiligen Entscheidung wurde zunächst für jeden Fall mithilfe einer Musterlösung ermittelt (vgl. Tab. 7.1). Korrekte Lösungen wurden mit 1 Punkt kodiert, inkorrekte mit 0 Punkten. Die über alle Studierenden gemittelte Lösungsrate pro Entscheidung und Fall kann im Sinne der Aufgabenschwierigkeit interpretiert werden (Bortz & Döring, 2016).

Subjektive Sicherheit. Neben der tatsächlichen Korrektheit der Lösung wurde für jede fallbezogene Entscheidung erfasst, inwiefern sich die Studierenden zutrauten, die jeweilige Entscheidung korrekt getroffen zu haben. Auf einer sechsstufigen Likert-Skala (1 = gar nicht sicher, 6 = sehr sicher) gaben sie an, wie sicher sie sich beim Treffen der jeweiligen Entscheidung fühlten.

Anteil genutzter Diagnoseinformationen. Die fähigkeitsbezogenen Diagnoseinformationen der Fallvignetten wurden jeweils in formeller, semiformeller und informeller Form (Kategorisierung sensu Hascher, 2008) vorgegeben. Die Studierenden wurden nach jeder diagnostischen Entscheidung um eine Angabe der jeweils genutzten Diagnoseinformationen gebeten. Hierzu wählten sie für jeden der drei Fähigkeitsbereiche (Lesen und Leseverständnis, Zahlen und Operationen, kognitive Fähigkeiten) einer Fallvignette aus, welche der drei zur Verfügung gestellten Diagnoseinformationen jeweils am relevantesten für die jeweilige Entscheidung war. Die Studierenden wurden hierbei nicht darüber informiert, um welche Informationsart es sich handelt. Sie gaben lediglich die Position der Information (von oben nach unten durchnummeriert) an oder wählten die Option „kann ich nicht genau sagen“ aus. Die Nutzungsanteile pro Entscheidung (Diagnose, Empfehlung) und Informationsart (formell, semiformell, informell) wurden berechnet, indem für jede*n Studierende*n die Anzahl der jeweils genutzten Informationsart in jeder Fallvignette (Wertebereich 0 - 3) ermittelt und für alle acht Fälle aufsummiert wurde (Wertebereich 0 - 24). Da die Nutzungssummen aller drei Informationsarten keine unabhängigen Ereignisse repräsentierten, betrug die Maximalsumme der Informationsangaben 24 (Summe formell + Summe semiformell + Summe informell = Gesamtsumme der Angaben zu genutzten Informationen). Bei Studierenden, die unbearbeitete Fälle oder fehlende Einzelangaben aufwiesen, betrug die maximale Nutzungssumme < 24 Informationsangaben. Bei der Berechnung der individuellen Nutzungsanteile der Studierenden wurde die Nutzungssumme jeder Informationsart durch die individuelle Maximalzahl an Informationsangaben dividiert.

7.3.2.4 Studiendesign und Vorgehen

Vorgehen bei der Datenerhebung

Die Untersuchungen wurden im Rahmen der Lehrveranstaltungen im Modul „Pädagogische Diagnostik“ implementiert. Zur Bearbeitung des Fallinventars wurde eine 90-minütige Seminarsitzung mit integrierter Befragung eingeplant. Die Studierenden bekamen in dieser Sitzung jeweils ein ausgedrucktes Fallheft und wurden gebeten, dieses in Einzelarbeit ohne Zuhilfenahme ihrer Seminarunterlagen oder Sitznachbar*innen zu bearbeiten. Die Eingabe der jeweiligen Aufgabenlösungen und die Beantwortung darüber hinausgehender Fragen erfolgte digital (iPads oder PCs) über die Online-Plattform „Blackboard“ der Freien Universität, die die Durchführung anonymer Umfragen ermöglicht.

Validierung der Musterlösungen für Diagnosen und Empfehlungen

Ausgehend von den Ergebnissen der Pilotierung wurden in dieser Studie u. a. Fragestellungen untersucht, die sich auf Schwierigkeiten bei der *Abgabe von Empfehlungen* beziehen. In der Pilotierung der Fallvignetten zeigten sich für Empfehlungsentscheidungen zumeist niedrigere Lösungsraten als für Diagnosen, wobei

nicht ausgeschlossen werden konnte, dass die festgelegten Musterlösungen diese Unterschiede begünstigten. Während bei der Festlegung der korrekten Diagnosen eindeutige Diagnosekriterien und prototypische Fähigkeitsprofile (vgl. Tabelle 6.4) herangezogen werden konnten, gestaltete sich die Festlegung der Musterlösungen für die Empfehlungen grundsätzlich weniger eindeutig: Zum einen waren teilweise Mehrfachnennungen unter den Antwortoptionen möglich, und zum anderen waren je nach Phänomen durchaus verschiedene korrekte Lösungsmuster denkbar.

Um die eigenen entwickelten Musterlösungen zu validieren, wurden Expert*innen aus Theorie ($N = 5$ Dozierende im Modul „Pädagogische Diagnostik“) und Praxis ($N = 3$ Schulpsycholog*innen aus Berliner SIBUZ⁷) gebeten, die acht Fallvignetten zu bearbeiten und dabei ihre Lösungen, Fragen und Kommentare schriftlich festzuhalten. Dabei bestätigten sich die Musterlösungen zu den Diagnosen, die als Vergleichsgrundlage ebenfalls abgefragt wurden. Hinsichtlich der Empfehlungen gaben die Expert*innen allerdings zusätzlich andere Lösungsvorschläge ab. Sobald mehr als ein/e Expert/in von dem ursprünglichen Lösungsvorschlag abwich, wurde deren Antwort ebenfalls als korrektes Lösungsmuster eingestuft. In drei von acht Fällen wurden dementsprechend jeweils zwei unterschiedliche Lösungsmuster als korrekt gewertet, wenn Studierende diese ankreuzten (vgl. Tabelle 7.1).

7.3.2.5 Statistische Analysen

Die Bereinigung und Auswertung der Daten erfolgte mithilfe der Statistik-Software SPSS Version 18 (Statistical Package for the Social Sciences; IBM Corporation, 2015).

1a. Rezeption der Lerngelegenheit durch die Studierenden

Die Angaben der Studierenden dazu, wie hilfreich sie die Lerngelegenheit für die Abschlussprüfung im Modul sowie die spätere Berufspraxis einschätzten, wurden ausschließlich deskriptiv analysiert. Ergänzt wurden diese Angaben um eine Auswahl an freien Kommentaren, die die Studierenden im Anschluss an die Bearbeitung des Fallinventars abgeben konnten.

1b. Eignung der Fallvignetten für die Lerngelegenheit

Zur Beantwortung der Fragen, ob die Fallvignetten für die Lerngelegenheit geeignet sind, wurden die *mittleren Lösungsraten* bei Diagnose- und Empfehlungsentscheidungen jedes Falles herangezogen. Durch die Kodierung korrekter Lösungen mit 1 bzw. inkorrekt gelöster Lösungen mit 0 entsprach die mittlere Lösungsrate pro Entscheidung und Fall der Aufgabenschwierigkeit. Ein Schwierigkeitsindex von $p = .50$ bedeutet dementsprechend, dass diese Aufgabe von 50 % der Studierenden korrekt gelöst wurde (vgl. Bortz & Döring, 2016). Bortz und Döring (ebd.) zufolge wurden Aufgaben mit $p \leq .20$ als sehr schwierig und solche mit $p \geq .70$ als sehr leicht eingeordnet. Als geeignet wurden die Fälle hier bezeichnet, wenn sie eine Aufgabenschwierigkeit von $p = .20$ nicht unter- bzw. den Wert $p = .90$ nicht überschritten. Um die

⁷ SIBUZ: Schulpsychologisches und Inklusionspädagogisches Beratungs- und Unterstützungszentrum. Dem Berliner Senat unterstellte Institution zur Unterstützung von Schulen, Lehrkräften und Eltern bei diagnostischen Fragen im Zusammenhang mit Lernbesonderheiten und sonderpädagogischen Förderbedarfen.

Schwierigkeiten der einzelnen Fälle inferenzstatistisch zu vergleichen, wurde die Summe der korrekten Entscheidungen pro Fall herangezogen (Wertebereich 0 - 2). Diese ordinal skalierten Punktesummen der acht Fälle wurden mithilfe einer Friedman-ANOVA auf einen Falleffekt hin untersucht. Anschließende paarweise Vergleiche wurden mittels Wilcoxon-Tests durchgeführt. Hierbei wurde das Signifikanzniveau a priori entsprechend der Anzahl der Einzelvergleiche auf $p = .05/36 = .001$ festgesetzt (Bonferroni-Korrektur).

Zur Bewertung der Eignung der Fallvignetten wurde außerdem das Ausmaß der *mittleren subjektiven Sicherheit* der Studierenden exploriert. Hier galt der Skalenmittelwert von $M = 3.5$ auf der sechsstufigen Likert-Skala als kritischer Wert, der bestenfalls nicht unterschritten werden sollte. Um zu prüfen, ob die subjektive Sicherheit der Studierenden eher durch die einzelnen Fallvignetten oder fallübergreifend durch die Entscheidungen beeinflusst war (Anknüpfung an Fragestellung 2a), wurde eine 2x2-ANOVA mit Messwiederholung mit den Faktoren Entscheidung (2 Stufen) und Fall (8 Stufen) durchgeführt. Die Annahme der Sphärizität war Mauchlys Test zufolge nicht verletzt. Für die Gruppenvergleiche wurde das partielle η^2 als Maß der Effektstärke berechnet. Zur Identifikation bedeutsamer Unterschiede wurden hier, wie für alle Vergleiche innerhalb dieser Studie, die folgenden Grenzen nach Cohen (1988) festgelegt: Alle Effekte mit $\eta^2 \geq .01$ werden als klein, mit $\eta^2 \geq .06$ als mittlerer Effekt und mit $\eta^2 \geq .14$ als großer Effekt bezeichnet. Anschließende paarweise Vergleiche der subjektiven Sicherheit für Diagnosen vs. Empfehlungen auf Einzelfallebene wurden mittels t-Tests für verbundene Stichproben geprüft. Diese Einzelvergleiche wurden entsprechend der Anzahl der Einzelvergleiche auf einem a priori festgelegten Signifikanzniveau von $p = .05/8 = .006$ geprüft (Bonferroni-Korrektur).

2a. Güte der diagnostischen Produkte

Zur Untersuchung der Fähigkeiten der Studierenden, *korrekte Diagnose- vs. Empfehlungsentscheidungen* zu treffen, wurden die fallübergreifenden Lösungsraten pro Entscheidung mittels eines t-Tests für verbundene Stichproben verglichen. Die mittleren Lösungsraten pro Fall und pro Entscheidung wurden hierfür ermittelt, indem die Summe der jeweils korrekt gelösten Fälle (Minimum = 0, Maximum = 8) durch die maximale Punktzahl geteilt wurde (Wertebereich 0 - 1). Nach anschließender Mittelung über alle Studierenden ergaben sich zwei Lösungsraten, die verglichen wurden. Anschließend wurde die Effektstärke Cohen's d_z für gepaarte t-Tests berechnet, indem die Mittelwertsdifferenz durch die gepoolte Standardabweichung dividiert wurde (Lakens, 2013). Zur Einordnung der Effektgröße wurden hier, wie bei allen weiteren Effektstärkenberechnungen in dieser Studie, die folgenden Grenzen für die Identifikation bedeutsamer Unterschiede nach Cohen (1988) festgelegt: kleiner Effekt für $|d| \geq .20$, mittlerer Effekt für $|d| \geq .50$ und großer Effekt für $|d| \geq .80$.

Tabelle 7.1 Validierte Lösungsvorschläge für Diagnosen und Empfehlungen

Fall- vignette	Diagnosen		Empfehlungen	
	Expert*innenlösung	Eigener Lösungsvorschlag		Expert*innen-Lösung
1. Luca / Jana	ein/e Schüler/in mit einer kombinierten Störung schulischer Fertigkeiten (c)	schulische Kleingruppen- oder lerntherapeutische Förderung im Bereich des Lesens (b) und schulische Kleingruppen- oder lerntherapeutische Förderung im Bereich des Rechnens (c)	–	Lösung identisch mit Lösungsvorschlag
2. Katrin / Felix	ein/e normal begabte/r Schüler/in mit eventuellen Stärken oder Schwächen in einem / mehreren Fächern (a)	Keine besonderen Maßnahmen sind erforderlich, die reguläre Förderung im Rahmen des Unterrichts ist ausreichend (f)	–	Erstlösung identisch mit Lösungsvorschlag; – Alternativlösung: schulische Kleingruppen- oder lerntherapeutische Förderung im Bereich des Lesens (b) und im Bereich des Rechnens (c)
3. Jan / Leonie	ein/e normal begabte/r Schüler/in mit eventuellen Stärken oder Schwächen in einem / mehreren Fächern (a)	Keine besonderen Maßnahmen sind erforderlich, die reguläre Förderung im Rahmen des Unterrichts ist ausreichend (f)	–	Erstlösung identisch mit Lösungsvorschlag; – Alternativlösung: Teilnahme an der AG "Club Einstein" für mathematisch sehr begabte Schüler*innen (d)
4. Laura / Thomas	ein/e hochbegabte/r Schüler/in mit guten bis hervorragenden Schulleistungen (e)	Teilnahme an der AG "Club Einstein" für mathematisch sehr begabte Schüler*innen (d) und an der AG "Club Goethe" für schriftsprachlich sehr begabte Schüler*innen (e)	–	Lösung identisch mit Lösungsvorschlag
5. Julia / Maximilian	ein/e normal begabte/r Schüler/in mit eventuellen Stärken oder Schwächen in einem / mehreren Fächern (a)	Keine besonderen Maßnahmen sind erforderlich, die reguläre Förderung im Rahmen des Unterrichts ist ausreichend (f)	–	Lösung identisch mit Lösungsvorschlag
6. Lena / Christian	ein/e hochbegabte/r Minderleister/in (f)	Teilnahme an der AG "Club Einstein" für mathematisch sehr begabte Schüler*innen (d) und an der AG "Club Goethe" für schriftsprachlich sehr begabte Schüler*innen (e)	–	Lösung identisch mit Lösungsvorschlag
7. Lukas / Sophie	ein/e Schüler/in mit einer spezifischen Lernstörung im Bereich des Lesens oder Rechnens (b)	schulische Kleingruppen- oder lerntherapeutische Förderung im Bereich des Lesens (b)	–	Lösung identisch mit Lösungsvorschlag
8. Hanna / Jonas	ein/e Schüler/in mit einer "Lernbehinderung" (d)	sonderpädagogische Förderung der allgemeinen/metakognitiven Fähigkeiten (a), schulische Kleingruppen- oder lerntherapeutische Förderung im Bereich des Lesens (b) und schulische Kleingruppen- oder lerntherapeutische Förderung im Bereich des Rechnens (c)	–	Erstlösung identisch mit Lösungsvorschlag – Alternativlösung: nur sonderpädagogische Förderung der allgemeinen/metakognitiven Fähigkeiten

Anmerkungen: Diagnosen wurden im Multiple-Choice-Format vorgegeben. Für Empfehlungen waren Mehrfachnennungen möglich (Mehrfach-Auswahl-Format).

Um Hinweise darauf zu erhalten, inwiefern die Studierenden bei der Vergabe von Empfehlungen von der korrekten Vergabe von Diagnosen profitieren, wurde zusätzlich der *Zusammenhang der Lösungsraten* für Diagnosen und Empfehlungen mithilfe der Pearson-Produkt-Moment-Korrelation geprüft. Neben den Lösungsraten an sich, die einen Hinweis auf tatsächliche Fähigkeiten geben, wurde schließlich analysiert, inwiefern diese mit der *subjektiven Sicherheit* für die beiden Entscheidungen übereinstimmen. Hierzu wurden die mittleren Lösungsraten für Diagnosen und Empfehlungen mit der jeweiligen mittleren fallübergreifenden subjektiven Sicherheit korreliert.

Zur Einordnung der Stärke von Zusammenhängen wurde in der vorliegenden Studie auf die Einteilung nach Cohen (1988) zurückgegriffen, der zufolge Zusammenhänge von $|r| \geq .10$ als schwach, $|r| \geq .30$ als moderat und $|r| \geq .50$ als stark bezeichnet werden.

2b. Abfolge der diagnostischen Entscheidungen

Zur Analyse der *Lösungsraten in Abhängigkeit von der Reihenfolge* der Entscheidungen wurde eine gemischte ANOVA mit dem Zwischensubjektfaktor Reihenfolge (normativ, nicht-normativ) sowie dem Innersubjektfaktor Entscheidung (Diagnose, Empfehlung) durchgeführt. Die Annahme der Sphärizität war Mauchlys Test zufolge nicht verletzt, der Levene-Test zur Prüfung der Varianzhomogenität war ebenfalls nicht signifikant. Für die Gruppenvergleiche wurde η^2 als Maß der Effektstärke berechnet. Mithilfe anschließender t-Tests für unabhängige Stichproben wurden die Unterschiede für Diagnosen und Empfehlungen getrennt untersucht, wobei die Irrtumswahrscheinlichkeit entsprechend der Anzahl der Einzelvergleiche a priori auf $p = .05/2 = .025$ festgesetzt wurde (Bonferroni-Korrektur). Hierfür wurde jeweils Cohen's d als Effektstärkenmaß für ungepaarte t-Tests mithilfe eines Online-Effektstärke-Rechners (Hemrich, 2015) berechnet.

2c. Nutzung formeller Diagnoseinformationen

Hinsichtlich der Nutzung der Diagnoseinformationen wurde zunächst deskriptiv analysiert, inwiefern die Studierenden für ihre Entscheidungen – wie intendiert – vermehrt formelle diagnostische Informationen nutzen. Hierfür wurden die *mittleren Nutzungsanteile* aller drei Informationsarten (formell, semi-formell, informell) herangezogen⁸. Ob sich die Nutzungsanteile der drei Informationsarten zwischen *Diagnosen und Empfehlungen* unterscheiden, wurde inferenzstatistisch mittels t-Tests für verbundene Stichproben überprüft. Das Signifikanzniveau wurde hierfür a priori auf $p = .05/3 = .017$ festgelegt. Anschließend wurde Cohen's d_z als Maß der Effektstärke für gepaarte t-Tests berechnet.

Zur Untersuchung der Frage, ob die vermehrte Nutzung formeller Informationen mit höheren Lösungsraten einhergeht, wurde für Diagnosen sowie Empfehlungen der Zusammenhang des jeweiligen

⁸ Da die Höhe des Nutzungsanteils einer Informationsart die höchstmögliche Nutzungsrate der beiden verbleibenden Informationen bestimmte (abhängige Ereignisse), erfolgte der Vergleich der fall- und entscheidungsübergreifenden Nutzungsraten ausschließlich deskriptiv.

Nutzungsanteils formeller Diagnoseinformationen mit der jeweiligen Lösungsrate mittels der Produkt-Moment-Korrelation nach Pearson analysiert.

Schließlich wurden die Nutzungsanteile der Diagnoseinformationen im *Kontext der Reihenfolge der Entscheidungen* untersucht (Anknüpfung an Fragestellung 2b). Hierzu wurden die mittleren Nutzungsanteile der drei Diagnoseinformationen zunächst getrennt nach Reihenfolge-Bedingung deskriptiv analysiert. Anschließend wurde mittels t-Tests für unabhängige Stichproben geprüft, ob sich Nutzungsanteile formeller Informationen zwischen den Reihenfolgebedingungen unterscheiden. Der Vergleich erfolgte für Diagnosen und Empfehlungen getrennt. Dabei wurde zur Reduzierung der Alpha-Fehler-Kumulierung das Signifikanzniveau für diese sechs Einzelvergleiche a priori auf $p = .05/6 = .008$ festgesetzt (Bonferroni-Korrektur). Anschließend wurde für alle Vergleiche Cohen's d als Effektstärkenmaß für ungepaarte t-Tests berechnet (Effektstärken-Rechner: Hemmerich, 2015).

7.3.3 Ergebnisse

Zugunsten der Nachvollziehbarkeit der Analysen werden in diesem Abschnitt nicht nur die wesentlichen Analyseergebnisse beschrieben, sondern auch eine erste Einordnung der Ergebnisse hinsichtlich der Hypothesen vorgenommen.

7.3.3.1 Untersuchung der Eignung der Lerngelegenheit (Fragestellung 1)

1a. Rezeption der Lerngelegenheit durch die Studierenden

Die Dozierenden, die mit dem Fallinventar im Rahmen ihrer Seminare arbeiteten, berichteten in informellen Gesprächen, dass die Studierenden sich intensiv mit den Fällen beschäftigt hätten und dadurch angeregte Diskussionen unter den Studierenden entstanden seien. Die Studierenden selbst äußerten in *offenen Kommentarfeldern* unter anderem positiv, dass die Lerngelegenheit nah an der Realität sei, eine Überprüfung des eigenen Wissens ermögliche und das Problem der Urteilsfehler bewusst mache. Dass mehrere Perspektiven auf die Schüler*innen gegeben werden, wurde von mehreren Studierenden lobend hervorgehoben. Auf der anderen Seite wurde kritisiert, dass für die Beurteilung der Fälle zu wenige Informationen vorgelegen hätten, eigentlich mehrere Informationen relevant gewesen seien und die Begründungen generell nicht erfasst wurden. Es wurde auch bemängelt, dass es für diese Übung nicht ausreichend Vorbereitung und Einführung gegeben habe, dass Vorübungen im Seminar hilfreich gewesen wären, dass das Verstehen der Testwerte nicht einfach sowie die Bearbeitung ohne die Unterlagen zu schwierig gewesen sei. Ferner wurde von einigen Studierenden rückgemeldet, dass sich die Bearbeitung eher wie ein Test und nicht wie eine Übung angefühlt habe und dass die Durchführung als Partnerübung statt am PC angenehmer gewesen wäre.

Subjektiver Nutzen der Lerngelegenheit für die Studierenden. Trotz dieser „Testatmosphäre“ schätzten die Studierenden die Lerngelegenheit auf einer sechsstufigen Rating-Skala im Mittel als „eher hilfreich“ für die Klausur ($M = 4.49, SD = 1.05, N = 124$) sowie als „hilfreich“ für die spätere Berufspraxis ($M = 4.93, SD = .87, N = 125$) ein. Die Untersuchung der Antworthäufigkeiten (vgl. Abb.

7.1) ergab, dass die Übung von mehr als der Hälfte der Studierenden als „hilfreich“ ($N = 64$, entspricht 51.6 %) bzw. von etwa 10 % als „sehr hilfreich“ ($N = 14$, entspricht 11.3 %) für die Modulabschlussprüfung empfunden wurde. Der Nutzen für die Praxis wurde ebenfalls von etwa der Hälfte der Studierenden als „hilfreich“ ($N = 64$, entspricht 51.2 %) bzw. von etwa einem Viertel der Studierenden als „sehr hilfreich“ ($N = 31$, entspricht 24.8 %) eingeschätzt. Diese Ergebnisse waren sehr zufriedenstellend.

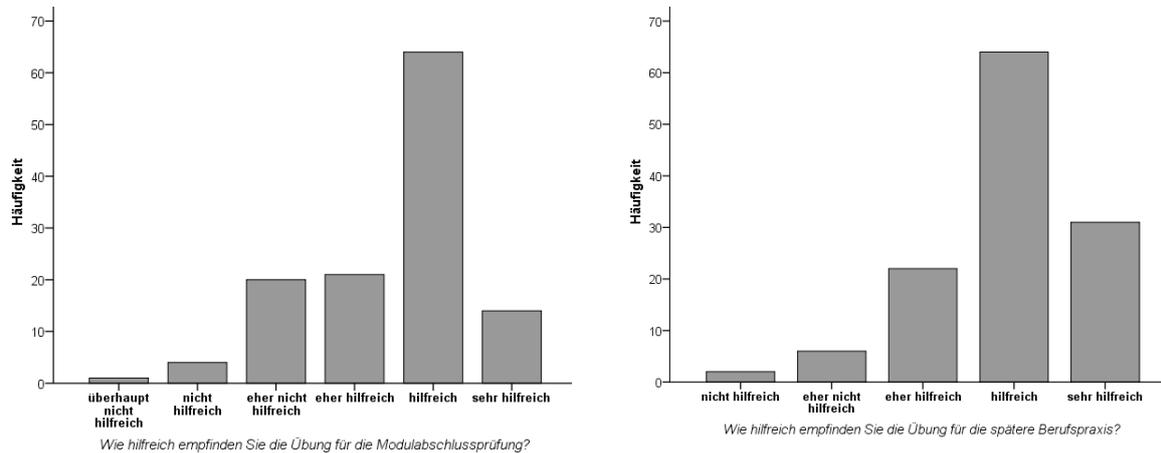


Abbildung 7.1 Häufigkeiten der Antwortoptionen zum subjektiv empfundenen Nutzen des Fallinventars als Lerngelegenheit (sechsstufige Antwortskala)

1b. Eignung der einzelnen Fallvignetten

Im Fokus der Analysen zur Eignung der Fallvignetten waren die Lösungsraten hinsichtlich der einzelnen diagnostischen Entscheidungen in den jeweiligen Fällen, die zur Bemessung der Aufgabenschwierigkeit genutzt wurden. Ergänzend dazu wurde die fallbezogene subjektive Sicherheit bei der jeweiligen Entscheidung analysiert. In Tabelle 7.2 werden die Aufgabenschwierigkeiten (mittleren Lösungsraten) sowie die mittleren subjektiven Sicherheitsratings der Studierenden dargestellt.

Deskriptive Analyse der fallbezogenen Aufgabenschwierigkeiten. Zur Einschätzung der Aufgabenschwierigkeiten wurde für alle Fälle die jeweils größtmögliche Anzahl an Datenpunkten verwendet, sodass die Anzahl der Messwerte hier für die fallbezogenen Entscheidungen variierte (vgl. Abschnitt II in Tabelle 7.2). Die deskriptive Analyse ergab, dass die Schwierigkeitsindices weitestgehend im erwünschten Bereich $.20 \leq p \leq .90$ lagen und die Fallvignetten daher für die Lerngelegenheit geeignet waren. Nur in einem Fall lag die Schwierigkeit geringfügig oberhalb des Grenzwertes (Fall 6: $p = .91$). Gegenüber der Pilotierung wichen die Schwierigkeitsindices in einigen Fällen mehr oder weniger stark von der ersten Erprobung ab (vgl. Abschnitt I in Tabelle 7.2): Die in der Pilotierung eher ungünstigen Fallszenarien „Johanna“ (hier Fall 4) sowie „Louisa“ (hier Fall 8) wiesen nun weniger extreme Schwierigkeitsindices auf und bewegten sich innerhalb der festgelegten kritischen Grenzen. Alle acht Fallvignetten wurden aufgrund dieser Ergebnisse als geeignet für die Lerngelegenheit erachtet.

Für *Diagnoseentscheidungen* variierten die Schwierigkeitsindices zwischen „schwer“ (Fall 1) und „sehr leicht“ (Fall 2, Fall 4, Fall 5, Fall 6). Drei Fälle (Fall 3, Fall 7, Fall 8) wiesen mittlere Aufgabenschwierigkeiten auf. Bei *Empfehlungsentscheidungen* bewegten sich die Schwierigkeitsindices zwischen „schwer“ (Fall 1) und „sehr leicht“ (Fall 3, Fall 7). Hier wurden für fünf Fälle mittlere Aufgabenschwierigkeiten ermittelt (Fall 2, Fall 4, Fall 5, Fall 6, Fall 8). Beim deskriptiven Vergleich der Schwierigkeitsindices entstand der Eindruck, dass die Lösungsanteile für die Empfehlungsentscheidungen in der Mehrheit der Fälle niedriger ausfielen als für Diagnoseentscheidungen, Diagnosen also leichter waren.

Deskriptive Analyse der fallbezogenen Sicherheitsratings. Zur Analyse der subjektiven Sicherheit wurden die gemittelten Ratings der subjektiven Sicherheiten für die jeweiligen Fällen analysiert. Die Studierenden fühlten sich demnach bei allen acht Fallvignetten ausreichend sicher (vgl. Tab. 7.2. Abschnitt II). Bei *Diagnosen* bewegten sich die Ratings zwischen „eher unsicher“ (Fall 1) und „sicher“ (Fall 4, Fall 6). In den verbleibenden fünf Fällen fühlten sich die Studierenden „eher sicher“ in ihren Diagnoseentscheidungen (Fall 2, Fall 3, Fall 5, Fall 7, Fall 8). Damit lag die mittlere subjektive Sicherheit bei nur einem Fall geringfügig unterhalb des kritischen Wertes von $p_{\text{subjektiv}} \geq 3.5$ (theoretischer Skalenmittelpunkt). Bei *Empfehlungen* hingegen lagen im Mittel alle Sicherheitsratings oberhalb des Grenzwertes und die Studierenden fühlten sich hier im Mittel in allen Fällen „eher sicher“. Insgesamt ergab die deskriptive Analyse der Sicherheitsratings den Eindruck, dass sich die Studierenden bei Diagnosen und Empfehlungen relativ ähnlich sicher fühlten, und das auch, wenn sich die Lösungsraten, wie z. B. in Fall 4 oder 6, recht stark unterschieden.

Untersuchung fallbezogener Effekte bei Aufgabenschwierigkeiten und Sicherheitsratings. Bei den deskriptiven Analysen im vorigen Abschnitt ergab sich der Eindruck, dass die Lösungsraten von den Sicherheitsratings der einzelnen Fälle abwichen. Zur inferenzstatistischen Untersuchung der Falleffekte wurden Personen mit fehlenden Einzeldaten listenweise ausgeschlossen und die Analysen somit nur für vollständige Datensätze durchgeführt (vgl. Tab. 7.2. Abschnitt III). Zur Analyse der Falleffekte für korrekte Lösungen wurden die *entscheidungsübergreifenden Lösungssummen* (0 - 2 korrekte Lösungen pro Fall) herangezogen. Eine Friedman-ANOVA mit Messwiederholung für den Faktor Fall bestätigte fallspezifische Schwankungen in den Lösungssummen: $\chi^2_{\text{F}}(7) = 141.28, p < .001$. Anschließend paarweise Vergleiche mittels Wilcoxon-Tests mit a priori festgelegtem Signifikanzniveau von $p = .001$ zeigten für Fallvignette 1 im Vergleich zu allen anderen Fällen eine signifikant geringere Lösungssumme (alle $ps < .001$). Für Fall 8 ergaben sich geringere Lösungssummen als für Fall 7 und Fall 5, Fall 5 wiederum evozierte signifikant höhere Lösungssummen als Fall 2 ($ps < .001$). Alle anderen Vergleiche waren nicht signifikant (alle $ps \geq .003$). Auf inferenzstatistischer Ebene zeigte sich damit deutlich, dass Fall 1 sowie tendenziell auch Fall 8 den Studierenden entscheidungsübergreifend am schwersten, Fall 5 und 7 hingegen tendenziell leichter fielen als die verbleibenden Fallvignetten. Die verbleibenden Fallvignetten waren dementsprechend in einem vergleichbaren mittleren Schwierigkeitsbereich einzuordnen. Eine Aufstellung der von den Studierenden in den einzelnen Fällen gewählten Antwortoptionen, die für eine qualitative Fehleranalyse herangezogen werden kann, ist in Anhang 2B einsehbar.

Tabelle 7.2 Fallbezogene Aufgabenschwierigkeiten und Sicherheitsratings

		<i>Fallvignetten (in weiblicher / männlicher Version)</i>							
		(1) Jana / Luca	(2) Katrin / Felix	(3) Leonie / Jan	(4) Laura / Thomas	(5) Julia / Maximilian	(6) Lena / Christian	(7) Sophie / Lukas	(8) Hanna / Jonas
<i>I Aufgabenschwierigkeiten aus der Pilotierung</i>									
Mittlere Lösungsraten (SD)	Diagnosen	.41 (.49)	ohne Werte	.69 (.46)	1.00 (.00)	.85 (.36)	.80 (.40)	ohne Werte	.13 (.34)
	Empfehlungen	.47 (.50)		.50 (.50)	.54 (.50)	.54 (.50)	.53 (.50)		.80 (.40)
<i>II Aufgabenschwierigkeiten und Sicherheitsratings: Vollständiger Datensatz mit fallweisem Ausschluss einzelner Proband*innen</i>									
Mittlere Lösungsraten (SD)	Diagnosen	.31 (.47) <i>N</i> = 127	.79 (.41) <i>N</i> = 126	.65 (.48) <i>N</i> = 125	.88 (.33) <i>N</i> = 126	.90 (.30) <i>N</i> = 125	.91 (.28) <i>N</i> = 127	.70 (.46) <i>N</i> = 128	.56 (.50) <i>N</i> = 126
	Empfehlungen	.26 (.44) <i>N</i> = 126	.52 (.50) <i>N</i> = 127	.74 (.44) <i>N</i> = 125	.45 (.50) <i>N</i> = 126	.70 (.46) <i>N</i> = 125	.46 (.50) <i>N</i> = 125	.80 (.41) <i>N</i> = 127	.61 (.49) <i>N</i> = 126
Mittlere subjektive Sicherheit (SD)	Diagnosen	3.63 (1.07) <i>N</i> = 126	3.67 (.50) <i>N</i> = 125	4.33 (.88) <i>N</i> = 125	4.46 (1.03) <i>N</i> = 126	4.14 (1.00) <i>N</i> = 125	4.23 (1.16) <i>N</i> = 127	4.33 (.98) <i>N</i> = 127	3.93 (1.00) <i>N</i> = 126
	Empfehlungen	3.38 (1.09) <i>N</i> = 127	3.79 (.97) <i>N</i> = 125	4.21 (1.06) <i>N</i> = 124	4.62 (1.12) <i>N</i> = 125	4.25 (1.04) <i>N</i> = 125	4.85 (1.13) <i>N</i> = 126	3.98 (1.06) <i>N</i> = 127	3.63 (1.04) <i>N</i> = 126
<i>III Aufgabenschwierigkeiten und Sicherheitsratings: Datensatz für Analysen zu Falleffekten mit listenweisem Ausschluss einzelner Proband*innen</i>									
Mittlere Lösungsraten (SD), <i>N</i> = 117	Diagnosen	.33 (.47)	.79 (.41)	.68 (.47)	.89 (.32)	.92 (.27)	.92 (.27)	.73 (.45)	.58 (.50)
	Empfehlungen	.27 (.45)	.52 (.50)	.74 (.44)	.44 (.50)	.69 (.46)	.47 (.50)	.81 (.39)	.61 (.49)
Mittlere Summe richtiger Lösungen pro Fall (SD)		.60 (.92)	1.31 (.91)	1.42 (.91)	1.33 (.82)	1.61 (.73)	1.39 (.77)	1.54 (.84)	1.19 (.99)
Mittlere subjektive Sicherheit (SD), <i>N</i> = 112	Diagnosen	3.40 (1.11)	3.77 (1.00)	4.24 (1.08)	4.64 (1.11)	4.26 (1.06)	4.88 (1.08)	4.00 (1.08)	3.63 (1.08)
	Empfehlungen	3.63 (1.11)	3.67 (.95)	4.36 (.91)	4.46 (1.00)	4.11 (.99)	4.26 (1.17)	4.35 (1.00)	3.92 (1.05)
Mittlere subjektive Sicherheit pro Fall (SD)		3.52 (1.11)	3.72 (.98)	4.30 (.99)	4.55 (1.06)	4.19 (1.03)	4.57 (1.13)	4.18 (1.04)	3.78 (1.07)

Anmerkungen. In den Fallvignetten dargestellte Lernbesonderheiten: (1) kombinierte Störung schulischer Fertigkeiten; (2) normal begabt mit schwachen schulischen Fähigkeiten; (3) normalbegabt mit Stärke mit mathematischen Bereich; (4) hochbegabte*r Hochleister*in; (5) normal begabt; (6) Hochbegabte*r Minderleister*in; (7) spezifische Lernstörung im Bereich des Lesens; (8) Schüler*in mit einer Lernbehinderung. Fall 2 sowie 7 waren in der Pilotversion des Fallinventars noch nicht enthalten. Die Reihenfolge der Präsentation der Fallvignetten variierte innerhalb verschiedener Fallheftversionen (vgl. Anhang 2A).

Für die *mittlere subjektive Sicherheit* wurde eine ANOVA mit Messwiederholung mit den Faktoren Fall und Entscheidung durchgeführt, die einen großen Haupteffekt für den Faktor Fall, $F(7, 777) = 30.12, p < .001$, $\eta^2_{\text{part}} = .21$, sowie einen mittelgroßen Interaktionseffekt für Fall*Entscheidung ergab, $F(7, 777) = 10.51, p < .001$, $\eta^2_{\text{part}} = .09$. Anschließende t-Tests für verbundene Stichproben, die die subjektive Sicherheit bei Diagnosen und Empfehlungen auf Einzelfallebene mit a priori festgelegtem Signifikanzniveau von $p = .006$ (Bonferroni-Korrektur) verglichen, ergaben für Fälle 1, 7 und 8 bei Diagnosen ein geringeres mittleres Sicherheitsempfinden als bei Empfehlungen (alle $ps \leq .004$), in Fall 6 hingegen evozierte die Diagnoseentscheidung im Mittel höhere Sicherheitsratings als die Empfehlungsentscheidung ($p < .001$). Für alle anderen Fälle ergaben sich keine signifikanten Unterschiede zwischen den Entscheidungen (alle $ps \geq .139$). Fallübergreifend unterschieden sich die mittleren Sicherheitsratings für Diagnosen ($M = 4.10, SD = 1.08$) und Empfehlungen ($M = 4.10, SD = 1.02$) allerdings nicht (kein Haupteffekt für Entscheidung, $F(1, 111) = .04, p = .840, \eta^2_{\text{part}} < .01$). Diese Ergebnisse zu den Sicherheitsratings deuteten zusammengenommen darauf hin, dass hier die einzelnen Fälle für die Studierenden einen größeren Unterschied ausmachten als die jeweiligen diagnostischen Entscheidungen. Wie der nachfolgende Abschnitt zeigt, war dieses Ergebnis abweichend von dem Muster, das sich für die Rate korrekter Lösungen für Diagnosen vs. Empfehlungen ergab.

7.3.3.2 Untersuchung des diagnostischen Prozesses der Studierenden (Fragestellung 2)

Fragestellung 2a. Güte der diagnostischen Produkte

Vergleich der Lösungsraten bei Diagnosen und Empfehlungen. Ob sich der Anteil korrekter Lösungen für *Diagnosen vs. Empfehlungen* unterscheidet, wurde mit einem t-Test für verbundene Stichprobe anhand der fallübergreifenden mittleren Lösungsraten verglichen. Dabei zeigte sich, dass die Studierenden im Mittel bei Diagnoseentscheidungen signifikant höhere Lösungsraten erreichten als bei Empfehlungsentscheidungen, $t(116) = 8.46, p < .001, d = .78$. Während die Studierenden im Mittel knapp drei Viertel der Diagnoseentscheidungen korrekt trafen ($M = .73, SD = .18$), konnten sie nur knapp 60 % der Empfehlungen korrekt vergeben ($M = .57, SD = .19$). Die Analyse der Lösungsraten zeigte damit, dass die Studierenden bei beiden Entscheidungen Schwierigkeiten hatten, und dass das Ableiten von Empfehlungen den Studierenden noch deutlich schwerer fiel als die Vergabe von Diagnosen.

Zusammenhang der Lösungsraten von Diagnosen und Empfehlungen. Die bivariate Korrelation der Lösungsraten beider Entscheidungen ergab einen signifikanten mittleren Zusammenhang: $r = .38, p < .001$ ($N = 117$). Dieser Befund belegte zwar, dass korrekte Diagnosen durchaus mit korrekten Empfehlungen einhergingen – aber es zeigte sich hierbei ebenfalls, dass die Studierenden auch ohne korrekte Diagnosen zu korrekten Empfehlungen kamen bzw. trotz korrekter Diagnosen durchaus inkorrekte Empfehlungen abgeleitet wurden. Dieser Zusammenhang ergab sich unabhängig davon, ob Diagnosen oder Empfehlungen zuerst vergeben wurden (vgl. Anhang 2C).

Zusammenhang von Lösungsraten und subjektiver Sicherheit. Die unter Fragestellung 1b beschriebenen Ergebnisse zu den Lösungsraten und Sicherheitsratings bei Diagnosen und Empfehlungen

deuteten darauf hin, dass die subjektive Sicherheit offenbar nur bedingt davon abhängig ist, ob die getroffene Entscheidung korrekt war. Diese Vermutung wurde hier noch einmal überprüft, indem die Zusammenhänge zwischen den mittleren Lösungsraten und den mittleren Sicherheitsratings bei beiden Entscheidungen analysiert wurden. Bivariate Produkt-Moment-Korrelationen ergaben bei Diagnoseentscheidungen dabei nur einen schwachen positiven Zusammenhang zwischen der mittleren Lösungsrate und der mittleren subjektiven Sicherheit der Studierenden, der allerdings nicht signifikant war: $r = .18, p = .058$ ($N = 115$). Auch für Empfehlungen zeigte sich ein schwacher Zusammenhang, ebenfalls nicht signifikanter Zusammenhang: $r = .12, p = .189$ ($N = 114$). Diese schwachen Zusammenhänge sprechen dafür, dass die Studierenden erst in Ansätzen realistisch einschätzen können, inwiefern die von ihnen getroffenen pädagogisch-diagnostischen Entscheidungen korrekt waren.

Um diesen Zusammenhang genauer zu explorieren, wurden die Zusammenhänge zwischen Lösungsraten und Sicherheitsratings auch *getrennt für beide Reihenfolgebedingungen* untersucht (Anschluss an Fragestellung 2b). Die nach Bedingungen getrennten bivariaten Korrelationsanalysen ergaben, dass sich nur für Diagnosen und zwar nur in der normativen Bedingung ein signifikanter Zusammenhang zwischen der mittleren Lösungsrate und der mittleren subjektiven Sicherheit ergab: $r = .47, p < .001$ ($N = 55$). Dieser Zusammenhang war mittelgroß. In der nicht-normativen Reihenfolgebedingung ergab sich für Diagnosen kein signifikanter Zusammenhang, $r = -.10, p = .472$ ($N = 60$). Bei Empfehlungen ergaben sich in keiner der Bedingungen bedeutsame Zusammenhänge: normativ: $r = .20, p = .148$ ($N = 52$); nicht normativ: $r = -.03, p = .800$ ($N = 62$). Die nach Bedingungen getrennte Analyse ergab zusammengefasst, dass die Studierenden bei Diagnosen in der normativen Bedingung relativ realistisch einschätzen konnten, ob ihre Entscheidungen richtig waren. Dies gelang ihnen in der nicht-normativen Reihenfolge allerdings noch nicht. Bei Empfehlungen gelang ihnen indes in keiner der Reihenfolgebedingungen eine realistische Einschätzung darüber, ob ihre Entscheidungen korrekt waren.

Fragestellung 2b. Schrittfolge diagnostischer Entscheidungen

Unterschiede in den Lösungsraten nach Reihenfolge der Entscheidungen. Zur Untersuchung der Frage, ob die Studierenden von der normativen Reihenfolge des diagnostischen Prozesses profitieren konnten, wurden die mittleren Lösungsraten für Diagnosen und Empfehlungen in Abhängigkeit von der Reihenfolge beider Entscheidungen herangezogen. Eine gemischte ANOVA mit dem Innersubjektfaktor Entscheidung und dem Zwischensubjektfaktor Reihenfolge der Entscheidung ergab einen großen Haupteffekt für den Faktor Entscheidung, $F(1, 115) = 99.8, p < .001, \eta^2_{\text{part}} = .47$ (Mittelwerte vgl. Fragestellung 2a). Für den Faktor Reihenfolge ergab sich kein signifikanter Haupteffekt, $F(1, 115) = 3.02, p = .085, \eta^2_{\text{part}} = .026$, d. h. gemittelt über beide Entscheidungen ergaben sich in etwa vergleichbare Lösungsraten für beide Bedingungen. Drittens ergab sich eine signifikante *Interaktion Reihenfolge * Entscheidung*, $F(1, 115) = 35.9, p < .001, \eta^2_{\text{part}} = .47$. Dieser große Effekt bestätigte, dass die Lösungsraten für Diagnosen und Empfehlungen in unterschiedlichem Ausmaß von der Reihenfolge der Entscheidungen beeinflusst waren. Während Diagnoseentscheidungen in beiden Bedingungen in etwa vergleichbare Lösungsraten nach sich zogen (vgl. Abb. 7.2; nicht-normativ: $M = .71, SD = .17, N = 62$;

normativ: $M = .76$, $SD = .18$, $N = 55$), waren die Lösungsraten bei den Empfehlungen von der Reihenfolge stärker beeinflusst: In der normativen Reihenfolge trafen die Studierenden im Mittel nur etwa die Hälfte der Empfehlungsentscheidungen korrekt ($M = .49$, $SD = .17$), wohingegen sie in der nicht-normativen Bedingung fast zwei Drittel der Fälle korrekt lösen konnten ($M = .64$, $SD = .17$).

Anschließende t-Tests für unabhängige Stichproben mit der Gruppenvariable Bedingung bestätigten, dass sich die Mittelwerte für Diagnosen nicht wesentlich zwischen den Bedingungen unterschieden, $t(115) = -1.57$, $p = .120$, $d = -.29$. Bei Empfehlungen hingegen ergab sich ein großer Effekt für den Unterschied in den Lösungsraten beider Reihenfolge-Bedingungen: $t(115) = 4.59$, $p < .001$, $d = .85$.

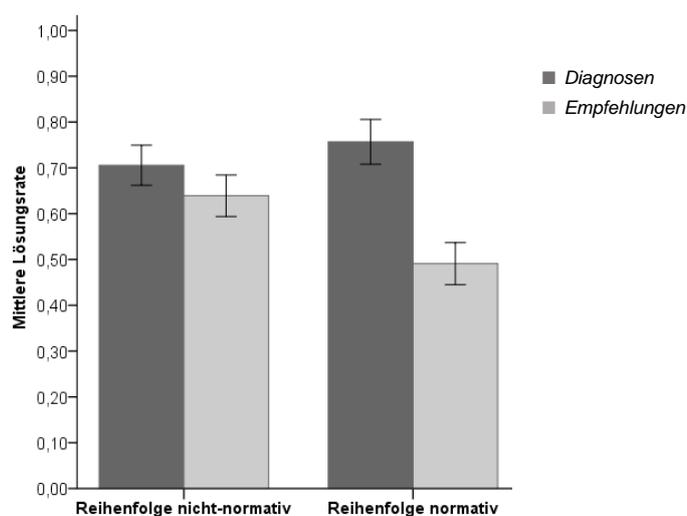


Abbildung 7.2 Mittlere Lösungsraten für Diagnosen und Empfehlungen in Abhängigkeit

nicht-normativ: Diagnose nach Empfehlung. Fehlerbalken: CI 95 %.

Fragestellung 2c. Nutzung formeller Diagnoseinformationen

Nutzung formeller Informationen für Diagnosen und Empfehlungen. Der *deskriptive Vergleich der Nutzungsanteile* für die einzelnen Diagnoseinformationen zeigt, dass die Studierenden für ihre pädagogisch-diagnostischen Entscheidungen überwiegendangaben, formelle Diagnoseinformationen genutzt zu haben (vgl. Tab. 7.2 Abschn. I). Sowohl bei Diagnosen als auch bei Empfehlungen lag der mittlere Nutzungsanteil formeller Informationen weit über 50 %. Der Anteil semiformeller Informationen lag deutlich darunter, und informelle Informationen wurden den Angaben der Studierenden zufolge am seltensten genutzt. Formelle Informationen wurden bei Diagnosen etwa fünfmal so häufig genutzt wie semiformelle Informationen, und bei Empfehlungen mehr als doppelt so häufig wie formelle Informationen. Informelle Informationen wurden zusammengenommen sogar nur in etwa einem Zehntel der Fälle herangezogen. Die Nutzung formeller Informationen, die im Fallinventar intendiert wurde, konnte damit zwar entscheidungsübergreifend festgestellt werden.

Die Nutzungsanteile unterschieden sich augenscheinlich allerdings für Diagnosen und Empfehlungen: Beim *Vergleich der Nutzungsanteile* beider Entscheidungen ergaben sich bedingungsübergreifend für alle drei Informationsarten signifikante Unterschiede (vgl. Tab. 7.2). Während die Studierenden bei Diagnosen vs. Empfehlungen deutlich häufiger angaben, auf formelle Informationen zurückgegriffen zu haben (ca. 76 % vs. 62 %, mittlerer Effekt), waren die Angaben zur Nutzung nicht-formeller Informationen genau gegensätzlich: Für Empfehlungen nutzten die Studierenden ihren Angaben zufolge sowohl semiformelle (ca. 24 % vs. 16 %, kleiner Effekt) als auch informelle Informationen (ca. 14 % vs. 8 %, mittlerer Effekt) deutlich häufiger als für Diagnoseentscheidungen (vgl. Tab. 7.3 Abschn. I). Auch wenn für Empfehlungsentscheidungen damit am häufigsten formelle Diagnoseinformationen genutzt wurden, wurde beim Ableiten von Anschlusshandlungen deutlich seltener auf diese Diagnoseinformationen höchster Güte zurückgegriffen als für Diagnosen.

Tabelle 7.3 Nutzungsanteile für die Arten diagnostischer Informationen im Vergleich

		Mittlerer Nutzungsanteil (SD)		Mittelwertsunterschiede			
(I) Diagnostische Entscheidungen		Diagnosen	Empfehlungen	<i>t</i>	<i>df</i>	<i>p</i>	<i>d_z</i>
Formelle Informationen		.76 (.17)	.62 (.22)	6.92	127	<.001	.61
Semiformelle Informationen		.16 (.12)	.24 (.14)	-5.23	127	<.001	.46
Informelle Informationen		.08 (.08)	.14 (.11)	-6.42	127	<.001	.57
(II) Reihenfolge der Entscheidungen		nicht-normativ	normativ	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Formelle Informationen	Diagnosen	.75 (.18)	.78 (.16)	-1.28	126	.202	-.23
	Empfehlungen	.68 (.19)	.55 (.22)	3.57	126	.001	.63
Semiformelle Informationen	Diagnosen	.17 (.13)	.15 (.12)	.64	126	.523	.11
	Empfehlungen	.21 (.14)	.27 (.14)	-2.50	126	.014	-.44
Informelle Informationen	Diagnosen	.09 (.08)	.06 (.07)	1.85	126	.067	.33
	Empfehlungen	.11 (.09)	.18 (.12)	-3.76	97.54	<.001	-.66

Anmerkungen: Für statistisch signifikante Unterschiede wurden die Effektstärken fett gedruckt. In Abschnitt (I) sind die mittleren Nutzungsanteile pro Informationsart getrennt für Diagnosen und Empfehlungen abgetragen. Zur inferenzstatistischen Prüfung wurden t-Tests für verbundene Stichproben durchgeführt, wobei das Signifikanzniveau entsprechend der Einzelvergleiche a priori auf $p = .017$ festgelegt wurde. In Abschnitt (II) wurden die Nutzungsanteile für beide Entscheidungen entsprechend der Reihenfolgebedingungen (normativ: Diagnose vor Empfehlung; nicht-normativ: Diagnose nach Empfehlung) verglichen. Zur inferenzstatistischen Prüfung der Unterschiede zwischen den Bedingungen wurden t-Tests für unabhängige Stichproben durchgeführt, wobei das Signifikanzniveau entsprechend der Einzelvergleiche a priori auf $p = .008$ festgesetzt wurde. Beim Vergleich der Nutzungsanteile informeller Informationen bei Empfehlungen wurden die Freiheitsgrade korrigiert, da die Varianzhomogenität verletzt war.

Zusammenhang zwischen der Nutzung formeller Informationen und Lösungsraten. Um der Frage nachzugehen, inwiefern diejenigen Studierenden, die für ihre Entscheidungen vermehrt formelle Informationen nutzen, auch häufiger korrekte pädagogisch-diagnostische Entscheidungen trafen, wurden bivariate Korrelationen zwischen den mittleren Nutzungsanteilen formeller Informationen mit den jeweiligen mittleren Lösungsraten berechnet. Für *Diagnosen* ergab sich hierbei ein schwacher Zusammenhang: $r = .25, p = .005$ ($N = 128$). Für *Empfehlungen* hingegen zeigte sich ein moderater

Zusammenhang, $r = .36, p = .002$ ($N = 117$), zwischen der Nutzung formeller Informationen und der Lösungsrate. In beiden Entscheidungen ging also eine vermehrte Nutzung formeller Diagnoseinformationen mit höheren Lösungsraten einher. Dieser Zusammenhang war allerdings bei Empfehlungen deutlich stärker ausgeprägt als bei Diagnosen. Die Scatterplots in Anhang 2C zeigen, dass sowohl die Lösungsraten als auch die Nutzungsanteile der Studierenden bei Diagnoseentscheidungen eher am oberen Rand der Bandbreite möglicher Werte lagen: Die Lösungsrate wie auch der Nutzungsanteil formeller Informationen lag bei der Mehrheit der Studierenden bei mindestens 50 %. Bei Empfehlungen war die Streuung der Lösungsraten und Nutzungsanteile formeller Informationen indes höher.

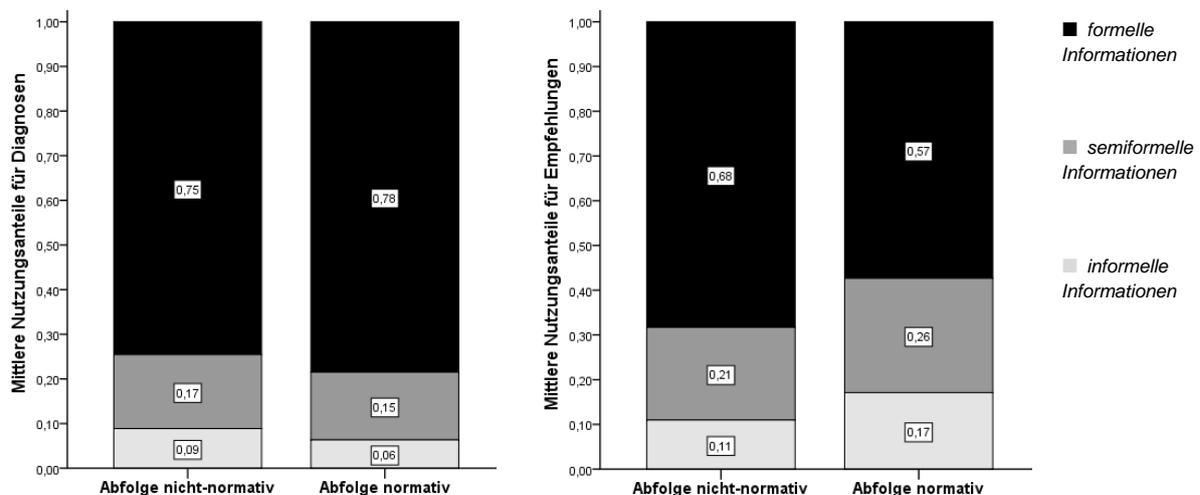


Abbildung 7.3 Nutzungsanteile der diagnostischen Informationen entsprechend der Reihenfolge von Diagnose- und Empfehlungsentscheidungen.

Normativ: Diagnose vor Empfehlung, $N = 58$.

Nicht-normativ: Diagnose nach Empfehlung, $N = 70$.

Nutzung diagnostischer Informationen in Abhängigkeit von der Reihenfolge. Die Analyse der Nutzungsanteile der drei diagnostischen Informationen getrennt für die beiden Reihenfolge-Bedingungen erfolgte aufgrund des Befundes, dass sich Lösungsraten für Diagnosen und Empfehlungen entsprechend der Reihenfolge dieser beiden Entscheidungen unterschieden (Anknüpfung an Fragestellung 2b). Die Analyse der Nutzungsanteile in Abhängigkeit von der Reihenfolge ergab hier für Diagnosen und Empfehlungen ein unterschiedliches Nutzungsmuster (vgl. Abb. 7.3). Während sich die jeweiligen *Nutzungsanteile der drei Informationsarten bei Diagnosen* in beiden Bedingungen kaum unterschieden (maximal kleine Effekte, vgl. Tab. 7.2 Abschn. II), waren bei *Empfehlungen bedingungsbezogene Unterschiede* in den Nutzungsanteilen ersichtlich: Beim Ableiten von pädagogischen Anschlusshandlungen zeigte sich eine Verschiebung bei der Nutzung der formellen und informellen Informationen. Formelle Informationen wurden in der nicht-normativen Bedingung (Diagnose nach Empfehlung) deutlich häufiger genutzt als in der normativen Bedingung (mittlerer Effekt), und informelle Informationen entsprechend in der

normativen Reihenfolge häufiger als in der nicht-normativen (mittlerer Effekt). Wenn also pädagogische Anschlusshandlungen nach der Diagnosestellung abgeleitet wurden, war bei den Studierenden ein anderes Muster bei der Nutzung formeller und informeller Informationen zu beobachten. Für semiformelle Informationen ergaben sich indes bei der Abgabe von Empfehlungen keine messbaren Unterschiede zwischen beiden Reihenfolgebedingungen (kleiner Effekt, nicht signifikant).

7.3.4 Diskussion der Ergebnisse

In dieser Studie wurde das diagnostische Fallinventar als Szenario basierte Umgebung eingesetzt. Damit wurde untersucht, wie die Studierenden die darin gestellten diagnostischen Aufgaben bearbeiten. Dabei wurden zwei Ziele verfolgt: Erstens sollte geprüft werden, inwiefern das Fallinventar mit seinen einzelnen Fallvignetten als Lerngelegenheit geeignet ist (Fragestellung 1). Zweitens war von Interesse, wie gut Lehramtsstudierende in der Lage sind, korrekte Diagnosen und Empfehlungen abzugeben, und inwiefern Aspekte des diagnostischen Prozesses zur Güte dieser Produkte beitragen können (Fragestellung 2). Beide Fragestellungen werden getrennt diskutiert und hinsichtlich des Forschungsstandes zu diagnostischen Kompetenzen bei Lehramtsstudierenden eingeordnet.

7.3.4.1 Fallinventar als Lerngelegenheit (Fragestellung 1)

1a. Eignung der Lerngelegenheit

Die Eignung der Lerngelegenheit wurde erstens daran untersucht, wie die Studierenden das Fallinventar rezipieren. Dass Teilnehmer*innen einer Intervention diese als relevant und nützlich erachten, ist eine zentrale Voraussetzung dafür, dass sie bereit sind, die angebotenen Inhalte aufzunehmen und zu verarbeiten (z. B. konkret für die Lehrkräftebildung: Lipowsky, 2014). In der vorliegenden Untersuchung nahmen die Studierenden die Übung mit dem Fallinventar insgesamt überwiegend positiv wahr und empfanden subjektiven Nutzen sowohl für die Modulabschlussprüfung als auch für die spätere Berufspraxis. Dieses Ergebnis ist erfreulich, weil hierdurch eine wichtige Voraussetzung erfüllt ist, um Lernprozesse überhaupt anzustoßen. Da die Lerngelegenheit sowohl konzeptuelles Wissen vertiefen als auch prozedurales Wissen anbahnen soll (vgl. Studie 1), ist es wichtig, dass die Studierenden für beide Anwendungsbereiche einen Nutzen empfinden und das konnte hier für beide Bereiche bestätigt werden. Vor dem Hintergrund, dass der Einsatz des Fallinventars in der vorliegenden Studie durchaus als Testsituation wahrgenommen wurde, wird die Resonanz hier möglicherweise sogar leicht unterschätzt. Ein Teil der Studierenden äußerte entsprechend, dass sie sich mehr Unterstützung und Anleitung bei der Bearbeitung der Übung gewünscht hätten. Der Einsatz als Szenario basierte Umgebung diene hier jedoch nicht nur der Validierung der Lerngelegenheit, sondern gleichzeitig auch als experimentelle Umgebung (vgl. Kaiser et al., 2015), mit deren Hilfe ein Beitrag zur Erforschung diagnostischer Produkt- und Prozessvariablen bei Studierenden geleistet werden sollte (Fragestellung 2). Daher waren hier weder eine systematische Instruktion durch die Dozierenden noch ein interaktives Setting mit Partnerarbeit möglich (anders als in Studie 3, s. Kap. 8).

1b. Eignung der Fallvignetten

Neben der ganzheitlichen Betrachtung des Fallinventars als Lerngelegenheit wurde zweitens die Eignung der Fallvignetten selbst untersucht. Dazu wurden erstens die Anteile der Studierenden analysiert, die einen Fall korrekt lösen konnten. Die so ermittelten *Itemschwierigkeiten* sollten einerseits überwiegend im mittleren (Aufgabenwahl nach Atkinson, 1957) bis leichten Bereich liegen, um Lernfreude und Erfolgserlebnisse bei den Studierenden zu ermöglichen, die für ein selbstständiges Lernen wichtig sind (z. B. Schwarzer & Warner, 2014). Auf der anderen Seite sollten auch schwierige Items enthalten sein, die potentiell geeignet sind, konzeptuelle Umstrukturierungen anzuregen (Stern et al., 2016; vgl. auch Schnotz, 2006). Dies ist im Rahmen fallbasierten Lernens vermutlich gegeben, wenn die Fälle nicht zu leicht sind und tatsächlich Problemlösungsprozesse auslösen (vgl. Wedel & Pfetsch, 2017). Für die acht Fallvignetten, die zum Einsatz kamen, bewegten sich die Schwierigkeitsindices überwiegend im erwünschten Bereich, wobei nur ein Fall sowohl bei den Diagnose- als auch bei den Empfehlungsentscheidungen geringe Lösungsraten aufwies und damit bedeutsam schwieriger zu lösen war als die verbleibenden Fälle. Umgekehrt wurden für zwei Fälle tendenziell höhere Schwierigkeitsindices gefunden als für die anderen Fälle, die damit als sehr leicht eingestuft werden können. Der schwerste Fall stellte eine*n Schüler*in mit einer kombinierten Störung schulischer Fertigkeiten dar, für welche*n nur etwa 30 % der Studierenden die korrekte Diagnose und Empfehlung ableiten konnten. Eine derart geringe Lösungsrate wäre für die Diagnostik dieses Phänomens in der Praxis absolut unzureichend. Hohe Trefferquoten wiederum wären vor diesem Hintergrund durchaus positiv zu betrachten. Die Unterschiede, die sich hinsichtlich des Anteils korrekter Lösungen für die Fälle ergibt, werden hier jedoch nicht phänomenspezifisch, sondern aufgabenspezifisch verstanden, d. h. als Lösungsraten in Bezug auf konkrete pädagogisch-diagnostische Entscheidungen in diesen spezifischen Fällen.

Inwiefern die beobachteten Fallschwierigkeiten durch die dargestellte Lernbesonderheit selbst oder eher durch die Fallgestaltung zustande kamen, konnte in der vorliegenden Studie nicht geklärt werden. Bemerkenswert war in diesem Zusammenhang, dass die Schwierigkeitsindices der Pilotierung teilweise stark von den hier erzielten Schwierigkeitsindices abwichen. Diese Abweichung zeigt einerseits, dass die Überarbeitung der Fälle gelungen war (vgl. Kap. 6.6). Sie kann andererseits als ein Hinweis darauf interpretiert werden, dass die Fallschwierigkeiten eben nicht nur durch die jeweilige Lernbesonderheit bedingt waren, die in den Fällen repräsentiert waren, sondern dass diese auch durch die konkrete Fallgestaltung modulierbar waren (s. a. Wedel & Pfetsch, 2017). Da sich die Methoden in der Pilotierung von den hiesigen unterschieden, kann natürlich nicht ausgeschlossen werden, dass ein Teil der Varianz auch auf die Stichprobenauswahl und die konkrete Operationalisierung der Erhebung zurückzuführen ist. Nichtsdestotrotz sind die variierenden Schwierigkeitsindices bei identischen Phänomenen ein Zeichen dafür, dass die Fälle, die leicht oder schwer zu lösen sind, nicht nur dadurch bedingt sind, dass die dargestellte Lernbesonderheit per se leicht oder schwer zu diagnostizieren ist. Die fallspezifischen Lösungsraten können dementsprechend nicht mit tatsächlichen Trefferquoten oder Aufdeckungsraten in der Praxis gleichgesetzt werden. Die systematische Untersuchung von Zusammenhängen zwischen

tatsächlichen Aufdeckungsraten eines diagnostischen Phänomens in der Praxis und den Trefferquoten in fallbasierten Umgebungen ist daher ein Desiderat für zukünftige Forschung. Dabei müssten mehrere Fälle zu einem Phänomen konstruiert werden, bei denen Fallmerkmale, wie bspw. Anzahl und Position kongruenter Informationen, systematisch variiert werden.

Zusätzlich zu den Trefferquoten bei Diagnose- und Empfehlungsentscheidungen wurde die *subjektive Sicherheit*, diese Entscheidungen korrekt getroffen zu haben, herangezogen, um Hinweise auf die Eignung der Fallvignetten zu erhalten. Die Exploration der subjektiven Sicherheitsratings sollte hauptsächlich sicherstellen, dass die Studierenden sich bei der Bearbeitung der Fälle keine Überforderungsgefühle erleben. Gefühle von Überforderung könnten der Lernfreude abträglich sein, und damit die Wirkung der Lerngelegenheit gefährden (Schwarzer & Warner, 2014; vgl. Atkinson, 1957). Die fallbezogene Analyse der Sicherheitsratings ergab für alle Fallvignetten im Mittel ein ausreichend hohes subjektives Sicherheitsempfinden, das in sieben von acht Fällen oberhalb des Skalenmittelpunktes lag. Bei dem Fall, der am schwersten war, war auch die subjektive Sicherheit beim Treffen der Diagnoseentscheidung etwas unterhalb des Skalenmittelpunktes anzusiedeln. In diesem Fall stimmte das Sicherheitsempfinden der Studierenden also mit der tatsächlichen Leistung beim Treffen der Entscheidung überein. Diese Übereinstimmung kann als realistische Einschätzung der eigenen Leistungen interpretiert werden, die wiederum als Zeichen von Kompetenz gedeutet werden kann (z. B. Thoren et al., 2020; aber siehe Schladitz et al., 2015). Auch im zweitschwersten Fall war die subjektive Sicherheit beim Treffen der pädagogisch-diagnostischen Entscheidungen relativ gering. In anderen Fallvignetten scheint jedoch das subjektive Sicherheitsempfinden relativ wenig mit der tatsächlichen Leistung einherzugehen: Ein Beispiel hierfür ist Fall 2, bei dem Studierenden im Mittel ähnliche Sicherheitsratings abgaben wie bei Fall 1, obwohl die Lösungsraten hier deutlich höher ausfielen. Auf diese Diskrepanz wird im nächsten Abschnitt unter Fragestellung 2a nochmals genauer eingegangen. An dieser Stelle sei noch darauf hingewiesen, dass die Sicherheitsratings durch die Fälle beeinflusst wurden (großer Falleffekt), und dass diese in der Hälfte der Fälle auch mit der jeweiligen Entscheidung interagierten (mittlerer Interaktionseffekt Fall*Entscheidung) – und zwar in der Form, dass sich die Studierenden bei diesen Fällen bei der Empfehlungsentscheidung sicherer fühlten als bei der Diagnosevermutung. Ein fallübergreifender Haupteffekt für die Entscheidungen war hier allerdings nicht zu verzeichnen. Diese Ergebnisse zeigen, dass das Sicherheitsempfinden nicht pauschal für eine Entscheidungsart höher ist, sondern dass es nur in bestimmten Fällen Diskrepanzen in der Sicherheit für Diagnosen vs. Empfehlungen gab. Bei der Analyse der Lösungsraten, die im nächsten Abschnitt erläutert wird, ergab sich jedoch ein anderes Bild (Fragestellung 2a).

7.3.4.2 Untersuchung des diagnostischen Prozesses (Fragestellung 2)

Fragestellung 2a. Güte der diagnostischen Produkte

Um Hinweise auf die Verarbeitungsprozesse der Studierenden im Fallinventar zu erhalten, wurden zunächst die diagnostischen Produkte untersucht. Als Endpunkte eines diagnostischen Prozesses lassen

sie Schlüsse auf die zugrunde liegenden Informationsverarbeitungsprozesse zu (vgl. Behrmann & Van Ophuysen, 2017). Im Kontext pädagogischer Diagnostik dürfen dabei nicht nur die Diagnosen selbst, sondern müssen auch die pädagogischen Anschlusshandlungen betrachtet werden (z. B. Hesse & Latzko, 2017; Schrader, 2009, 2013). Bei der Analyse der *absoluten Lösungsraten der Diagnose- und Empfehlungsentscheidungen* ergab sich erstens, dass die Studierenden bei beiden Entscheidungen noch ein deutliches Entwicklungspotential haben. Bei den Diagnoseentscheidungen lag der mittlere Anteil korrekter Lösungen bei knapp 80 %, was gemessen an der erforderlichen Trefferquote in der Praxis nicht zufriedenstellend ist. Wenn an Diagnostiker*innen die gleichen Anforderungen wie an Tests mit dichotomen Ergebnissen gestellt werden, sollten die Trefferquoten bei mind. 90 % liegen. Das heißt, 90 % der Personen, die eine Lernbesonderheit aufweisen, sollten auch als solche identifiziert werden (Sensitivität; vgl. Faller, 2005). Da die Diagnostik von Lernbesonderheiten den verbindlichen Diagnoseanlässen zuzuordnen ist, kann dieses streng erscheinende Kriterium m. E. hier herangezogen werden. Auch wenn für die Ableitung von Empfehlungen kein solches Kriterium festgesetzt werden kann, ist eine mittlere Trefferquote von knapp 60 % auch hier nicht zufriedenstellend. Bei dieser Einordnung der Trefferquoten muss allerdings bedacht werden, dass diese Ergebnisse nicht direkt auf Trefferquoten für reale Diagnosesituationen übertragen werden können (s. Abschnitt 7.3.4.1). Sie sind zunächst nur als Leistungen in Bezug auf konkrete Entscheidungen in dieser Szenario basierten Umgebung interpretierbar. Erste Vergleiche zwischen der Urteilsgenauigkeit in Szenario basierten Umgebungen und im Feld lassen zwar den Verdacht zu, dass die Urteilsgenauigkeit in komplexitätsreduzierten Umgebungen höher ausfällt (vgl. Kaiser et al., 2015). Dementsprechend könnten die Fähigkeiten der Studierenden hier womöglich überschätzt werden und in der Praxis sogar noch niedriger ausfallen. Diese Schlussfolgerung muss jedoch als vorläufig betrachtet werden. Weitere Studien sind notwendig, um die Übereinstimmung der Urteilsgenauigkeit im Feld vs. Szenario zu ermitteln. Diese Studien würden außerdem Untersuchungen zur Urteilsgenauigkeit bei der Diagnostik von Lernbesonderheiten im Feld voraussetzen, die aktuell noch ein Desiderat sind.

Unabhängig davon, ob die Höhe der absoluten Trefferquoten mit realen Praxisbedingungen übereinstimmt, ist ein zweiter zentraler Befund dieser Studie hervorzuheben, der von dieser Einschränkung unberührt ist: Die *Ableitung von den Empfehlungen* fiel den Studierenden deutlich schwerer als das Abgeben von Diagnosevermutungen. Dieses Ergebnis ist konform mit den Ergebnissen aus der Pilotierung des Fallinventars sowie Befunden einer Studie von Julia Klug und Kolleg*innen (2016), die Fähigkeiten von (angehenden) Lehrkräften in Bezug auf die verschiedenen Phasen des diagnostischen Prozesses erhob. Auch in dieser Studie zeigten die Studierenden deutliche Schwierigkeiten bei der Ableitung pädagogischer Anschlusshandlungen. Obwohl der Diagnosegegenstand (Lernverhalten) sowie der Handlungsrahmen für Anschlusshandlungen (Vorschläge zur Unterstützung von Lernprozessen) dort nicht direkt mit den hiesigen vergleichbar waren, ergab sich übereinstimmend, dass das Ableiten pädagogischer Anschlusshandlungen eine Herausforderung für angehende Lehrkräfte ist. Die Empfehlungen für pädagogische Anschlusshandlungen in der vorliegenden Studie richteten sich auf die Zuweisung zu außerunterrichtlichen Fördermaßnahmen (vgl. Kap. 1.2.1. sowie Kap. 6.4). Hierfür wurden

allgemeingültige Musterlösungen verwendet, wie sie auch in pädagogischen Handreichungen (SenBJF, 2018) zu finden sind, und die entsprechend zum Handlungsrepertoire von Lehrkräften in der Praxis gehören sollten. Dieses Ergebnis ist ein deutlicher Hinweis darauf, dass das Ableiten von Empfehlungen eine pädagogisch-diagnostische Fähigkeit ist, die explizit eingeübt werden sollte.

Um Hinweise auf die Möglichkeiten und Grenzen bei der Förderung dieser Fähigkeiten zu erhalten, können Überlegungen nützlich sein, die *Gründe für das schlechte Abschneiden* der Studierenden bei den Empfehlungsentscheidungen finden. Ein erster Grund dafür, dass die Empfehlungen den Studierenden schwerer fielen als die Diagnosen, könnte zumindest teilweise schlicht die geringere Wahrscheinlichkeit für korrekte Lösungen sein: Während es bei den Diagnosen i. d. R. nur eine richtige Lösung gab⁹, waren bei den Empfehlungen zumeist Muster aus mehreren Antworten korrekt. Diese Möglichkeit der Mehrfachauswahl bei den Antwortoptionen spiegelt wider, dass auch in der Praxis potentiell mehrere korrekte Anschlusshandlungen möglich sind. Die geringeren Lösungsraten sind dementsprechend ökologisch valide, da es auch in der Praxis schwieriger ist, aus den vielen Möglichkeiten eine korrekte Anschlusshandlung auszuwählen. Dennoch ist anzunehmen, dass die Lösungsraten bei Praktiker*innen höher lägen als die der Studierenden. Ein Hinweis darauf sind die Lösungsmuster der Praktiker*innen, die zur Validierung der Musterlösung beitrugen (vgl. Tab. 7.1). Der Abgleich der Lösungsmuster der Studierenden mit dieser Musterlösung ergab eindeutig, dass die Studierenden zu diesem Zeitpunkt ihrer Ausbildung noch nicht in der Lage waren, in gleichem Maße wie die Praktiker*innen richtige Lösungen zu produzieren. Dass Studierende bei der Ableitung von Anschlusshandlungen schlechter abschneiden als praktisch tätige Lehrkräfte, zeigte auch die o. g. Studie von Julia Klug und Kolleg*innen (2016). Ein Grund für diese vergleichsweise schlechten Leistungen der Noviz*innen ggü. den Expert*innen könnte darin begründet sein, dass es den angehenden Lehrkräften eben noch an Praxiserfahrungen mangelt. Praxiserfahrungen spielen zweifelsfrei eine Schlüsselrolle bei der Entwicklung diagnostischer Expertise (Hascher, 2008; Schrader, 2017).

Eine andere Erklärung dafür, dass Empfehlungen eine besondere Herausforderung für die Studierenden waren, könnte ein weiterer Befund dieser Studie sein. Die Untersuchung des *Zusammenhangs der Lösungsraten von Diagnosen und Empfehlungen* ergab, dass die Studierenden diese beiden Schritte offenbar noch unzureichend aufeinander bezogen: Es zeigte sich, dass korrekte Diagnosen nur bedingt mit korrekten Empfehlungen einhergingen. Der moderate Zusammenhang bedeutet einerseits, dass die Studierenden trotz korrekter Diagnoseentscheidung nicht immer die korrekten Empfehlungen ableiteten. Umgekehrt bedeutet dieses Ergebnis auch, dass die Studierenden auch bei inkorrekten Diagnosen zu korrekten Empfehlungen kommen konnten. Dieses Muster könnte ein Hinweis darauf sein, dass bei Diagnosen und Empfehlungen unterschiedliche Verarbeitungsprozesse stattfinden, die sich auf unterschiedliche

⁹ Bei der Gestaltung des Fallinventars wurde darauf geachtet, dass die Phänomene eindeutig sind und nur eine Lösung bei Diagnosen korrekt ist (vgl. Kap. 6.4 und 6.5). In der Praxis ist es bspw. möglich, dass Diagnosen in Bezug auf umschriebene Entwicklungsstörungen schulischer Fertigkeiten gleichzeitig mit Einschränkungen im kognitiven Bereich (bis zu einem gewissen Grad) auftreten (s. DGKJP, 2015; 2018).

Informationen beziehen. Die Frage der Informationsverarbeitung wird weiter unten noch diskutiert, kann jedoch in dieser Studie nicht abschließend beantwortet werden.

Der letzte Aspekt, der hinsichtlich der diagnostischen Produkte diskutiert werden soll, ist die *subjektive Sicherheit beim Treffen der diagnostischen Entscheidungen*. Das Ausmaß der Sicherheit wurde untersucht, um Hinweise auf die subjektiv wahrgenommenen Fähigkeiten der Studierenden zu erhalten. Hierzu wurde der Zusammenhang für die Lösungsraten von Diagnosen und Empfehlungen mit der jeweiligen mittleren Sicherheit beim Treffen dieser Entscheidungen berechnet. Eine positive Korrelation bedeutet hier, dass die Studierenden sich bei falschen Entscheidungen unsicher und umgekehrt bei richtigen Entscheidungen sicher fühlten. Tatsächlich zeigte sich, dass die subjektive Sicherheit kaum mit den erzielten Lösungsraten korrelierte, sondern teilweise sogar relativ stark von diesen abwich (vgl. Fragestellung 1b). Dieses Ergebnis spricht dafür, dass die Studierenden ihre eigenen diagnostischen Fähigkeiten noch nicht realistisch einschätzen können. Dabei kann eine realistische Einschätzung der eigenen Fähigkeiten neben den tatsächlichen Fähigkeiten als Gradmesser der professionsbezogenen Kompetenz von Lehrkräften betrachtet werden (z. B. Thoren et al., 2020; aber siehe Schladitz et al., 2015). Insbesondere bei Lehramtsstudierenden kann allerdings oft beobachtet werden, dass diese Selbsteinschätzung noch nicht wirklichkeitsnah ist, vermutlich deshalb, da es an korrektiven praktischen Erfahrungen mangelt (Müller, 2010). Lernerfahrungen sind allgemein notwendig, um eigene Fähigkeiten realistisch einschätzen zu können – denn Selbsteinschätzungskompetenz muss eben auch erlernt werden (Ihme & Senkbeil, 2017). Vor diesem Hintergrund sprechen die vorliegenden Ergebnisse zur subjektiven Sicherheit erneut dafür, dass die Studierenden noch Lernerfahrungen beim Treffen pädagogisch-diagnostischer Entscheidungen sammeln müssen. Aufschlussreich ist hierbei auch die Betrachtung des Zusammenhangs getrennt nach Reihenfolge-Bedingungen. Hier zeigte sich, dass es den Studierenden unter einer Bedingung doch bereits relativ gut gelang, die Korrektheit ihrer getroffenen Entscheidungen einzuschätzen: Wenn sie zuerst Diagnosen und anschließend Empfehlungen abgaben, also die normative Reihenfolge des diagnostischen Prozesses (Jäger, 2006) befolgten, gingen korrekte Diagnoseentscheidungen recht häufig auch mit höheren Sicherheitsratings einher. In der umgekehrten Reihenfolge ergab sich kein Zusammenhang, und auch für Empfehlungen zeigte sich dieses Muster nicht analog. Dieses Ergebnis spricht dafür, dass die Studierenden zwar auch bei Diagnosen noch Lernerfahrungen benötigen, um die Korrektheit ihrer Entscheidungen besser einschätzen zu können. Bei Empfehlungen erscheint der Bedarf an Lernerfahrungen jedoch noch um einiges höher. Hier zeigte sich in der normativen Bedingung allenfalls ein sehr schwacher Zusammenhang zwischen Lösungsraten und Sicherheitsratings.

Fragestellung 2b. Reihenfolge der pädagogisch-diagnostischen Entscheidungen

Die Ergebnisse, die unter dieser und der folgenden Fragestellung diskutiert werden, widmen sich der Erforschung des diagnostischen Prozesses der Studierenden im engeren Sinne. Hierzu wurde zunächst untersucht, ob die Manipulation des diagnostischen Prozesses einen Einfluss auf die darin ablaufenden Verarbeitungsprozesse hat, die sich wiederum auf die Güte der diagnostischen Produkte niederschlagen sollten (vgl. Behrmann & Van Ophuysen, 2017). Im Rahmen normativer (Hesse & Latzko, 2009, 2017)

sowie heuristischer (Herppich et al., 2017, 2018) Prozessmodelle der Diagnostik bilden Diagnosen die Grundlagen für pädagogische Anschlusshandlungen (s. erweitertes Prozessmodell Kap. 1.2.4.3). Auch im Rahmen von heuristischen Modellen der datengestützten Unterrichtsentwicklung (z. B. Helmke, 2004; Hosenfeld & Groß-Ophoff, 2007; Mandinach & Gummer, 2016) werden pädagogische Anschlusshandlungen auf Basis der Interpretation sowie Integration von (bspw. diagnostischen) Daten abgeleitet. Dieses Vorgehen spiegelt professionelles Verhalten von Lehrkräften wider, das im Laufe der professionellen Entwicklung erlernt werden muss. Wie und wann genau diese Entwicklung erfolgt, ist angesichts weniger Forschungsarbeiten hierzu (aber siehe Klug et al., 2016) aktuell unklar. Die vorliegende Studie sollte daher einen Beitrag leisten, die Befundlage der Entwicklung prozessbezogener Kompetenzen in der ersten Bildungsphase zu erweitern. Untersucht wurde daher, inwiefern Lehramtsstudierende im ersten Mastersemester bereits Anzeichen für dieses professionelle Vorgehen zeigen. Die Reihenfolge der pädagogisch-diagnostischen Entscheidungen wurde hierzu experimentell variiert. Es bestätigte sich, dass die Reihenfolge der beiden Entscheidungen einen deutlichen *Effekt auf die Rate korrekter Empfehlungen* hatte: In der normativen Bedingungen, in der die Empfehlungen vor den Diagnosen abgeleitet wurden, konnten die Studierenden weniger Fälle korrekt lösen als in der Bedingung, in der sie Empfehlungen zuerst vergeben konnten. Für Diagnosen zeigte sich umgekehrt nur eine minimale Tendenz dahingehend, dass in der normativen Bedingung mehr korrekte Lösungen produziert wurden. Bei Diagnosen waren die Lösungsraten also kaum von der Reihenfolge beeinflusst – anders als bei Empfehlungsentscheidungen. Dass die Studierenden mehr Fehler machten, wenn sie Empfehlungen nach der Diagnose ableiteten, passt zu dem Befund, der sich bei der Korrelation der Lösungsraten beider Entscheidungen ergab (Fragestellung 2a). Offenbar konnten die Studierenden von der normativen Reihenfolge noch nicht ausreichend profitieren, weil sie noch nicht in der Lage waren, Anschlusshandlungen auf der Grundlage von Diagnosen abzuleiten. Das könnte ein Zeichen dafür sein, dass die Studierenden zumindest bei den Empfehlungen unterschiedliche Informationsverarbeitungsprozesse wählten, wenn sie diese vor oder nach einer Diagnoseentscheidung ableiten sollten. Diese Vermutung fand Bestätigung bei der Analyse der gewählten Diagnoseinformationen, die im folgenden Abschnitt erläutert wird.

Fragestellung 2c. Nutzung formeller Diagnoseinformationen

Der zweite Aspekt, der hinsichtlich des diagnostischen Prozesses im engeren Sinne untersucht wurde, war die Nutzung der diagnostischen Informationen, die den Studierenden zum Treffen ihrer pädagogisch-diagnostischen Entscheidungen zur Verfügung standen. Es wurde angenommen, dass die Güte der diagnostischen Informationen, die bei der Urteilsbildung einfließen, maßgeblich für die Güte der diagnostischen Produkte ist, die anhand dieser Daten integriert werden (für Diagnosen s. a. Behrmann & Van Ophuysen, 2017). Tatsächlich belegen Studien, die Informationsverarbeitungsprozesse von Noviz*innen und Expert*innen miteinander vergleichen, dass die Urteilsbildung von Noviz*innen sich noch recht stark von verzerrenden Informationen beeinflussen lässt, wobei dies auch teilweise bei erfahrenen Lehrkräften in der Praxis zu beobachten ist (z. B. Kaiser et al., 2015; Krolak-Schwerdt et al., 2009, 2013).

In der vorliegenden Studie wurde untersucht, inwiefern sich die Auswahl diagnostischer Informationen sowohl auf die Lösungsrate der Diagnosevermutungen als auch auf die der Empfehlungen auswirkte. Dabei zeigte sich erstens, dass die Studierenden für beide Entscheidungen überwiegendangaben, formelle Informationen genutzt zu haben. Formelle Informationen waren die Diagnoseinformationen mit der höchsten Güte, die systematisch und unter Nutzung evidenzbasierter Methoden erhoben wurden (vgl. formelle Diagnose; Schrader, 2010; Hascher, 2008). Es zeigte sich zweitens, dass diese *hochwertigen Diagnoseinformationen bei Diagnosen häufiger* genutzt wurden als bei Empfehlungen. Die Informationsauswahl und –verarbeitung gestalteten sich hier also bei beiden Entscheidungen nicht in gleichem Maße: Bei Empfehlungen flossen nach Angaben der Studierenden häufiger informelle und semiformelle Diagnoseinformationen in die Entscheidungsfindung mit ein. Informelle und semiformelle Diagnoseinformationen sind solche, die im Rahmen der Diagnostik von Lernbesonderheiten nicht zur Urteilsfindung relevant sein sollten, da sie sich allgemein weder durch eine hohe Objektivität noch durch eine ausreichende Reliabilität auszeichnen (für eine kritische Bewertung informeller und semiformeller Diagnosen s. Hascher, ebd.). In der vorliegenden Studie war bei einem Teil der Informationen daneben auch die Validität dieser nicht-formellen Informationen verletzt, da sie inkongruent zur formellen Diagnoseinformationen waren. Zu erwarten war also, dass die diagnostischen Entscheidungen häufiger korrekt sein müssten, wenn vermehrt formelle Diagnoseinformationen genutzt würden.

Tatsächlich zeigte sich als drittes Ergebnis zur Informationsnutzung bei *Empfehlungen ein moderater Zusammenhang* zwischen dem Nutzungsanteil formeller Diagnoseinformationen und der Lösungsrate. Je mehr formelle Diagnoseinformationen ein*e Studierende*r also für ihre Empfehlungsentscheidung nutzte, desto mehr korrekte Lösungen konnte sie/er hier tendenziell erzielen. Dieses Ergebnis erweitert Befunde, die belegen, dass die Nutzung irrelevanter Informationen die Urteilsgenauigkeit verzerren kann (z. B. Kaiser et al., 2015). In der vorliegenden Studie zeigte sich, dass die Nutzung irrelevanter Informationen auch mit einer geringeren Lösungsrate bei pädagogischen Anschlusshandlungen einherging. Bei den Diagnosen war dieser Zusammenhang zwar nicht nachweisbar – hier ergab sich nur ein schwacher Zusammenhang für den Nutzungsanteil formeller Diagnoseinformationen und den Lösungsraten der Studierenden. Dieses Ergebnis kann jedoch durchaus durch Deckeneffekte erklärt werden, denn bei Diagnosen waren sowohl die Lösungsraten als auch die Nutzungsanteile formeller Informationen bei der Mehrheit der Studierenden sehr hoch, sodass sich hier weniger Varianz ergab. Dieses Ergebnis könnte dementsprechend auch als Bestätigung der Hypothese verstanden werden: Dass die Diagnoseentscheidungen so häufig korrekt waren, könnte dadurch zustande gekommen sein, dass überwiegend formelle Diagnoseinformationen genutzt wurden.

Der Befund, dass die vermehrte Nutzung formeller Diagnoseinformationen mit mehr korrekten Entscheidungen einherging, ist insbesondere hinsichtlich des Vergleichs der *Nutzungsanteile getrennt für die Reihenfolgebedingungen* des diagnostischen Prozesses hervorzuheben. Diese vierte Analyse ergab, dass sich die Nutzungsanteile der Diagnoseinformationen bei den Empfehlungen in Abhängigkeit von der Reihenfolge durchaus unterschieden: Während formelle Informationen für Empfehlungen besonders

dann deutlich häufiger genutzt wurden, wenn die Empfehlung als erste Entscheidung getroffen wurde (nichtnormative Bedingung), verhielt es sich bei informellen Informationen genau umgekehrt. Informelle Informationen wurden deutlich häufiger genutzt, wenn die Diagnose der Empfehlung vorausging (normative Bedingung). Für Diagnosen ergaben sich indes keine bedeutsamen Unterschiede in den Nutzungsanteilen der Diagnoseinformationen zwischen den beiden Reihenfolge-Bedingungen. Dieser Befund bestätigt damit den Eindruck, der sich bei der Untersuchung der Lösungsraten in Abhängigkeit von der Reihenfolge der Entscheidungen ergab (Fragestellung 2b) – nämlich dass die Abgabe von Diagnosen den Angaben der Studierenden zufolge in beiden Bedingungen hauptsächlich auf den formellen Informationen beruhten. Da die formellen Informationen die relevanten Informationen waren, konnten die Studierenden hier häufiger korrekte Entscheidungen treffen. Bei der Ableitung von Empfehlungen jedoch gestalteten sich die Nutzungsmuster unterschiedlich, wenn diese Entscheidung der Diagnose vorausging oder nachgelagert war. Dass die Studierenden für Empfehlungen in der normativen Bedingung weniger formelle und dafür mehr informelle Informationen zu nutzen angaben, könnte zumindest teilweise erklären, warum die Lösungsraten für Empfehlungen in der normativen Bedingung niedriger ausfielen. Sollte die Empfehlung nach der Diagnose vergeben werden, nutzen die Studierenden vermehrt auch informelle Informationen, jedoch standen diese oftmals in Widerspruch zu den formellen Informationen und konnten daher zu falschen Lösungen führen.

Die Schlussfolgerung, dass durch die Reihenfolge der Entscheidungen *unterschiedliche Informationsverarbeitungsprozesse* ausgelöst werden, kann aufgrund der vorliegenden Daten nicht endgültig geprüft und muss daher als vorläufig betrachtet werden. Um diese Vermutung weitergehend zu untersuchen, sind Folgestudien notwendig, in denen kongruente und inkongruente Informationen systematisch variiert werden. Dabei könnten die tatsächlichen Informationsverarbeitungsprozesse besser analysiert werden, wenn neben Selbstberichtsmaßen auch objektive Messungen zum Einsatz kommen. Insbesondere Blickbewegungsmessungen (Eyetracking) könnten eine geeignete Methode sein, um zu verfolgen, welche Informationen die Probanden wie lange betrachten. Zwar wurde in der vorliegenden Studie darauf geachtet, die Studierenden sehr offen nach den von ihnen genutzten Informationen zu fragen. Die Studierenden wurden nicht instruiert, dass formelle Informationen bevorzugt zu nutzen seien und woran sie diese Informationsart hätten erkennen können. Um die Wahrscheinlichkeit möglicher Wiedererkennungseffekte zu reduzieren, wurde die Position der Informationsarten außerdem variiert. Einem Teil der Studierenden gelang es jedoch gut, formelle Informationen zu erkennen und zu nutzen. Es kann hierbei nicht ausgeschlossen werden, dass die Studierenden zum Teil andere Informationen nutzten, als sie angaben – denn sie hatten im Modul „Pädagogische Diagnostik“ bereits gelernt, dass Werte aus standardisierten Tests bei der Urteilsbildung bevorzugt werden sollten, und könnten hier sozial erwünscht geantwortet haben. Um die Angabe der Studierenden zu validieren, wären o. g. Blickbewegungsmessungen geeignet.

In der vorliegenden Studie kann außerdem nicht geklärt werden, welche *Gründe die Studierenden* dafür hatten, dass sie die Informationsauswahl bei Empfehlungen ändern, wenn diese nach den Diagnosen

abgegeben werden sollten. Für eine umfangreiche Förderplanung in der Praxis wären selbstverständlich neben fähigkeitsbezogenen Informationen auch Informationen zum häuslichen Umfeld, zur Lernmotivation oder zum Fähigkeitsselbstkonzept wichtig (z. B. Heine et al., 2012; Landerl et al., 2017), da diese auch der praktischen Ressourcenplanung dienen. Das war in dieser Szenario basierten Umgebung jedoch gar nicht erforderlich. Und wenn eine derartige Planung erforderlich gewesen wäre, dann hätte sie doch gleichermaßen in beiden Reihenfolgebedingungen erfolgen müssen. Hier war es jedoch so, dass diese zusätzlichen Informationen vor allem in der normativen Bedingung herangezogen wurden.

Möglicherweise ist dieses unsystematische Verhalten darauf zurückzuführen, dass Diagnosen und Empfehlungen noch nicht ausreichend aufeinander bezogen werden (vgl. auch Fragestellung 2a) und daher für Empfehlungen entsprechend vermehrt auch andere Informationen herangezogen wurden als für die Diagnosen. Um diese Vermutung zu stützen, wäre eine differenzierte Analyse der Nutzungsmuster erforderlich. Dabei könnten möglicherweise unterschiedliche Nutzungstypen diagnostischer Informationen herausgestellt werden. Die Augenscheininspektion der Informationsnutzung schien in die Richtung zu deuten, dass ein Teil der Studierenden bevorzugt eine Informationsart nutzte, und andere in ihren Nutzungsmustern inkonsistent waren. Auch eine Befragung der Studierenden zu ihren Intentionen bei der Informationsnutzung wäre hilfreich, um mehr über die Hintergründe der Datennutzung der Studierenden zu erfahren. Hierzu könnte die Methode des lauten Denkens zur Anwendung kommen, die auch mit Blickbewegungsmessungen kombinierbar wäre. Schließlich wäre eine Untersuchung an praktisch tätigen Lehrkräften sicher sehr aufschlussreich, da zum aktuellen Zeitpunkt auch unklar ist, ob sich das Nutzungsverhalten für diagnostische Informationen bei erfahrenen Lehrkräften genauso verhält und ob diese bei der Ableitung von Anschlusshandlungen von einer expliziten Diagnose profitieren könnten oder, wie auch die Studierenden, bei der Nutzung der Diagnose Schwierigkeiten hätten, und ihre Empfehlungen ohne diese besser ausfallen würden.

Zwischenfazit aus Studie 2

Beim Einsatz des *Fallinventars als Szenario basierte Umgebung* zeigte sich, dass den Studierenden die Abgabe von Diagnosevermutungen und Empfehlungen noch Schwierigkeiten bereitet. Insbesondere die Ableitung von pädagogischen Anschlusshandlungen gelang ihnen noch verhältnismäßig schlecht. Diese Ergebnisse standen in Zusammenhang mit der Gestaltung des diagnostischen Prozesses. Die Manipulation der Reihenfolge beider Entscheidungen zeigte, dass die Studierenden bei Empfehlungen noch nicht gut genug von der normativen Reihenfolge der Schritte im diagnostischen Prozess profitieren. In diesem Zusammenhang wurde auch offensichtlich, dass die Informationsnutzung der Studierenden noch nicht systematisch genug erfolgt, um für beide Entscheidungen in beiden Bedingungen vergleichbare Lösungsraten zu erzielen. Außerdem gelang es den Lehramtsstudierenden noch nicht ausreichend, ihre Diagnosevermutungen für die Ableitung von Anschlusshandlungen heranzuziehen. Diese Ergebnisse unterstützen erstens die Annahme, dass die Güte von Diagnosen und Empfehlungen in engem Zusammenhang mit Variablen des diagnostischen Prozesses steht (vgl. Behrmann & Van Ophuysen, 2017). Sowohl die Variation der Abfolge diagnostischer Schritte als auch die Analyse der

Informationsnutzung erwiesen sich als fruchtbare Untersuchungsansätze und könnten in Folgestudien dazu beitragen, diagnostische Produkte in Zusammenhang mit Variablen des diagnostischen Prozesses weiter zu explorieren. Für den Diagnoseanlass der Feststellung von Lernbesonderheiten ist mit dieser Studie ein Anfang gemacht.

Die Ergebnisse sprechen zweitens dafür, dass Lehramtsstudierende noch deutlichen Unterstützungsbedarf haben, um ihre prozessbezogenen Fähigkeiten weiterzuentwickeln. Dabei sollten nicht nur die diagnostischen Leistungen, sondern auch selbsteingeschätzte Fähigkeiten bedacht werden. Zur Förderung dieser Fähigkeiten könnte das *Fallinventar als Lerngelegenheit* einen wichtigen Beitrag leisten, da es zum Bedarf der Studierenden passt und sie darin unterstützen kann, die o. g. Schwierigkeiten zu reflektieren und zu überwinden. Die hier vorgestellten ersten Ergebnisse zur Eignung des Fallinventars als Lerngelegenheit deuten darauf hin, dass das Fallinventar zumindest eine Lerngelegenheit ist, die bei den Studierenden auf positive Resonanz stößt. Ferner belegen die Ergebnisse hier, dass alle konstruierten Fallvignetten grundsätzlich geeignet sind und daher im Fallinventar verwendet werden können. In der dritten Studie dieser Arbeit wird das Fallinventar als Lerngelegenheit systematisch eingesetzt und evaluiert (Kap. 8). Dort wird untersucht, welchen Beitrag diese praxisorientierte Übung leisten kann, um die Studierenden auf ihre pädagogisch-diagnostischen Aufgaben im Zusammenhang mit der Feststellung von Lernbesonderheiten vorzubereiten.

8 Evaluation des diagnostischen Fallinventars als fallbasierte Lerngelegenheit (Studie 3)

8.1 Hintergrund und Überblick

Im vorausgehenden Kapitel 7 (Studie 2) wurde das Fallinventar als szenariobasierte Umgebung eingesetzt, die der Erforschung der Fähigkeiten der Studierenden bei der Vergabe von Diagnosevermutungen und Empfehlungen für pädagogische Anschlusshandlungen diente. Dabei zeigte sich, dass die Studierenden bei beiden Entscheidungen, insbesondere aber bei der Ableitung von Empfehlungen, deutliche Schwierigkeiten hatten. Diese Schwierigkeiten konnten damit in Zusammenhang gebracht werden, dass die Studierenden noch besser auf die *Durchführung des diagnostischen Prozesses* vorbereitet werden müssen: Den Studierenden mangelte es offenbar an prozeduralem Wissen darüber, auf welche Weise Diagnosen zur Ableitung von Empfehlungen genutzt werden können und wie formelle diagnostische Informationen systematisch zu einer hohen Treffsicherheit bei beiden Entscheidungen beitragen können.

In der vorliegenden Studie wird geprüft, inwiefern das diagnostische Fallinventar zum *Aufbau prozeduralen Wissens* im Kontext der Feststellung von Lernbesonderheiten beitragen kann. Anders als in Studie 2 erfolgte der Einsatz des Fallinventars hier in einer Lernsituation, in der die Studierenden klare Instruktionen, Übungsbeispiele sowie Hilfestellungen während der Bearbeitung erhielten. Der Beitrag, den das Fallinventar für den Erwerb diagnostischen Wissens leistete, wurde im Vergleich zu einer konventionellen aufgabenbasierten Übung evaluiert, welche wiederum dem Aufbau deklarativen Wissens diente. Dabei stand die Frage im Zentrum, welchen spezifischen Beitrag zum Wissenserwerb jede der beiden Übungen leisten und ob das Fallinventar eine sinnvolle Ergänzung oder sogar ein Ersatz für die konventionelle Übung sein kann. Neben diagnostischem Wissen wurden auch affektiv-emotionale Aspekte bei der Evaluation berücksichtigt, da beide Aspekte diagnostischer Kompetenz Voraussetzungen für professionelles diagnostisches Verhalten sind (Kompetenz allgemein: Blömeke et al., 2015; diagnostische Kompetenz im Speziellen: Herppich et al., 2018).

8.2 Aufbau diagnostischen Wissens bei angehenden Lehrkräften

Der Erwerb von Wissen steht im Allgemeinen im Vordergrund der universitären Ausbildungsphase und ist im Speziellen ein zentraler Aspekt bei der Entwicklung diagnostischer Expertise (Hascher, 2008; s. a. Baumert & Kunter, 2006; vgl. Kap. 3). Das zentrale Anliegen bei der Entwicklung des Fallinventars war, angehende Lehrkräfte bereits im Lehramtsstudium auf eine anspruchsvolle diagnostische Aufgabe vorzubereiten, für die eine fundierte Wissensbasis besonders wichtig ist – die Feststellung von Lernbesonderheiten (vgl. Kap. 6). Zur Feststellung von Lernbesonderheiten ist insbesondere Wissen über die Durchführung eines systematischen, *methodisch kontrollierten diagnostischen Prozesses notwendig*, denn die Diagnostik von Lernbesonderheiten ist ein verbindlicher Diagnoseanlass mit weitreichenden Folgen für die betroffenen Schüler*innen (Karst et al., 2017). Die Abgabe von Förderempfehlungen setzt daher eine formelle Diagnose voraus (Hesse & Latzko, 2017; Schrader, 2010; vgl. Kap. 7). Neben dem Wissen, das

zur Durchführung der einzelnen Phasen des diagnostischen Prozesses notwendig ist (Auflistung einzelner Wissensaspekte: z. B. Jäger, 2006; s. a. Kap. 3), brauchen Lehramtsstudierende auch Wissen, um aus den Diagnosen entsprechende Anschlusshandlungen abzuleiten (Auflistung von Kompetenzfacetten zur Nutzung diagnostischer Daten zur Unterrichtsentwicklung; Mandinach & Gummer, 2016).

In Bezug auf die Bewältigung der verschiedenen Phasen des diagnostischen Prozesses weisen angehende Lehrkräfte gegenüber praktisch tätigen Lehrkräften geringere Fähigkeiten auf (Klug et al., 2016): Die Schwierigkeiten der Lehramtsstudierenden beziehen sich dabei auf alle Phasen des Prozesses. Besonders große Probleme bereiten ihnen die methodisch begründete Auswahl von diagnostischen Instrumenten sowie die Interpretation und Nutzung der diagnostischen Informationen zur Entscheidungsfindung. Auch in der zweiten Studie der vorliegenden Arbeit konnten weitere Hinweise darauf gefunden werden, dass die Studierenden formelle diagnostische Informationen noch nicht ausreichend gut nutzen können, um daraus treffsichere Diagnosen und passende pädagogische Anschlusshandlungen abzuleiten (vgl. Kap. 7).

Ursächlich für diese Schwierigkeiten könnte ein *Mangel an Wissen* zur Durchführung jener Phasen des diagnostischen Prozesses sein. Genau an diesem Punkt setzt das diagnostische Fallinventar an: Diese fallbasierte Lerngelegenheit will insbesondere dazu beitragen, den Aufbau von Wissen zu unterstützen, das den systematischen Umgang mit diagnostischen Daten sowie die regelgeleitete Durchführung diagnostischer Prozesse ermöglicht (vgl. auch Thoren et al., 2020).

Dieses prozessbezogene Wissen ist für die Entwicklung diagnostischer Expertise essentiell (Hascher, 2008). Diagnostische Expertise kann umschrieben werden als die Kompetenz, diagnostische Daten zu erheben, zu interpretieren und zu nutzen (vgl. Helmke, 2009). Hierzu bedarf es Wissens dazu, welche Diagnosen mithilfe welcher diagnostischen Daten auf welche Weise erhoben werden können und was sich daraus ableiten lässt (vgl. Hascher, 2008). Dieses Wissen wird auch als *prozedurales diagnostisches Wissen* bezeichnet. Prozedurales Wissen beinhaltet Kenntnisse über Prozeduren und Methoden, die zur systematischen Lösung von – hier diagnostischen – Problemen erforderlich sind (für den hochschuldidaktischen Kontext vgl. Glessmer & Lüth, 2016). Prozedurales Wissen ist außerdem ein handlungsbezogenes Wissen, das durch häufige Anwendung und Wiederholung entsteht, und schließlich automatisiert und implizit abgerufen werden kann (kognitionspsychologisches Verständnis prozeduralen Wissens, vgl. Stern et al., 2016). Damit prozedurales Wissen sinnvoll eingesetzt und auch auf unbekannte Probleme übertragen werden kann, muss es mit konzeptuellem Wissen verknüpft werden (vgl. Kap. 3).

Konzeptuelles Wissen ist verbalisierbares Wissen über Zusammenhänge von grundlegenden Fakten und Begriffen, das ein tiefergehendes Verständnis von theoretischem Grundlagenwissen ermöglicht (Stern et al., 2016; Glessmer & Lüth, 2016). Konzeptuelles Verständnis entsteht durch Vernetzung, Umstrukturierung und Re-Organisation von Wissen (vgl. Kap. 3.2.3). Auf diese Weise entsteht grundlegendes, umfassendes theoretisches Wissen, das auf praktische Probleme angewendet werden kann. Auf diese Weise können Theorie und Praxis verknüpft werden – ein wichtiges Anliegen bei der Lehrkräftebildung, das in den letzten fünf Jahren im deutschen Raum besonders stark vorangetrieben wird (BMBF, 2016). Um Lehramtsstudierenden diese Theorie-Praxis-Verknüpfungen zu ermöglichen, kommen

in jüngerer Zeit zunehmend fallbasierte Lerngelegenheiten zum Einsatz. Die Förderung diagnostischer Kompetenz im Rahmen fallbasierter Lerngelegenheiten scheint ersten empirischen Befunden zufolge besonders gut zu gelingen, wenn in einem ersten Schritt konzeptuelles Wissen etabliert und gefestigt wird, und dieses Wissen anschließend fallbasiert zur Anwendung kommt (vgl. Glogger-Frey & Renkl, 2017; Herppich et al., 2017a).

Ein großer Vorteil fallbasierter Lerngelegenheiten liegt darin, dass sie eine schrittweise Annäherung an die Praxis erlauben, indem die Komplexität realer praktischer Probleme sowie der Zeitdruck, der mit deren Lösung i. d. R. verbunden ist, reduziert werden können (Approximation-to-Practice Framework: Grossman et al., 2009; vgl. Kap. 6). Bereits kurze und schematische schriftliche Fallvignetten sind dabei geeignet, um realistische Szenarien von Schüler*innen entstehen zu lassen, die realitätsnahe diagnostische Verarbeitungsprozesse auslösen (z. B. Schülerinventar; Kaiser et al., 2015). Auch die ersten Erfahrungen mit dem diagnostischen Fallinventar haben gezeigt, dass die Studierenden die Übung als realitätsnah und hilfreich für die spätere Berufspraxis empfanden. Die Studierenden setzten sich zudem intensiv mit den Fällen auseinander und gerieten bei der Auswertung in angeregte Diskussionen darüber, welche Lösungen für die jeweiligen Schüler*innen richtig seien. Diese Beobachtungen sprechen für eine relativ hohe ökologische Validität, die im Rahmen szenariobasierter Umgebungen erreicht werden kann (vgl. Praetorius et al., 2017). Fallbasierte Lerngelegenheiten scheinen also besonders geeignete Möglichkeit zu sein, prozedurales diagnostisches Wissen im Lehramtsstudium zu fördern.

8.3 Kognitive Veränderungen und weitere Zielvariablen in der Lehrkräftebildung

Die mehr oder weniger praxisnahe Bearbeitung diagnostischer Probleme im Rahmen fallbasierter Lerngelegenheiten kann – neben der Anbahnung prozeduralen Wissens – auch dazu beitragen, professionelle Einstellungen zur Diagnostik zu entwickeln (Wedel & Pfetsch, 2017) und auf diese Weise für eine professionelle pädagogische Diagnostik zu motivieren (vgl. Westphal et al., 2018 zur Nutzung datenbasierter Rückmeldungen). Diese Wirkung fallbasierter Lerngelegenheiten könnte vermittelt sein durch die Freude beim Lernen mit Praxisfällen sowie dem Gefühl, dass diese Übung für die Berufspraxis relevant sein wird – zwei wichtige Wirkfaktoren, damit Lerngelegenheiten auch tatsächlich Lernprozesse anregen (vgl. Lipowsky, 2014; Kittel et al., 2018). Diese Reaktionen der Teilnehmer*innen sind neben den ausgelösten Lernprozessen zwei von vier Indikatoren, an denen die Wirksamkeit einer Intervention für (angehende) Lehrkräfte festgemacht werden kann. Die vier Ebenen, die im Rahmen von Angebot-Nutzungs-Modellen für (Fort-)Bildungsangebote für Lehrkräfte Beachtung finden, werden nachfolgend vorgestellt.

8.3.1 Vier Ebenen der Evaluation für Interventionen in der Lehrkräftebildung

Im Rahmenmodell von Lipowsky (2014, S. 401), das sich an Lehrkräften in der Praxis orientiert, wird der *Rezeption der Lerngelegenheit* durch die Teilnehmer*innen ein großer Stellenwert für den Erfolg einer

Intervention beigemessen. Die Rezeption ist auf der ersten Ebene der Wirkung anzusiedeln und umfasst z. B. Akzeptanz und Zufriedenheit mit der Fortbildung. Jene –bestenfalls positiven – Wahrnehmungen des Bildungsangebots interagieren mit Voraussetzungen der Lehrkräfte, wie z. B. deren Vorwissen oder Überzeugungen, und wirken sich auf diese Weise auf die Teilnahme- und Lernmotivation der Teilnehmer*innen aus. Diese Motivation wird außerdem vermittelt durch den subjektiv wahrgenommenen Wert der Maßnahme für die Lehrkraft, der sich durch Interesse sowie empfundene Nützlichkeit und Relevanz der Maßnahme ausdrückt. Die Teilnahme- und Lernmotivation wiederum wirkt sich Lipowsky zufolge auf die Transfermotivation aus, welche sich schließlich darauf auswirkt, ob die Teilnehmer*innen das Gelernte in der Praxis anwenden. Dieser gewünschte Erfolg der Maßnahme wird auf dieser ersten Ebene allerdings noch nicht erfasst.

Messbar werden die Effekte einer Lerngelegenheit erst auf der zweiten Ebene im Modell, die *Veränderungen hinsichtlich kognitiver und affektiv-motivationaler Voraussetzungen* der Lehrkräfte abbildet. Zu den kognitiven Voraussetzungen zählt Lipowsky professionelles Wissen sowie Überzeugungen, zu den affektiv-motivationalen Aspekten rechnet er beispielsweise Selbstwirksamkeitserwartungen. Auf der zweiten Ebene wird zusammengefasst, ob die Maßnahme bei den Teilnehmer*innen einen Lernprozess auslösen konnte, der sich in Form von Wissenszuwachs, einer professionellen Haltung oder der Annahme, in der Zukunft professionelles Verhalten zeigen zu können, ausdrückt.

Ob die (Fort-)Bildungsmaßnahme sich tatsächlich auf das Verhalten der Teilnehmer*innen auswirkt, wird erst auf der dritten Ebene fokussiert. Dort wird analysiert, inwiefern das Gelernte in die Praxis transferiert werden kann und ob es zu *Änderungen im unterrichtspraktischen Handeln* kommt. Auf der vierten und letzten Ebene werden die Effekte auf Seiten der Schüler*innen angesehen, die von den Lehrkräften unterrichtet werden. Ob eine Maßnahme den *Schulerfolg der Schüler*innen*, gemessen an Schulleistungen sowie motivationalen Aspekten, positiv beeinflusst, wird allerdings selten analysiert.

Die vier vorgestellten Ebenen strukturieren verschiedene Aspekte professioneller Kompetenz, die im Rahmen der Lehrkräftebildung berücksichtigt werden können (vgl. auch Angebot-Nutzungs-Modell: Kunter et al., 2011). Sie formulieren verschiedene Erfolgskriterien, die dafür sprechen, dass ein Lernangebot tatsächlich von den Teilnehmer*innen genutzt wird. Es eignet sich daher als Rahmenmodell zur Evaluation von Interventionen, bei dem die Ebenen als *hierarchisch aufeinander aufbauende Evidenzstufen* zur Untersuchung der Wirksamkeit von Maßnahmen in der Erwachsenenbildung betrachtet werden können (Naugle et al., 2000; Kirkpatrick & Kirkpatrick, 2006; Naugle et al., 2006): Positive Reaktionen (Ebene 1) sind die notwendige Voraussetzung dafür, dass Interventionen überhaupt wirken können, da sie die Bereitschaft der Teilnehmenden, die Trainingsinhalte zu erlernen, schaffen. Der Lernprozess selbst (Ebene 2) ermöglicht eine Verhaltensänderung, sobald die Teilnehmer*innen einer Intervention einen Zuwachs an Kompetenzen erleben und über das Wissen verfügen, diese auch in der Praxis anzuwenden. Dieser Transfer des Erlernten, der sich in Form professionellen Verhaltens (Ebene 3) manifestiert, ist wiederum die Voraussetzung dafür, dass sich Ergebnisse im System zeigen: Veränderungen bei den Schülerinnen und Schülern, die von den trainierten Lehrkräften unterrichtet werden, sind das höchste Ziel

von Interventionen für Lehrkräfte, weil sie einerseits für eine große Reichweite der Maßnahme sprechen. Andererseits stellt diese Evidenzstufe auch die höchste methodische Anforderung an die Evaluation, wie der nachfolgende Abschnitt verdeutlicht.

8.3.2 Evaluation von Interventionen für Lehramtsstudierende: Kognitive und affektiv-emotionale Voraussetzungen im Zentrum

Die vier genannten Ebenen, die zur Strukturierung von Evaluationsstudien herangezogen werden können, sind mit *aufsteigenden methodischen Anforderungen* verbunden (Naugle et al., 2006): Die Erfassung der Rezeption der Teilnehmer*innen auf Ebene 1 bietet eine niedrighschwellige Möglichkeit der Erfassung von Interventionseffekten, da hier Selbstberichte der Teilnehmer*innen ausreichen. Die Veränderungsprozesse auf Ebene 2 hingegen sollten mittels Selbsteinschätzungs- und Einstellungsskalen sowie mithilfe standardisierter Instrumente erfasst werden, die bestenfalls vor und nach dem Training vorgegeben werden, um den Veränderungsprozess bestmöglich abbilden zu können. Die Einschätzung des Transfers des Gelernten auf Ebene 3 erfolgt in der Regel über die Erfassung des Verhaltens der Trainierten vor und nach der Intervention. Dieser Vergleich sollte unter Beachtung einer Kontrollgruppe realisiert werden. Am aufwändigsten und gleichzeitig am aussagekräftigsten ist es, wenn Verhaltensänderungen der Empfänger*innen des Verhaltens der Trainierten auf Ebene 4 nachgewiesen werden könnten, was beispielsweise anhand der Verbesserung der Leistungen oder der Lernmotivation der von Lehrkräften unterrichteten Schüler*innen angezeigt werden könnte.

Anhand dieser Auflistung wird deutlich, dass die Erfassung aller vier Ebenen nur bei praktisch tätigen Lehrkräften möglich ist, die Verhalten im Unterricht zeigen können (Ebene 3), welches wiederum Effekte auf deren Schüler*innen hat (Ebene 4). Die fallbasierte Lerngelegenheit, die in der vorliegenden Studie evaluiert wird, richtet sich an *Lehramtsstudierende zu Beginn des Masterstudiums*, die keinen Zugang zur diagnostischen Praxis haben und daher nicht in der Praxis handeln können. Bei der Evaluation von Interventionen für Lehramtsstudierende sind daher insbesondere die beiden ersten Ebenen im Fokus. Die zweite Ebene ist hierbei die zentrale Ebene der Evaluation, da sie die Voraussetzungen kompetenten Verhaltens abbildet (vgl. Kompetenzmodell Blömeke et al., 2015; Kap. 3). Auf dieser Ebene stehen kognitive und affektiv-motivationale Aspekte diagnostischer Kompetenz im Vordergrund. Nachfolgend werden Möglichkeiten der Erfassung dieser zwei zentralen Voraussetzungen diagnostischer Kompetenz skizziert.

Zur *Erfassung der kognitiven Voraussetzungen* von diagnostischen Kompetenzen haben sich Wissenstests etabliert, die das diagnostische Grundlagenwissen abfragen (z. B. Klug et al., 2013; Linninger et al., 2015). In diesen klassischen Wissenstests wird in der Regel deklaratives Wissen in Multiple-Choice-Formaten abgefragt (vgl. auch Kap. 7). Eine Erweiterung dieser klassischen Wissenstests sind sogenannte *Szenario-Tests*, in denen die Wissensabfrage in realistische Fallschilderungen eingebettet ist (für eine Übersicht s. Praetorius et al., 2017). Insbesondere für außerunterrichtliche Diagnoseanlässe kann eine relativ hohe ökologische Validität angenommen werden, bei der gleichzeitig auch die interne Validität nicht verletzt

wird (vgl. Kap. 7). Das Verhalten, das die Studierenden im Rahmen von Szenario-Tests zeigen, kann daher valide Hinweise auf diagnostisches Verhalten in der Praxis geben (Klug et al., 2016). Szenariobasierte Tests eignen sich daher besonders zur Erfassung praxisrelevanten Anwendungswissens und das Verhalten, das Studierende in einem Szenariotest zeigen, ist ein guter Schätzer für die Professionalität des Verhaltens, das Lehrkräfte in der realen Berufspraxis auch zeigen würden (Klug et al., 2013; 2016).

Neben diagnostischem Wissen geben auch affektiv-motivationale Faktoren bzw. Voraussetzungen (Blömeke et al., 2015; Lipowsky, 2014) Hinweise darauf, inwiefern Lehrkräfte professionelles Verhalten zeigen: Diese affektiv-emotionalen Voraussetzungen tragen dazu bei, dass das vorhandene Wissen in konkreten Situationen in professionelle Wahrnehmungs-, Interpretations- und Entscheidungsprozesse überführt werden kann, die schließlich anhand kompetenter Handlungen sichtbar werden (vgl. Blömeke et al., 2015). Auch wenn die empirische Evidenz zu diesen Faktoren noch recht schmal und die Erweiterung der Befundlage daher ein Desiderat ist, können einige Prädiktoren professionellen Verhaltens benannt werden, die i. d. R. in Form von Selbstberichten erfasst werden. Annika Ohle und Kolleg*innen (2015) konnten zeigen, dass sich die Zeit, die Lehrkräfte mit diagnostischen Tätigkeiten verbringen, durch diagnostikbezogene Einstellungen, Selbstwirksamkeitsüberzeugungen sowie Selbstreflexion in Bezug auf Diagnostik vorhersagen lässt. Andrea Westphal und Kolleginnen (2018) fanden bei der Suche nach Faktoren für eine Praxis evidenzbasierter Unterrichtsentwicklung Hinweise darauf, dass insbesondere die Motivation zur Nutzung eines Diagnoseinstruments (z. B. VERA) sowie die Einstellung zu Diagnostik im Allgemeinen einen entscheidenden Einfluss darauf haben, ob datenbasierte Rückmeldungen für die Unterrichtsentwicklung genutzt werden. Julia Klug und Kollegen (2016) ermittelten mithilfe eines Szenariotests, der prozessbezogene Kompetenzen bei der Diagnostik von Lernprozessen untersucht, dass sich für unterschiedliche Phasen der Professionalisierung unterschiedliche Prädiktoren benennen lassen: Während es bei Studierenden insbesondere das diagnostische Wissen war, das eine Vorhersage der Leistung im Szenariotest erlaubte, konnte die Einstellung bei Referendar*innen vorhersagen, in welchem Ausmaß diagnostisch kompetent gehandelt wird. Bei erfahrenen Lehrkräften wiederum waren die Motivation sowie das vorhandene Wissen positive Prädiktoren. Diese Befunde sprechen dafür, dass bei der Untersuchung diagnostischer Kompetenzen neben kognitiven Voraussetzungen auch affektive und motivationale Aspekte untersucht werden sollten. Diese bieten sich nicht nur aufgrund ihres prädiktiven Wertes für professionelles diagnostisches Verhalten an, sondern auch, weil sie in Form von Fragebögen relativ niedrighschwellig erhoben werden können.

8.4 Evaluation des diagnostischen Fallinventars

Im Wintersemester 2017/18 kam das diagnostische Fallinventar zur Förderung diagnostischer Expertise von Lehramtsstudierenden im Modul „Pädagogische Diagnostik“ zum Einsatz. Hierbei sollte primär untersucht werden, welchen Beitrag diese fallbasierte Übung zum Aufbau prozeduralen Wissens leisten kann. Kontrastiert wurden die Effekte auf die kognitiven Voraussetzungen mit denen einer klassischen aufgabenbasierten Übung, die vor der Entwicklung des Fallinventars im Modul eingesetzt wurde. Die aufgabenbasierte Übung sollte deklaratives Grundlagenwissen vermitteln, das im Fallinventar praxisnah

zur Anwendung kommt. Bei der Untersuchung der übungsspezifischen Effekte auf deklaratives bzw. prozedurales Wissen wurden die Ergebnisse einer Expert*innen-Befragung berücksichtigt, die der Validierung der Wissensarten in den Lerngelegenheiten sowie den Wissenstests diente. Ergänzt wurde die Untersuchung der übungsspezifischen Effekte auf diagnostisches Wissen durch die Analyse der Rezeption der beiden Lerngelegenheiten (Ebene 1 nach Lipowsky, 2014) sowie die Wirkungen auf weitere Voraussetzungen kompetenten diagnostischen Verhaltens (Ebene 2).

8.4.1 Fragestellungen und Hypothesen

8.4.1.1 Kognitive Prozessdimensionen der Lerngelegenheiten

Fragestellung 1a. Inhaltliche Validierung der Lerngelegenheiten

Das erste Ziel der Untersuchung zur Lerngelegenheit war die Validierung der einzelnen Übungen in Bezug auf die intendierten kognitiven Prozessdimensionen. Damit sind zum einen die *Wissensarten* gemeint, auf die die jeweiligen Übungen fokussieren: Die fallbasierte Übung mit dem Fallinventar wurde konzipiert, um maßgeblich prozedurales diagnostisches Wissen zur Feststellung von Lernbesonderheiten zu fördern. Die aufgabenbasierte Übung sollte das hierfür notwendige Grundlagenwissen liefern, das in Form von Fakten- und Konzeptwissen etabliert wird. Der Aufbau prozeduralen Wissens sollte hierdurch vorbereitet werden.

Für einen strukturierten Wissensaufbau ist neben der Betrachtung der Wissensarten auch die Analyse von *Lernzielebenen* relevant, um die Studierenden systematisch beim Wissenserwerb zu unterstützen (vgl. Anderson et al., 2001; Glessmer & Lüth, 2016; Stern et al., 2016). Glessmer & Lüth (ebd.) definieren innerhalb jeder Wissensart sechs Lernziele, deren Komplexität sukzessiv zunimmt, wodurch die notwendigen kognitiven Prozesse zur Lösung der entsprechenden Aufgaben zunehmend schwieriger werden (vgl. Kap. 3.2.3.2, Tab. 3.1). Um Hinweise zur möglichen Verarbeitungstiefe der Lerngelegenheiten zu erhalten, sollten die einzelnen Aufgaben der Übungen ebenfalls hinsichtlich der Lernzielebenen eingestuft werden.

Fragestellung 1b. Inhaltliche Validierung des Wissenstests

Schließlich beziehen sich hochschuldidaktische Überlegungen zur Validität auch darauf, inwiefern die in den Übungen intendierten Wissensaspekte auch in den Prüfungsaufgaben abgebildet werden (vgl. Kap. 3.2; Glessmer & Lüth, 2016; Halbherr et al., 2016). Um in der vorliegenden Studie eine Passung zwischen Übungs- und Prüfungsaufgaben herzustellen, war auch von Interesse, ob der eingesetzte *Wissenstest* die in den Übungen angesprochenen Wissensarten und Lernzielebenen in gleicher Weise abbildet. Daher wurden auch die Aufgaben des Wissenstests einem Expert*innen-Gremium vorgelegt.

8.4.1.2 Effekte der Lerngelegenheiten auf die Studierenden

Das zweite zentrale Ziel war es zu überprüfen, inwiefern die einzelnen Übungen zum Kompetenzaufbau der Studierenden beitragen können. Oberstes Interventionsziel des Fallinventars war es, den Aufbau prozeduralen Wissens bei den Studierenden anzuregen sowie weitere Aspekte diagnostischer Kompetenz

zu fördern. Diese Lerneffekte sowie Veränderungen hinsichtlich affektiv-motivationaler Variablen werden auf Ebene zwei des Angebot-Nutzungs-Modells von Lipowsky (2014) evaluiert. Da ein Kompetenzzuwachs nur dann zu erwarten ist, wenn eine Lerngelegenheit auf positive Resonanz stößt, wurde zunächst die Rezeption (Ebene 1) sowie Indikatoren für Lernzuwachs (Ebene 2) untersucht. Hierzu wurden die Wirkungen der fallbasierten Übung mit denen der aufgabenbasierten Übung verglichen (Fragestellung 2a). Anschließend wurden die Antworten der Studierenden, die jeweils eine bzw. beide Lerngelegenheiten absolvierten, hinsichtlich kognitiver und affektiv-motivationaler Aspekte analysiert (Fragestellung 2b).

Fragestellung 2a. Vergleich der Lerngelegenheiten (Ebene 1 und 2)

Evaluationsebene 1 (Rezeption). In einem ersten Schritt wurden die Voraussetzungen untersucht, die Lernprozesse sowie den Aufbau von Kompetenzen in einer Lerngelegenheit ermöglichen: die Reaktionen der Studierenden auf die Lerngelegenheit (Evaluationsebene 1, vgl. Lipowsky, 2014; Kirkpatrick & Kirkpatrick, 2006). Hierzu wurden die aufgabenbasierte und die fallbasierte Übung hinsichtlich kognitiver sowie affektiver Aspekte von den Studierenden bewertet. Es wurde vermutet, dass sich die *Rezeption des Fallinventars* auf beiden Dimensionen von der Wahrnehmung der aufgabenbasierten Übung unterscheidet. Hinsichtlich der kognitiven Dimension wurde angenommen, dass sich der empfundene Nutzen für die Klausur sowie die spätere Berufspraxis unterscheiden, und zwar kongruent zu den Überzeugungen der Studierenden, welchem Übungszweck die jeweilige Übung diene: Es wurde angenommen, dass beim Fallinventar der Nutzen für die spätere Berufspraxis größer empfunden wird, weil diese Übung mehr Praxisnähe aufweist, und umgekehrt die aufgabenbasierte Übung als relevanter für die modulabschließende Klausur empfunden wird, da sie eine Übung zur Vertiefung theoretischen und methodischen Grundlagenwissens darstellt. Diese Kongruenz zwischen den Annahmen der Studierenden und den Trainingsinhalten kann die Akzeptanz einer Intervention steigern und sich positiv auf den empfundenen Nutzen auswirken (vgl. Tillema, 1995). Dementsprechend wurde auch vermutet, dass das Fallinventar auf der affektiven Dimension positiver bewertet wird als die aufgabenbasierte Übung: Als eine praxisnahe Lerngelegenheit empfinden die Studierenden bei der Bearbeitung des Fallinventars vermutlich mehr Freude und Interesse, weil es als berufsvorbereitend erlebt wird (Rothland, 2019).

Evaluationsebene 2 (Lernen). Eine weitere Frage beim Vergleich der fallbasierten mit der aufgabenbasierten Übung war, ob sich das subjektive Kompetenzerleben während der Übungen unterschied. Das Kompetenzerleben sollte exploriert werden, um Hinweise auf die intrinsische Bearbeitungsmotivation (vgl. Ryan & Deci, 2000; Kittel & Rollett, 2018; Kittel et al., 2017) sowie einen Eindruck darüber zu erhalten, ob die Studierenden durch die Bearbeitung der Übung an Sicherheit gewinnen konnten, um Aufgaben der gleichen Art auch in Zukunft bewältigen zu können (vgl. Kap. 3.3.2). Diese Variablen dienten als Indikatoren für einen potentiellen Lernzuwachs. Außerdem wurde untersucht, ob die Studierenden durch die Bearbeitung der Übungen einen subjektiven Lernzuwachs erlebten. Die Abschätzung des Lernzuwachses über Selbsteinschätzungen war wichtig, da das Setting, in

dem die Studie durchgeführt wurde, keine Prä-Post-Messung erlaubte und Vorher-Nachher-Vergleiche auf der Wissensdimension entsprechend nicht möglich waren.

Fragestellung 2b. Vergleich der Interventionsgruppen (Ebene 2)

Kognitive Voraussetzungen. Neben diesen subjektiven Maßen zum Lernzuwachs und Kompetenzerleben bei den Übungen wurde auf der zweiten Ebene untersucht, inwiefern die Übungen Effekte auf die tatsächlichen Fähigkeiten der Studierenden haben. Die zentrale Frage der Studie, ob die Übungen den Wissensaufbau der Studierenden in unterschiedlicher Weise unterstützen, wurde mit einem diagnostischen Wissenstest untersucht. Dabei wurde erwartet, dass die Bearbeitung der *aufgabenbasierten Übung* die Studierenden vor allem beim Aufbau deklarativen Wissens unterstützen kann, da diese Übung die Bildung von Begriffsnetzwerken und damit konzeptuelles Wissen fördern sollte (Stern et al., 2016). Die Arbeit mit dem Fallinventar in der fallbasierten Übung hingegen sollte vor allem prozedurales Wissen fördern, da hier eine Wiederholung eines sorgfältigen diagnostischen Vorgehens anhand verschiedener Fälle zu einer Prozeduralisierung führen kann (Stern et al., 2016; Hetmanek & Van Gog, 2017). Zur Evaluation der Effekte auf der zweiten Ebene wurde zusätzlich eine dritte Gruppe von Studierenden analysiert, die beide Übungen hintereinander absolvierte. Die Einführung dieser *kombinierten Gruppe* sollte es ermöglichen, die spezifischen Beiträge der einzelnen Übungen auf die kognitiven Voraussetzungen von multiplikativen Effekten abzugrenzen, die im Zusammenspiel beider Übungen entstehen könnten: Es wurde angenommen, dass die Kombination aus aufgabenbasierter und fallbasierter Übung die günstigsten Bedingungen für den Wissenserwerb schaffen kann, da sich die Konsolidierung deklarativen Wissens in Kombination mit einer anschließenden Anwendung dieses Wissens bereits in anderen Interventionsstudien als erfolgreich für die Förderung prozeduralen diagnostischen Wissens erwies (vgl. Tutoring-Studie: Herppich et al., 2017a; beispielbasiertes Kurztraining: Glogger-Frey & Renkl, 2017). Generell entwickelt sich prozedurales Wissen am besten auf der Basis (Glessmer & Lüth, 2016) bzw. im Zusammenspiel mit dem Aufbau konzeptuellen Wissens (Stern et al., 2016). Beide Übungen hintereinander zu bearbeiten könnte die Studierenden also besonders gut dabei unterstützen, tiefgreifendes konzeptuelles Wissen aufzubauen, welches durch die bewusste, wiederholte Anwendung des Prozedurenwissens der Studierenden fördert.

Affektiv-motivationale Voraussetzungen. Für die kombinierte Gruppe wurde darüber hinaus erwartet, dass die affektiv-motivationalen Voraussetzungen für kompetentes diagnostisches Verhalten durch die Verknüpfung beider Übungen am besten gefördert werden können. Untersucht wurden hierzu diagnostikbezogene Einstellungen sowie diagnostikbezogene Selbsteinschätzungen. Sowohl von Einstellungen als auch von Selbstwirksamkeitswartungen kann angenommen werden, dass diese positiv mit der Wissensbreite von Lehrkräften korreliert sind (Ernest, 1989; Kunter et al., 2011). Dementsprechend sollte die kombinierte Gruppe, da sie nach beiden Übungen über eine breitere Wissensbasis verfügen sollte, auch hinsichtlich der affektiv-motivationalen Aspekte höhere Werte aufweisen. Zweitens kann vermutet werden, dass auch die Gruppe der Studierenden, die nur mit dem Fallinventar arbeitete, nach dieser Übung auf der affektiv-motivationalen Dimension höher punktet als

diejenigen Studierenden, die nur mit der aufgabenbasierten Übung arbeiteten: Fallbasierte Lerngelegenheiten sind vermutlich besonders gut geeignet, um die Motivation für eine professionelle Diagnostik zu erhöhen, da sie mehr Lernfreude erzeugen (vgl. Wedel et al., 2020). In der fallbasierten Übung können die Studierenden außerdem erste praxisnahe Erfahrungen sammeln. Eigene Erfahrungen und Praxiskonfrontationen sind ein entscheidender Faktor für Einstellungsänderungen von Lehrkräften (für inklusionsbezogene Einstellungen: Bosse & Spörer, 2014), und möglicherweise zeigt sich ein ähnlicher Effekt bereits bei praxisnahen Erfahrungen. Durch die praktische Anwendung von testtheoretischen und methodischen Grundlagen beim Fallinventar könnte die Akzeptanz standardisierter Tests steigen und sich entsprechend Einstellungen gegenüber Tests oder Diagnostik im Allgemeinen erhöhen – eine wichtige Voraussetzung für die Entwicklung diagnostischer Expertise (Helmke, 2004; 2009).

8.4.2 Methoden

Zunächst werden die Inhalte und Methoden der aufgabenbasierten sowie der fallbasierten Übung vorgestellt. Den Beschreibungen der Interventionsinstrumente folgt die Beschreibung der Methoden zur Validierung der Lerngelegenheiten (Fragestellung 1, Abschnitt 8.4.2.2). Anschließend wird das methodische Vorgehen bei der Evaluation der Lerngelegenheiten vorgestellt (Fragestellung 2, Abschnitt 8.4.2.3).

8.4.2.1 Instrumente der Intervention

Aufgabenbasierte Übung

Im Rahmen der aufgabenbasierten Übung wurden klassische Übungsaufgaben präsentiert, die den Aufbau deklarativen Wissens unterstützen sollen. Diese konventionelle Übung wurde eingesetzt, um Fakten-, Begriffs- und Konzeptwissen zu ausgewählten diagnostischen Themen zu etablieren, zu vertiefen und zu vernetzen. Im Zentrum der Übung standen testtheoretische Grundlagen standardisierter Tests, die Interpretation von Normwerten sowie Diagnosekriterien für umschriebene Entwicklungsstörungen schulischer Fertigkeiten („Lernstörungen“, z. B. Lese-Rechtschreibstörung) und Besonderheiten hinsichtlich kognitiver Lernvoraussetzungen von Schüler*innen („Intelligenz“, „Hochbegabung“, „Lernbehinderung“). Mit den Übungsaufgaben erhielten die Studierenden die Gelegenheit, die Inhalte aus der Vorlesung zu wiederholen und zu vertiefen. Die Übungseinheit umfasste insgesamt sechs Aufgabenblöcke, die die Studierenden überwiegend in Partnerarbeit bearbeiteten. Die Lösungen wurden aufgabenweise direkt nach der Bearbeitung im Plenum diskutiert. In Tabelle 8.1 wird ein Kurzüberblick über die einzelnen Teilaufgaben der Übung gegeben. Eine Darstellung der didaktischen Planung der Einheit (Wißmann & Kinder, 2018), die Aufgabenstellungen sowie die entsprechenden Musterlösungen können auf der Homepage des Arbeitsbereiches eingesehen werden (K2teach, 2018).

Im ersten Teil der Übungseinheit (Aufgaben 1 - 3) wurde die Anwendung von Kennwerten der gängigen Standardnormwertskalen (z-Skala, T-Skala, IQ-Skala) sowie Prozenträngen eingeübt. Neben der Einordnung beispielhafter Standardnormwerte in die Standardnormalverteilung sollten verschiedene Werte in Normwerte anderer Skalen umgerechnet werden. Im zweiten Teil der Übung (Aufgaben 4a - d)

wurde die Anwendung von Diagnosekriterien zur Feststellung von Lernbesonderheiten wie z. B. Rechenstörung oder intellektueller Hochbegabung thematisiert. Hierzu wurde den Studierenden eine Tabelle mit fiktiven Ergebnissen aus einem Intelligenztest, einem Lese-, einem Rechen- sowie einem Rechtschreibtest präsentiert. Die Aufgabe der Studierenden bestand darin, diese fiktiven Befundmuster mit jeweils einer der spezifischen Lernbesonderheiten in Zusammenhang zu bringen sowie Förderschwerpunkte zu benennen, die für das jeweilige fiktive Muster im Rahmen des schulischen Unterrichts angemessen wären. Nach der Berechnung von Diskrepanzen zwischen einzelnen Teilleistungen und Intelligenzwerten (Aufgabe 4e und 5) waren die Studierenden im letzten Teil der Übung (Aufgabe 6) dazu aufgefordert, verschiedene Lernbesonderheiten in eigenen Worten zu definieren und Implikationen für die Unterrichts- und Schulgestaltung abzuleiten. Dieser letzte Teil der Übungseinheit war als Zusatzaufgabe konzipiert, die als freiwillige Hausaufgabe beendet werden konnte.

Tabelle 8.1 Kurzbeschreibung der Inhalte der aufgabenbasierten Übung

Aufgabe	Kurzbeschreibung
1a	Abbildung einer Standardnormalverteilung interpretieren
1b	Merkmale von Standardnormwerten (3 Teilaufgaben)
2	Interpretation von Normwerten in Bezug zur Standardnormalverteilung
3	Umrechnung von Normwerten
4a	Begriff der Diskrepanz in Bezug auf Teilleistungen (2 Teilaufgaben)
4b	Bedeutung des Diskrepanzkriteriums bei der Diagnostik von Lernstörungen
4c	Leistungsprofile von Schüler*innen: Diagnosevermutungen, Förderempfehlungen (2 Teilaufgaben)
4d	Bedeutung des Diskrepanzkriteriums bei der Diagnostik von Hochbegabung (2 Teilaufgaben)
4e	Berechnung einer Leistungsdiskrepanz
5	Bedeutung des Konfidenzintervalls von Testwerten in Bezug auf Hochbegabung
6	Lernbesonderheiten in eigenen Worten (Rollenspiel: Erklärung für Eltern; 2 Teilaufgaben)

Die Aufgaben dieser Übung wurden seit einigen Semestern zur Vorbereitung der Abschlussprüfung im Modul „Pädagogische Diagnostik“ eingesetzt und für den Einsatz in dieser Studie geringfügig überarbeitet sowie um die sechste Aufgabe ergänzt, die als didaktische Reserve diente. Auch wenn die Übung klausurrelevant für die Studierenden sein sollte, schilderten die Dozierenden im Modul, dass die Aufgaben von den Studierenden mit wenig Interesse oder Freude bearbeitet wurden. Die Befürchtung der Dozierenden war daher, dass die erarbeiteten Inhalte nach der Klausur schnell in Vergessenheit geraten könnten, da sie für die Praxis nicht als relevant erachtet würden. Um den Praxisbezug zu verstärken und um positive affektive Reaktionen der Studierenden hinsichtlich der Diagnostik von Lernbesonderheiten zu evozieren, kam die fallbasierte Lerngelegenheit zum Einsatz.

Fallbasierte Übung

Das Herzstück der fallbasierten Übung bildete das Fallinventar (Kap. 6.). Die Studierenden bearbeiteten hier acht realitätsnahe Fallvignetten von Schüler*innen mit und ohne Lernbesonderheiten, die auf je einer Seite in einem Fallheft beschrieben waren (vgl. Abb. 6.2). Die Fälle waren so zusammengestellt, dass

zunächst Fälle mit hohen Lösungswahrscheinlichkeiten bearbeitet werden sollten, zur Mitte hin nahmen die Lösungswahrscheinlichkeiten ab und zum Ende der Übung wieder zu (vgl. Tab. 6.6). Damit sollte gewährleistet werden, dass die Studierenden verschiedene Schwierigkeitsniveaus bearbeiten, auch wenn sie nicht alle Fälle in der Seminarsitzung schaffen. Die Fallgeschichten enthielten je drei Informationen zu den Rechen-, Lese- und kognitiven Fähigkeiten der fiktiven Schüler*innen, die unterschiedlich gut zur Diagnostik von Lernbesonderheiten geeignet und teilweise untereinander widersprüchlich waren. Formelle Diagnoseinformationen (z. B. Ergebnisse teil- und standardisierter Schulleistungstests) waren gegenüber semiformalen (z. B. selbstentwickelter Leistungstest einer Lehrkraft) sowie informellen Informationen (z. B. beiläufige Gespräche) zu bevorzugen (zur Unterscheidung von Diagnosen nach Formalisierungsgraden vgl. Hascher, 2008). Die Studierenden wurden instruiert, dass sie diagnostische Informationen unterschiedlicher Güte erhielten. Aufgabe der Studierenden war es, auf Grundlage der gegebenen Informationen eine Diagnosevermutung abzugeben, und entsprechend der Diagnosevermutung eine Förderempfehlung abzuleiten. Dieses Vorgehen wurde zu Beginn von den Dozierenden an einem Fall demonstriert, anschließend wurde ein weiterer Fall im Plenum erarbeitet. Die verbleibenden Fälle bearbeiteten die Studierenden in Partnerarbeit in zwei Blöcken mit jeweils max. drei Fällen. Nach jedem Block wurden die Lösungen im Plenum diskutiert. Dabei wurde auch diskutiert, inwiefern die nichtformellen Diagnoseinformationen bei der Lösungsfindung berücksichtigt werden könnten, und welche zusätzlichen Informationen oder diagnostischen Schritte nötig wären, um die fiktiven Schüler*innen bestmöglich in ihrer Entwicklung zu unterstützen.

8.4.2.2 Validierung hinsichtlich kognitiver Prozessdimensionen

Validierung der Lerngelegenheiten (Fragestellung 1a)

Um Hinweise darauf zu erhalten, inwiefern die Übungen geeignet sind, die angestrebten kognitiven Prozessdimensionen zu fördern, wurde eine inhaltliche Validierung durch ein Expert*innen-Gremium vorgenommen. Dem Gremium gehörten Hochschuldozierende an, die die Lehre im Modul „Pädagogische Diagnostik“ aktiv gestalteten und mit beiden Lerngelegenheiten bereits Erfahrungen im Rahmen ihrer Seminare gesammelt hatten ($N = 5$). Die Befragung der Expert*innen fand im Wintersemester 2018/19 statt und erfolgte in zwei Schritten: Zunächst wurden die Dozierenden gebeten, die einzelnen Aufgaben der Übungen für sich einzuordnen und diese Einordnung zu notieren. Anschließend wurde in Form eines offenen Diskurses eine einheitliche Lösung auf Basis der individuellen Einordnungen angestrebt. Dieses interaktive Rating nahm insgesamt drei Sitzungen à 90 Minuten in Anspruch, da die Zuordnungen durchaus kontrovers diskutiert wurden. Am Ende wurde für jede Aufgabe ein Konsens gebildet, der bestenfalls von allen, mindestens aber von drei Expert*innen (absolute Mehrheit) geteilt wurde.

Die Einordnung der Aufgaben erfolgte auf Grundlage der hochschuldidaktischen Adaption der vielfach verwendeten *Lernzieltaxonomie* von Anderson et al. (2001) durch Glessmer & Lüth (2016). Diese berücksichtigt zum einen drei *Wissensarten* (Fakten-, Konzept- und Prozedurenwissen) und zum anderen sechs *Lernzielebenen*, die hierarchisch aufeinander aufbauende Verarbeitungsstufen abbilden (vgl. Tab. 3.2). Um eine Wissensart nachhaltig zu fördern, sollten die kognitiven Verarbeitungsprozesse entsprechend

tiefgehend sein. Die Expert*innen wurden gebeten, jede Teilaufgabe der aufgabenbasierten sowie der fallbasierten Übung hinsichtlich der beiden kognitiven Prozessdimensionen einzuordnen. Auf diese Weise sollte eine Einschätzung darüber möglich sein, ob die Verarbeitungstiefe der Lerngelegenheiten ausreicht, um die intendierten Wissensarten zu fördern. Um die Einordnung der Aufgaben zu erleichtern, wurde ein Entscheidungsschema entwickelt, das den Expert*innen zur Verfügung gestellt wurde (s. Tab. 8.2).

Tabelle 8.2 Schema zur Entscheidung über Lernzielebenen

<i>Lernzielebene</i>	<i>Voraussetzung zur Lösung einer Aufgabe</i>					
	Aufgabe und Lösung bekannt	Antwort direkt ablesbar	Lösungsweg vorgegeben	Lösungskriterien vorgegeben	Wissensstrukturen bekannt	Wissen neu strukturieren
VI Erschaffen	--	--	--	--	--	✓
V Evaluieren	--	--	--	--	✓	
IV Analysieren	--	--	--	✓		
III Anwenden	--	--	✓			
II Verstehen	--	✓				
I Erinnern	✓					

Anmerkungen: Im Schema wird ersichtlich, dass die Voraussetzungen, die zur Lösung einer Aufgabe gegeben sind, auf den höheren Lernzielebenen komplexer werden (vgl. Glessmer & Lüth, 2016). Die Häkchen (✓) bedeuten, dass die Lösung einer Aufgabe auf der jeweiligen Lernzielebene unter den oben stehenden Voraussetzungen möglich ist. Wenn die Lernenden beispielsweise eine Aufgabe lösen, bei der sowohl die Aufgabe als auch die Lösung bereits aus der Lehrveranstaltung bekannt sind, reicht zur Lösung dieser Aufgabe aus, dass sich die Studierenden erinnern. Die Lösung dieser Aufgabe findet auf der niedrigsten Lernzielebene erreicht statt und die höheren Anforderungen in den nachfolgenden grauen Feldern sind zur Lösung dieser Aufgabe irrelevant. Die Striche (--) bedeuten, dass die Voraussetzungen für die jeweilige Lernzielebene nicht gegeben sind und die Aufgabe entsprechend höhere Anforderungen an die Studierenden stellt. Um eine Aufgabe der Lernzielebene II zu lösen, reicht bloßes Erinnern nicht mehr aus, da die Aufgabe sowie die Lösung nicht bekannt sind. In diesem Fall ist die Antwort allerdings direkt ablesbar, daher sind keine höheren Anforderungen zur Lösung der Aufgabe gefragt. Auf der höchsten Lernzielebene VI werden die höchsten Anforderungen an die Studierenden gestellt: Hier müssen sie Wissen neu strukturieren und alle anderen Voraussetzungen sind nicht mehr gegeben.

Validierung des Wissenstests (Fragestellung 1b)

Unter Zuhilfenahme dieses Entscheidungsschemas ordneten die Expert*innen ebenfalls die Aufgaben des Wissenstests ein, der zur Evaluation der kognitiven Voraussetzungen der Studierenden zum Einsatz kam. Hierbei sollte überprüft werden, ob die Übungs- und Prüfungsaufgaben hinsichtlich der Wissensarten und Lernzielebenen zueinander passen. Diese Passung wurde angestrebt, da der Wissenstest erstens das zentrale Evaluationskriterium dieser Studie war und daher inhaltlich valide die Wissensaspekte der Lerngelegenheiten erfassen sollte. Zweitens diente er den Studierenden selbst zur Reflektion ihres Wissensstandes in Bezug auf die Diagnostik von Lernbesonderheiten. Die Ergebnisse dieses Tests geben den Studierenden also relevante Hinweise darauf, welche Aspekte sie zur Klausurvorbereitung noch aufarbeiten sollten. Eine Übersicht über die im Wissenstest abgefragten Inhalte kann im Anhang 3.A eingesehen werden.

8.4.2.3 Evaluation der Lerngelegenheiten (Fragestellung 2)

Stichprobe

Die Stichprobe zur Evaluation der Lerngelegenheit setzte sich aus insgesamt $N = 280$ *Lehramtsstudierenden* der Freien Universität Berlin zusammen, die überwiegend im ersten Fachsemester eines Master-of-Education-Studienganges ($N = 249$; 88.9 %) studierten oder sich am Ende des Bachelorstudienganges befanden und die Veranstaltung innerhalb ihres Studienverlaufsplanes vorzogen ($N = 29$). Von allen Teilnehmer*innen der Studie waren etwa ein Drittel ($N = 103$) Studierende der Grundschulpädagogik und die verbleibenden Studierenden strebten das Lehramt für Gymnasien oder integrierte Sekundarstufen an. Etwa zwei Drittel der Studierenden ($N = 188$) waren weiblich. Gut die Hälfte der Studierenden war 20 - 25 Jahre alt ($N = 153$; 26 - 30 Jahre: $N = 56$; > 30 Jahre: $N = 66$). Etwa die Hälfte der Studierenden ($N = 147$) verfügte bereits über berufspraktische Erfahrungen, und nunmehr eine Minderheit gab an, bereits Erfahrungen mit standardisierten Testverfahren gemacht zu haben ($N = 25$; 8.9 %).

Ein Teil der Befragung wurde auch den *Dozierenden in den Seminaren* vorgelegt ($N = 5$). Die Dozierenden machten Angaben zu ihrem subjektiven Eindruck hinsichtlich der Wirkungen der Übungen. Sie gaben an, wie sie die Rezeption der Studierenden bei der Bearbeitung der Übung erlebt haben (Evaluationsebene 1) und inwiefern sie einen Lernzuwachs bei den Studierenden vermuten (Evaluationsebene 2). Außerdem wurden sie gebeten, für jedes Seminar anzugeben, ob die geplante Übung vollständig durchgeführt werden konnte, wie die Resonanz im Seminar war und ob es Besonderheiten bei der Durchführung im jeweiligen Kurs gab. Hierüber wurde die Implementierung der Übungen überprüft.

Instrumente der Evaluation

Die zur Evaluation der Interventionseffekte verwendeten Indikatoren und Instrumente orientierten sich an den ersten beiden Ebenen des Angebot-Nutzungs-Modells für Interventionen für Lehrkräfte (Lipowsky, 2014; vgl. auch Kirkpatrick & Kirkpatrick, 2006). Auf beiden Ebenen wurden vornehmlich selbstentwickelte Instrumente verwendet, die die spezifischen Aspekte diagnostischer Kompetenz konkret für die Diagnostik von Lernbesonderheiten abbildeten (vgl. Überlegungen Kap. 6.3). Die selbstentwickelten Skalen wurden vor dem Einsatz zu Evaluationszwecken pilotiert (Wintersemester 2016/17). Zur Absicherung der Reliabilität wurde für die vorliegende Erhebung für jede Skala die interne Konsistenz überprüft. Eine Übersicht über die eingesetzten Evaluationsinstrumente sowie die Reliabilität der Skalen wird in Tab. 8.3 (Fragestellung 2a) sowie Tab. 8.4 (Fragestellung 2b) angegeben. Die Skalen- und Itemstatistiken sowie die konkreten Itemformulierungen sind in Anhang 3.B einzusehen.

Fragestellung 2a. Vergleich der Lerngelegenheiten

Zum Vergleich der Lerngelegenheiten auf der *ersten Evaluationsebene* („*Reaktion*“) wurden drei Maße erhoben, die den subjektiven Eindruck und die Bewertung der jeweiligen Übung durch die Teilnehmer*innen erfassen. Auf der kognitiven Ebene wurde erfasst, inwiefern die Studierenden die jeweilige Übung als relevant und nützlich empfanden – zum einen für die Klausur im Modul, und zum anderen für die spätere Berufspraxis (2x2 Items). Auf der affektiven Ebene wurden das persönliche Interesse an den Übungsinhalten sowie die Freude beim Bearbeiten der Übung erfragt (je ein Item). Die Items für beide Aspekte wurden in Anlehnung an ein Evaluationsinstrument von Katharina Thoren und Kolleg*innen (Thoren et al., 2020) formuliert, das diese vier Erwartungs-mal-Wert-Aspekte als relevante Bewertungsdimensionen einer Lerngelegenheit zur Förderung von Datennutzungskompetenzen definiert (in Anlehnung an Wigfield & Eccles, 2002).

Der Vergleich der Lerngelegenheiten auf der *zweiten Evaluationsebene* („*Lernen*“) erfolgte zum einen anhand von Indikatoren für potentiellen Lernzuwachs. Hierzu sollte das Kompetenzerleben während der Übung sowie für die Bearbeitung ähnlicher Aufgaben in der Zukunft (je ein Item) angegeben werden. Zudem wurde der selbstwahrgenommene Lernzuwachs über drei Items erfasst, die eine Veränderung der Wissenstiefe und –breite sowie die Verbesserung der eigenen Beurteilungskompetenz beschreiben.

Tabelle 8.3 Indikatoren für den Vergleich der Lerngelegenheiten (Fragestellung 2a)

Instrumente und Items	Inhalt (Anzahl der Items)	Range der Antworten
<i>Ebene 1 (Rezeption)</i>		
Subjektive Bewertung kognitiv -- Klausur	Relevanz und Nützlichkeit der Übung für die Modulabschlussprüfung (2 Items)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)
Subjektive Bewertung kognitiv -- Praxis	Relevanz und Nützlichkeit der Übung für die spätere Berufspraxis (2 Items)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)
Subjektive Bewertung affektiv	Persönliches Interesse und Freude an Übungsinhalten (2 Items)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)
<i>Ebene 2 (Lernen)</i>		
Kompetenzerleben aktuell	Empfundene Leichtigkeit bei der Bearbeitung der Übung (1 Item)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)
Kompetenzerleben prognostisch	Angenommene Bewältigung solcher Übungen in der Zukunft (1 Item)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)
Subjektiver Lernzuwachs	Empfundene Vertiefung und Erweiterung des Wissens, subjektive Verbesserung der Beurteilungskompetenz (3 Items)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)

Anmerkungen. Die Items zum Vergleich der Lerngelegenheiten wurden auch den Dozierenden vorgegeben. Diese wurden gebeten, die mittlere Zustimmung der Studierenden auf Grundlage ihrer Beobachtungen bei der Durchführung der Übungen im Seminar einzuschätzen.

Fragestellung 2b. Vergleich der Interventionsgruppen

Kognitive Voraussetzungen. Zur Erfassung der Effekte auf der *zweiten Evaluationsebene (Lernen)* wurde erstens ein zweidimensionaler Wissenstest eingesetzt, der pädagogisch-diagnostisches Wissen zur Feststellung von Lernbesonderheiten und zur Ableitung entsprechender Anschlusshandlungen erfasst (vgl. Kap. 6.3). Dieser Wissenstest bestand aus zwei Teilen: Der Teil zum Grundlagenwissen fragte verbalisierbares Fakten- und Konzeptwissen ab, das zur Bearbeitung der aufgabenbasierten Übung explizit notwendig ist sowie in der fallbasierten Übung implizit zur Anwendung kommt. Das deklarative Wissen bezieht sich auf drei Themen des Moduls „Pädagogische Diagnostik“ („kognitive Lernvoraussetzungen“, „Lernstörungen“ und „standardisierte Tests“; vgl. Anhang 3.A). Die Fragen wurden überwiegend im Multiple-Choice-Format gestellt und stammten aus einem Set von Selbst-Testaufgaben, die im Modul optional zur Klausurvorbereitung eingesetzt wurden. Die Items des Wissenstests wurden im Wintersemester 2016/17 an größeren Stichprobe von Studierenden im Modul „Pädagogische Diagnostik“ ($N = 258$) pilotiert und selektiert (29 Items mit Schwierigkeitsindizes $.20 \leq p \leq .80$ und Trennschärfen $r_{it} \geq .30$)¹⁰.

Der Anwendungsteil war ein Szenariotest, der aus früheren Klausuraufgaben des Moduls konstruiert wurde. Hier wurde pädagogisch-diagnostisches Anwendungswissen erfasst, indem die Studierenden insgesamt sechs fiktive kurze Fallgeschichten bearbeiteten, in denen Schülerinnen und Schüler mit Lernbesonderheiten beschrieben wurden¹¹ (vgl. Anhang 3.A). Neben einer Skizzierung der Leistungen in verschiedenen Unterrichtsfächern wurde für jeden Fall eine tabellarische Zusammenstellung von Ergebnissen aus standardisierten Tests vorgegeben, die Angaben zu schulischen Fähigkeiten (Lesen, Rechtschreiben und Rechnen) sowie allgemeinen kognitiven Fähigkeiten enthielten. Auf Basis dieser Informationen sollten die Studierenden zunächst eine Diagnose vergeben und anschließend eine mögliche Förderempfehlung für die/den fiktive/n Schüler/in vorschlagen. Die Vergabe der Diagnose erfolgte im Multiple-Choice-Format, wobei die Antwortalternativen hier für jeden Fall entsprechend der simulierten Lernbesonderheit angepasst wurden. Die Ableitung der Fördermaßnahme erfolgte anschließend im offenen Antwortformat. Anders als in der fallbasierten Übung sind in den hier verwendeten Fallbeispielen ausschließlich relevante Informationen zusammengestellt.

Die Punktevergabe im Wissenstest erfolgte anhand einer zuvor erstellten Musterlösung. Dabei wurde jede korrekte Antwort mit jeweils einem Punkt bewertet. Für alle nicht bearbeiteten Aufgaben wurden 0

¹⁰ Die Skala „Grundlagenwissen“ enthielt ursprünglich 29 Items. Die interne Konsistenz dieser Skala mit der vollen Itemanzahl war zufriedenstellend (Cronbach's $\alpha = .78$). Zwei Aufgaben mit je vier Items wurden aufgrund zu extremer Schwierigkeitsindizes ($p \leq .1$ oder $p \geq .9$) sowie zu geringer Trennschärfen ($r_{it} \leq .1$) von den Analysen zur Evaluation ausgeschlossen. Die ausgeschlossenen Items stammten aus den Themenbereichen „Intelligenz“ und „Lernstörungen“, bei denen Diagnosen zu Fähigkeitsprofilen zugeordnet werden sollten. .

¹¹ Die Skala „Anwendungswissen“ enthielt 12 Items. Die interne Konsistenz war für Gruppenvergleiche ausreichend (Cronbach's $\alpha = .66$).

Punkte vergeben, es sei denn, dass der Abbruch der Bearbeitung ab einer bestimmten Aufgabe erkennbar war – in diesem Falle wurden die unbearbeiteten Aufgaben als fehlende Werte behandelt.

Zur Überprüfung der Validität des Wissenstests wurde – neben dem bereits beschriebenen Expertenrating (vgl. Abschnitt 8.4.2.2) – anhand der Lösungsraten in der vorliegenden Studie die Faktorenstruktur des Wissenstests analysiert. Dabei ergab sich eine von der ursprünglichen Konzeptualisierung abweichende Zuordnung der Items zu den Subskalen: Die Items zu den Diagnosevermutungen aus dem Anwendungsteil wurden den Ergebnissen zufolge demselben Faktor wie die Items des Grundlagenteils zugeordnet (vgl. Abschnitt 8.4.3.1). Dieser Faktor wird nachfolgend als „deklaratives Wissen“ bezeichnet. Die Items zu den Empfehlungen im Szenariotest werden der Dimension „prozedurales Wissen“ zugeordnet. Diese Subskalen des Wissenstests wurden in untenstehender Zusammenfassung der Evaluationsinstrumente (Tab. 8.4) übernommen. Eine Kurzbeschreibung der einzelnen Items beider Subskalen findet sich in Anhang 3.B (deklaratives Wissen: Tab. 3.B.2.6; prozedurales Wissen: Tab. 3.B.2.7).

Affektiv-motivationale Voraussetzungen. Zur Analyse affektiv-motivationaler Voraussetzungen diagnostisch kompetenten Verhaltens wurden diagnostikbezogene Selbsteinschätzungen sowie diagnostikbezogene Einstellungen erhoben (vgl. Tab. 8.4). *Diagnostikbezogene Selbsteinschätzungen* wurden einerseits mit der Skala selbsteingeschätzter diagnostischer Kompetenz (adaptiert nach Kunter et al., 2014) erhoben, die die subjektive Fähigkeit zur Abgabe genauer Leistungsurteile sowie zur Beurteilung der Aufgabenschwierigkeit misst. Die Skala wurde um zwei Items ergänzt (vgl. Tab.3.B.2.2 im Anhang 3.B) und die vierstufige Rating-Skala auf sechs Stufen erweitert. Zum anderen wurde die selbsteingeschätzte diagnostische Expertise erfasst, die den Umgang mit diagnostischen Verfahren und Daten abbildet. Diese Skala prozessbezogener diagnostischer Fähigkeiten wurde auf Grundlage einer Auflistung von Kompetenzfacetten nach Helmke (2009) entwickelt.

Diagnostikbezogene Einstellungen wurden einerseits über die Einstellung zu Diagnostik im Allgemeinen erhoben, die Aussagen wie „Diagnostik ist einer der wichtigsten Bestandteile der täglichen Arbeit einer Lehrkraft“ enthielt. Zur Vermeidung von Antworttendenzen wurden die Items hier abwechselnd positiv (4 Items) und invers formuliert (5 Items; z. B. „Für die Unterrichtsgestaltung kann man meist auf Diagnostik verzichten“). Die inversen Items wurden vor der Auswertung rekodiert, sodass hohe Skalenwerte eine hohe positive Einstellung gegenüber diagnostischen Tätigkeiten im Unterricht anzeigen. Ergänzend hierzu wurde die Einstellung zu standardisierten Tests erfasst. Eine hohe positive Einstellung zu standardisierten Tests sollte sich bei hoher Zustimmung auf Items wie „Alle Lehrkräfte sollten, wann immer möglich, auf standardisierte Verfahren zurückgreifen“ zeigen. Auch hier wurden zur Reduzierung von Antworttendenzen abwechselnd positive wie inverse Items (3 Items; z. B. „Die Verwendung standardisierter Tests im Schulalltag ist unnötig“) verwendet, wobei inverse Items auch hier vor der weiteren Auswertung rekodiert wurden.

Tabelle 8.4 Instrumente zur Analyse der Interventionsgruppen (Fragestellung 2b)

Instrumente und Items	Inhalt (Anzahl der Items)	Range der Antworten	Cronbach's α
<i>Ebene 2 (Lernen) – kognitive Voraussetzungen</i>			
Wissen deklarativ	Grundlagenwissen zur Diagnostik von Lernbesonderheiten (21 Items) sowie Wissen um Diagnosen im Rahmen kurzer Fallgeschichten ¹ (6 Items)	0 – 27 Punkte	.82
Wissen prozedural	Vergabe von Empfehlungen ¹ anhand kurzer Fallgeschichten (6 Items)	0 – 6 Punkte	.70
<i>Ebene 2 (Lernen) – affektiv-motivationale Voraussetzungen</i>			
Selbsteingeschätzte Diagnostische Kompetenz ²	Selbsteingeschätzte Kompetenz bei der Einschätzung von Schüler*innenleistungen und Schwierigkeit von Aufgaben (9 Items)	1 – 6 (kann ich gar nicht, kann ich sehr gut)	.83
Selbsteingeschätzte Diagnostische Expertise ³	Selbsteingeschätzte Kompetenz bei der Auswahl, Auswertung und Interpretation diagnostischer Daten (9 Items)	1 – 6 (kann ich gar nicht, kann ich sehr gut)	.78
Einstellung zu Diagnostik	Überzeugung, dass Diagnostik eine wichtige und interessante Aufgabe für Lehrkräfte ist (9 Items)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)	.73
Einstellung zu standardisierten Tests	Überzeugung, dass standardisierte Tests in der Schulpraxis unverzichtbar sind (9 Items)	1 – 6 (ich stimme gar nicht zu, ich stimme voll zu)	.83

Anmerkungen: ¹ Diagnosevermutungen und Empfehlungen der Studierenden wurden mit Hilfe eines Szenariotests erfasst. ² adaptiert nach Kunter et al. (2014). ³ Eigenentwicklung in Anlehnung an Helmke (2009)

Design und Datenerhebung

Die Durchführung der Übungen der Lerngelegenheit (Paper-Pencil-Format) sowie die darin eingebettete Datenerhebung (schriftliche Paper-Pencil-Befragung) erfolgten im Januar des Wintersemesters 2017/18. Zu diesem Zeitpunkt hatten die Studierenden das Modul „Pädagogische Diagnostik“, bestehend aus einer Vorlesung und einem Begleitseminar, zu etwa 2/3 abgeschlossen. Im Rahmen eines *Interventions-Kontrollgruppen-Planes mit einmaliger Postmessung* umfasste der gesamte Komplex aus Übungen und Erhebung drei aufeinanderfolgende Seminarsitzungen, d. h. insgesamt drei Wochen. Während die Übungen jeweils eine ganze Seminarsitzung à 90 min füllten, nahm die Datenerhebung für die Evaluation ca. 60 min in Anspruch. Um die Effekte der einzelnen Übungen von denen der Kombination beider Übungsformen unterscheiden zu können, wurden die Studierenden entsprechend ihrer Seminare verschiedenen Untersuchungsgruppen zugeordnet, in denen die Reihenfolge der drei Sitzungen variiert wurde.

Das Evaluationsdesign beinhaltete drei Bedingungen: zwei Interventionsgruppen und eine Kontrollgruppe (vgl. Tab. 8.5). Die Studierenden in der Gruppe „Aufgaben“ (Kontrollgruppe) erhielten zunächst nur die konventionelle aufgabenbasierte Übung, bevor in der zweiten Sitzung die Erhebung stattfand, und

schließlich erst in der dritten Sitzung die fallbasierte Übung vorgegeben wurde. Um den singulären Effekt der fallbasierten Übung analysieren zu können, bearbeiteten die Studierenden in der *Gruppe „Fallinventar“* (Interventionsgruppe 1) zunächst nur die fallbasierte Übung. In der darauffolgenden Woche fand in dieser Gruppe die Erhebung statt, und erst nach der Erhebung, in der dritten Woche, erhielten sie die aufgabenbasierte Übung. Die inhaltlichen Übungssitzungen, die nach der Erhebung stattfanden, flossen nicht in die Untersuchung ein, sondern stellten nur sicher, dass die Teilnehmer*innen alle zur Klausurvorbereitung notwendigen Übungen vollständig angeboten bekamen. In der *„kombinierten“ Gruppe* erhielten die Studierenden zunächst die aufgabenbasierte Übung, in der folgenden Woche die fallbasierte Übung, und in der dritten Woche fand schließlich die Erhebung statt. Für diese Gruppe sollte untersucht werden, inwiefern die Teilnahme an beiden Übungen einen Mehrwert gegenüber der Teilnahme an nur einer der beiden Übungen zeigt.

Die drei Untersuchungsbedingungen wurden gleichmäßig auf vier Dozierende mit insgesamt 15 Lehrveranstaltungen verteilt. Drei der Dozierenden leiteten jeweils drei Seminare, sodass hier jede Bedingung jeweils einmal zugeordnet wurde. Eine Dozierende leitete sechs Seminare, sodass dort jede Bedingung zweimal zugeordnet werden konnte. Damit war sichergestellt, dass die Bedingungen regelmäßig über die Dozierenden verteilt waren, um mögliche Dozierendeneffekte zu kontrollieren. Außerdem wurden die Bedingungen ungefähr gleich auf die unterschiedlichen Wochentage und Tageszeiten verteilt, sodass eventuelle Verzerrungen durch die spezifische Veranstaltungszeit möglichst gering gehalten werden sollten. Die Zuordnung der Versuchspersonen zu den Bedingungen erfolgte auf Basis eines *quasiexperimentellen Untersuchungsdesigns*. Aufgrund der Lehrplanung war es nicht möglich, die Zuordnung der Studierenden zu den drei Untersuchungsbedingungen (kombiniert, Fallinventar, Aufgaben) zufällig zu gestalten, sondern die Studierenden wurden seminarweise den Bedingungen zugeordnet. Die Zuordnung der Studierenden zu den Seminaren erfolgte vor Semesterbeginn durch Selbstzuweisung im Rahmen eines Campus-Management-Systems (unsystematisch in Bezug auf die Interventionsbedingungen).

Tabelle 8.5 Evaluationsdesign zur Analyse der Interventionsgruppen

Gruppe	Reihenfolge der Sitzungen		
	Sitzung 1	Sitzung 2	Sitzung 3
„Aufgaben“ (KG)	Aufgabenbasierte Übung	Erhebung	Fallbasierte Übung
„Fallinventar“ (IG 1)	Fallbasierte Übung	Erhebung	Aufgabenbasierte Übung
„Kombiniert“ (IG 2)	Aufgabenbasierte Übung	Fallbasierte Übung	Erhebung

Anmerkungen: KG = Kontrollgruppe; IG = Interventionsgruppe

Die *Einrichtung einer Wartekontrollgruppe* war im Rahmen der regulären Seminarplanung für das Lehrmodul nicht möglich. Es gab allerdings eine kleine Gruppe von Studierenden, die vor der Erhebung an keiner Übung teilgenommen haben, weil sie in dieser Sitzung fehlten. Aufgrund der geringen Zellbesetzung kann sie nicht als repräsentative Gruppe bezeichnet werden, über die generalisierbare Aussagen getroffen

werden können. Daher werden die Daten dieser Gruppe bei den inferenzstatistischen Analysen der Untersuchungsgruppen außen vor gelassen. Die deskriptiven Ergebnisse dieser Gruppe werden in Anhang 3.F berichtet und auch im Ergebnisteil erwähnt, um die gefundenen Effekte hinsichtlich ihrer praktischen Bedeutsamkeit besser einordnen zu können.

Statistische Analysen

Die *Bereinigung und Auswertung der Daten* erfolgte mithilfe der Statistik-Software SPSS Version 18 (Statistical Package for the Social Sciences; IBM Corporation, 2015). Zur Berechnung der Interventionseffekte auf den Evaluationsebenen konnten insgesamt 280 korrekte Datensätze herangezogen werden (25 von $N = 305$ Proband*innen wurden aufgrund fehlerhafter Angaben bezüglich ihrer Teilnahme an den Übungen ausgeschlossen¹²). Von diesen nahmen $N = 257$ Studierende an denjenigen Übungen teil, denen sie entsprechend ihres Seminars zugeordnet waren, und 23 nahmen entweder an keiner Übung ($N = 11$) oder an nur einer Übung teil, obwohl sie ursprünglich der kombinierten Gruppe zugeteilt waren ($N = 12$). Diese Datensätze wurden nicht zur inferenzstatistischen Auswertung der Gruppenanalysen herangezogen, da für diese Studierenden nicht bekannt war, ob sie die Inhalte der verpassten Übung in der Zwischenzeit nachgeholt hatten oder nicht. Für die verbleibenden Versuchspersonen wurde kontrolliert, ob durch die Selbstzuordnung der Studierenden zu den einzelnen Seminaren signifikante Ungleichverteilungen hinsichtlich möglicher Störvariablen oder Kovariaten entstanden waren. Die Analyse der Variablen Altersgruppe, Studienphase, Studiengang, Erfahrung mit Diagnostik, Erfahrung in Praxis sowie Seminarleitung mittels Kreuztabellen und χ^2 -Tests über die Randsummen der jeweiligen Stufen des Faktors „Gruppe“ ergab keine signifikanten Ungleichverteilungen (vgl. Anhang 3.C). Alle Gruppenvergleiche zu den Hypothesenprüfungen erfolgten daher ohne Einbeziehung von Störvariablen.

Vergleich der Lerngelegenheiten (Fragestellung 2a)

Die inferenzstatistische Prüfung von Unterschieden zwischen den Lerngelegenheiten erfolgte auf zwei Evaluationsebenen (vgl. Tab. 8.3). Auf der ersten Ebene (Reaktion) wurden die kognitive Bewertung für die Klausur, die kognitive Bewertung für die Praxis sowie die affektive Bewertung der aufgaben- vs. fallbasierten Übung verglichen. Diese Indikatoren wurden durch Mittelung von je zwei Items gebildet. Auf der zweiten Ebene wurden das aktuelle sowie das prognostizierte Kompetenzerleben bei der Bearbeitung der Übungen jeweils über Einzelitems verglichen. Die Variable des subjektiven Lernzuwachses wurde durch Mittelung der drei verwendeten Items berechnet.

Diese sechs Variablen wurden erstens mittels *t-Tests für unabhängige Stichproben* hinsichtlich vorhandener Mittelwertsunterschiede geprüft. Hierzu wurden die Gruppen „Aufgaben“ und „Fallinventar“ verglichen. Unterschiedliche Gruppenvarianzen wurden durch die Reduzierung der Freiheitsgrade korrigiert (dies war

¹² Jene Studierende gaben an, nicht an der Übung teilgenommen zu haben, und machten dennoch Angaben zur subjektiven Bewertung der Übung. Hier konnte nicht mit Sicherheit festgestellt werden, ob sie tatsächlich nicht an der Übung teilgenommen haben bzw. worauf ihre Bewertung denn begründet ist, wenn sie nicht teilgenommen haben.

nur für die Variable „Bewertung kognitiv – Klausur“ nötig). Zweitens wurden zum Vergleich der Lerngelegenheiten *t-Tests für gepaarte Stichproben* innerhalb der kombinierten Gruppe durchgeführt, die Unterschiede der Einschätzung derselben Probanden nach beiden Übungen abbilden können. Auf der ersten Evaluationsebene erfolgte die Testung jeweils einseitig, da hier gerichtete Hypothesen vorlagen (vgl. Abschnitt 8.4.1.2).

Zur Korrektur der Alphafehler-Kumulierung wurden die p-Werte auf jeder Evaluationsebene um die Anzahl der durchgeführten Tests adjustiert (Eid et al., 2015; Sedlmeier & Renkewitz, 2008). Für alle Paarvergleiche wurden schließlich Effektstärken nach Cohen (1988) auf Grundlage der Mittelwerte und Standardabweichungen berechnet (für ungepaarte Stichproben: $d = (M1 - M2) / s_{pooled}$ mittels Effektstärken-Rechner: Hemmrich, 2015; für gepaarte Stichproben: $d_z = M / SD$ der gepaarten Differenzen). Nach Cohen (ebd.) wurden die folgenden Grenzen zur Identifikation bedeutsamer Unterschiede festgelegt: Alle Effekte mit $d \geq |.2|$ werden als klein, mit $d \geq |.5|$ als mittelgroß und mit $d \geq |.8|$ als groß bezeichnet.

Vergleich der Interventionsgruppen (Fragestellung 2b)

Zum Vergleich der drei Interventionsgruppen (Aufgaben, Fallinventar, kombiniert) wurden univariate Varianzanalysen mit dem Zwischensubjektfaktor Gruppe durchgeführt. Für die Berechnungen wurde der jeweilige Skalenmittelwert der kognitiven bzw. affektiv-motivationalen Variablen zugrunde gelegt. Die Annahme der Varianzhomogenität wurde über den Fmax-Test nach Hartley geprüft, bei dem ein Quotient aus größter und kleinster Gruppenvarianz gebildet wird. Bei ≥ 60 Freiheitsgraden liegt der kritische Wert bei 1.85, der jedoch in keinem der berechneten Vergleiche erreicht wurde (alle Werte ≤ 1.29). Eine zusätzliche Analyse der Stichprobengrößen, bei der die Fallzahl der größten Gruppe (Gruppe „Fallinventar“, $N = 109$) mit der der kleinsten Gruppe („Aufgaben“, $N = 76$) ins Verhältnis gesetzt wird, ergab zudem ein Verhältnis von 1.33 : 1, was einem vergleichbaren Gruppenverhältnis entspricht und dementsprechend robuste Ergebnisse in Varianzanalysen annehmen lässt (Tabachnick & Fidell, 2006). Für alle Gruppenvergleiche wurden Effektstärken in Form von η^2 berechnet, wobei nach Cohen (1988) die folgenden Grenzen zur Identifikation bedeutsamer Unterschiede festgelegt wurden: Alle Effekte mit $\eta^2 \geq .01$ werden als klein, mit $\eta^2 \geq .06$ als mittlerer Effekt und mit $\eta^2 \geq .14$ als großer Effekt bezeichnet.

Da vermutet wurde, dass insbesondere die kombinierte Gruppe, die beide Übungsbestandteile absolvierte, in allen Aspekten die höchste Ausprägung zeigen sollte, wurden in Anschluss an jede ANOVA, sofern diese einen signifikanten Gruppeneffekt anzeigte, Post-Hoc-Vergleiche mit Bonferroni-Korrektur berechnet.

8.4.3 Ergebnisse und Interpretation

8.4.3.1 Validierung der kognitiven Prozessdimensionen

Fragestellung 1a. Validierung der Lerngelegenheiten

Die Ergebnisse des Expert*innenratings zu den Aufgaben der jeweiligen Übungen sind in Tabelle 8.6 zusammengestellt. In der Tabelle wird deutlich, dass die *aufgabenbasierte Übung* nach Einschätzung der Expert*innen zu einem Großteil deklaratives Wissen abbildet, das sich etwa gleichermaßen aus Fakten- und Konzeptwissen zusammensetzt. Diese Übung enthält der Einschätzung der Dozierenden zufolge allerdings auch einen nicht unerheblichen Anteil von Aufgaben, die prozedurales Wissen abdecken. Sämtliche Aufgaben wurden, unabhängig von der jeweiligen Wissensart, auf den Lernzelebenen I (Erinnern) bis IV (Analysieren) eingeordnet. Bei der *fallbasierten Übung* waren sich die Expert*innen einig, dass diese ausschließlich prozedurales Wissen erfordere, welches auf den Lernzelebenen V (Evaluieren) und VI (Erschaffen) einzuordnen sei. Die Zuordnung der Lernzelebenen war für die verschiedenen Fälle und pädagogisch-diagnostischen Entscheidungen nicht immer übereinstimmend. Die Ergebnisse sprechen zusammengenommen dafür, dass die fallbasierte Übung zur Förderung prozeduralen Wissens und die aufgabenbasierte Übung zur Unterstützung deklarativen Wissens geeignet ist.

Fragestellung 1b. Validierung des Wissenstests

Die Ergebnisse des Expert*innen-Ratings zum Wissenstest sind in Tabelle 8.7 dargestellt. Hier kamen die Expert*innen zu dem Schluss, dass die Aufgaben im *Grundlagenteil* zu rund 52 % Fakten- und zu 48 % Konzeptwissen abfragten. Dieses Ergebnis war erwartungskonform. Dem Urteil der Expert*innen zufolge werde das Faktenwissen hierbei auf den Lernzelebenen I (Erinnern) und III (Anwenden) abgefragt, und das Konzeptwissen auf den Lernzelebenen II (Verstehen) bis IV (Analysieren). Der Grundlagenteil des Wissenstests prüft damit ähnliche Inhaltsbereiche ab wie die aufgabenbasierte Übung, die allerdings insbesondere beim Faktenwissen auf niedrigeren Lernzelebenen anzusiedeln sind.

Abweichend von der ursprünglichen Konzeptualisierung kamen die Expert*innen in Bezug auf den *Anwendungsteil* des Wissenstests zu der Einschätzung, dass nur bei der Abgabe der Empfehlungen prozedurales Wissen abgefragt werde. Bei den Diagnoseentscheidungen allerdings konnten die Expert*innen in den vorliegenden Fallbeispielen schwerlich zu einer Einigung kommen, ob diese tatsächlich Prozedurenwissen oder doch eher Konzeptwissen abbilden. So wurden 2/5 der Aufgaben im Anwendungsteil beim Konzeptwissen eingestuft und nunmehr 3/5 der Aufgaben entfielen auf das intendierte prozedurale Wissen. Die Diagnoseentscheidungen werden unabhängig vom Wissensbereich mehrheitlich auf der Lernzelebene IV (Analysieren) eingeordnet.

Tabelle 8.6 Ergebnisse des Expert*innenratings zu den Übungen

Aufgabe	Wissenart			Lernzielebene					
	Fakten	Konzepte	Prozeduren	I	II	III	IV	V	VI
<i>Aufgabenbasierte Übung</i>									
1a		✓					✓		
1b_1	✓				✓				
1b_2	✓			✓					
1b_3	✓			(✓)	✓				
2	✓					✓			
3	✓					✓			
4a_1		✓					✓		
4a_2	✓			✓					
4b		✓			✓				
4c	✓			✓					
4d_1			✓				✓		
4d_2		✓					✓		
4e			✓				✓		
5		✓				✓			
6_1		✓					✓		
6_2		✓					✓		
Anteil	<i>7/16</i>	<i>7/16</i>	<i>2/16</i>	<i>3/16</i>	<i>3/16</i>	<i>3/16</i>	<i>7/16</i>	<i>0/16</i>	<i>0/16</i>
(in %)	<i>(43.8)</i>	<i>(43.8)</i>	<i>(12.5)</i>	<i>(18.8)</i>	<i>(18.8)</i>	<i>(18.8)</i>	<i>(43.8)</i>	<i>(0.0)</i>	<i>(0.0)</i>
<i>Fallbasierte Übung</i>									
1_D			✓						✓
1_E			✓					✓	
2_D			✓					(✓)	✓
2_E			✓						✓
3_D			✓					(✓)	✓
3_E			✓						✓
4_D			✓					✓	
4_E			✓					✓	
5_D			✓					(✓)	✓
5_E			✓						✓
6_D			✓						✓
6_E			✓					✓	
7_D			✓					(✓)	✓
7_E			✓					✓	
8_D			✓					✓	
8_E			✓					✓	
Anteil	<i>0/16</i>	<i>0/16</i>	<i>16/16</i>	<i>0/16</i>	<i>0/16</i>	<i>0/16</i>	<i>0/16</i>	<i>7/16</i>	<i>9/16</i>
(in %)	<i>(0.0)</i>	<i>(0.0)</i>	<i>(100.0)</i>	<i>(0.0)</i>	<i>(0.0)</i>	<i>(0.0)</i>	<i>(0.0)</i>	<i>(43.8)</i>	<i>(56.3)</i>

Anmerkungen: Fett gedruckte Häkchen bedeuten, dass die Mehrheit der Expert*innen diese Einordnung favorisiert. Wenn Zweifel bei der Einordnung bestanden, wurde die mögliche Alternativlösung in Klammern vermerkt. Teilweise gab es mehrere Arbeitsaufträge bei einer Aufgabe - zur Ermittlung des jeweiligen Anteils wurde die Gesamtzahl der Arbeitsaufträge zugrunde gelegt sowie die Mehrheitsmeinung verwendet. Abkürzungen: D = Diagnose, E = Empfehlung

Tabelle 8.7 Ergebnisse des Expert*innenratings zum Wissenstest

Aufgabe	Wissenart			Lernzielebene					
	Fakten	Konzepte	Prozeduren	I	II	III	IV	V	VI
<i>Grundlagenteil</i>									
Int_3	✓					✓			
Int_5	✓			✓					
Int_6	✓			✓					
Int_18		✓			✓				
Int_7 (1-2)	2✓	(2✓)		2✓					
Int_10		✓				✓			
Int_15		✓				(✓)	✓		
Int_4	✓					✓			
Int_12		✓			✓				
Lern_6		✓			✓				
Lern_3		✓			✓				
Lern_11		✓			✓				
Lern_10		✓			✓				
Test_3	✓					✓			
Test_5	✓					✓			
Test_7	✓					✓			
Test_9		✓				✓			
Test_4		✓			✓				
Test_2	✓					✓			
Test_11	✓			✓					
Anteil	11/21	10/21	0/21	5/21	7/21	8/21	1/21	0/21	0/21
(in %)	(52.4)	(47.6)	(0.0)	(23.8)	(33.3)	(38.1)	(4.8)	(0.0)	(0.0)
<i>Anwendungsteil</i>									
Fall_1_D		✓	(✓)				✓		
Fall_1_E			✓					✓	
Fall_3_D		✓	(✓)				✓		
Fall_3_E			✓					✓	
Fall_4_D		(✓)	✓				✓		
Fall_4_E			✓					✓	
Fall_5_D		✓	(✓)				✓		
Fall_5_E			✓					✓	
Fall_6_D		✓	(X)				✓	(✓)	
Fall_6_E			✓					✓	
Fall_2_D		✓	(✓)				✓		
Fall_2_E			✓					✓	
Anteil	0/12	5/12	7/12	0/12	0/12	0/12	6/12	6/12	0/12
(in %)	(0.0)	(41.7)	(58.3)	(0.0)	(0.0)	(0.0)	(50.0)	(50.0)	(0.0)

Anmerkungen: Inhaltsbereiche im Grundlagenteil spiegeln Vorlesungsthemen wider. Abkürzungen: Int = Intelligenz, Lern = Lernstörungen, Test = standardisierte Tests. Nichteindeutige Zuordnungen wurden so abgetragen, dass Mehrheitsmeinung fett gedruckt und alternative Einordnung in Klammern gesetzt ist. Zur Ermittlung des Anteils wurden die Anzahl der einzelnen Arbeitsaufträge sowie die Mehrheitsentscheidungen zugrunde gelegt.

Eine *explorative Faktorenanalyse* (Hauptkomponentenanalyse; $N = 305$; $KMO = .78$) bestätigte die Einordnung der Expert*innen: Entsprechend des Scree-Kriteriums ergab sich eine Zweifaktorenlösung (Varianzaufklärung: 23 %), bei der erwartungsgemäß sämtliche Items aus dem Grundlagenteil auf dem ersten Faktor luden (14 von 20 Faktorladungen $> .4$, alle anderen $> .1$). Dieser erste Faktor konnte bei einem Eigenwert von 5.23 ca. 16.6 % der Varianz aufklären und wurde als „deklaratives Wissen“ bezeichnet. Auf dem zweiten Faktor, der mit einem Eigenwert von 2.07 nunmehr 6.5 % der Varianz erklärte, luden fünf der sechs Items aus dem Anwendungsteil, die die Abgabe der Empfehlungen für die Fallgeschichten abbildeten (alle Faktorladungen $\geq .35$), eindeutig. Ein Item lud gleichermaßen auf beiden Faktoren und wurde aus theoretischen Gründen der zweiten Subskala zugeordnet, die „prozedurales Wissen“ genannt wurde. Die sechs Diagnose-Items aus dem Anwendungsteil luden allerdings nicht auf diesem, sondern auf dem ersten Faktor (Faktorladungen .28 bis .40) und wurden entsprechend dem deklarativen Wissen zugeordnet. Die Faktorladungen der einzelnen Items können Anhang 3.D entnommen werden.

Vor der Faktorenanalyse wurden die *Itemschwierigkeiten für Diagnose- und Empfehlungsentscheidungen* im szenariobasierten Teil des Wissenstests geprüft, um ungeeignete Items von weiteren Analysen auszuschließen. Die Analyse zeigte, dass alle Fälle geeignet waren (vgl. Tab. 8.8). Es zeigte sich ferner, dass die Schwierigkeitsindices bei den Diagnosen eine breitere Range aufwiesen als bei den Empfehlungen. Während sich die Schwierigkeitsindices bei den Empfehlungen sämtlich im mittleren Bereich befanden, lagen bei den Diagnosen zwei Items im sehr leichten Bereich ($p > .80$). Grundsätzlich ergab sich ein ähnliches Bild hinsichtlich der Schwierigkeiten wie in Studie 2 (Kap. 7): In den meisten Fällen fielen die Lösungsraten für Diagnosen tendenziell höher aus als für Empfehlungen.

Tabelle 8.8 Lösungsraten für Diagnose- und Empfehlungsentscheidungen im Szenariotest

Falldarstellung	Diagnose	Empfehlung
Fall 1: Hochbegabter Minderleister („Underachiever“)	0.94 (0.23)	0.47 (0.24)
Fall 3: Schüler mit Lese-Rechtschreibschwäche	0.79 (0.41)	0.54 (0.23)
Fall 4: Schülerin mit Rechenschwäche	0.35 (0.48)	0.41 (0.23)
Fall 5: Schüler mit allgemeiner Lernschwäche („Lernbehinderung“)	0.85 (0.35)	0.49 (0.28)
Fall 6: Schülerin mit Hochbegabung (ohne Minderleistung)	0.34 (0.47)	0.38 (0.30)
Fall 2: Schülerin mit kombinierter Störung schulischer Fertigkeiten	0.60 (0.49)	0.43 (0.24)

Anmerkung: Angegeben sind die mittleren Lösungsraten (SD) pro Falldarstellung.

8.4.3.2 Evaluation der Lerngelegenheiten

Rückmeldungen der Dozierenden

Die Dozierenden meldeten zurück, dass die *Implementation der Übungen* in den Seminaren wie geplant erfolgte und die Mitarbeit der Studierenden zu ihrer Zufriedenheit ausfiel. Sie meldeten außerdem zurück, dass die Arbeitszeit für die Mehrheit der Studierenden nicht ausreichte, um die Übungen vollständig zu bearbeiten. Bei der aufgabenbasierten Übung entfiel in den meisten Seminaren die letzte Aufgabe, bei der

im Partnergespräch eigene Definitionen und mögliche Anschlusshandlungen für die einzelnen Lernbesonderheiten gefunden werden sollten. Diese war allerdings ohnehin als Zusatzaufgabe angedacht gewesen. Bei der fallbasierten Übung wurden in den meisten Seminaren sechs von acht Fällen bearbeitet, d. h. zwei Fälle wurden gemeinsam im Plenum besprochen und anschließend vier Fälle in Partner- und anschließender Plenumsarbeit erarbeitet. Aus der tabellarischen Übersicht der Anteile bearbeiteter Arbeitsaufträge in den Interventionsgruppen (vgl. Tabelle 8.9) geht hervor, dass in der kombinierten Gruppe sowohl bei der aufgaben- als auch bei der fallbasierten Übung tendenziell weniger Aufgaben bearbeitet wurden als in den Einzelbedingungen.

Die Einschätzungen der Dozierenden, die die Übungen im Modul „Pädagogische Diagnostik“ begleiteten, sind in Anhang 3.E für die einzelnen Dozierenden aufgeschlüsselt. Sie schätzten die aufgabenbasierte Übung mehrheitlich als sehr nützlich und relevant für die Klausur ein, die fallbasierte Übung bewerteten sie hingegen überwiegend als sehr nützlich und relevant für die spätere Berufspraxis. Auf der affektiven Dimension schätzten sie das Fallinventar sehr positiv, die aufgabenbasierte Übung als weniger positiv ein – sie erlebten diese Übung ihren Angaben zufolge als recht technisch, zahlenlastig und abstrakt. Die Arbeit mit der fallbasierten Übung wurde insgesamt als lebendiger, aber auch als herausfordernder empfunden, da hier die Lösungen weniger eindeutig gewesen seien als bei der aufgabenbasierten Übung. Die Erwartungen hinsichtlich des Lernzuwachses in beiden Übungen waren gemischt, und sie hatten überwiegend den Eindruck, dass die Studierenden sich bei der Bearbeitung beider Übungen relativ kompetent fühlten.

Tabelle 8.9 Anteil bearbeiteter Arbeitsaufträge in den Interventionsgruppen

Lerngelegenheit	Bearbeitete Aufträge	Interventionsgruppe		
		Aufgaben (N = 81)	Fallinventar (N = 109)	Kombiniert (N = 79)
Aufgabenbasierte Übung (max. 6 Aufgabenblöcke)	< 4	0		0
	4	0		13 (16.7 %)
	5	59 (72.8 %)		53 (67.9 %)
	6	22 (27.2 %)		12 (15.4 %)
Fallbasierte Übung (max. 8 Fallvignetten)	< 6		0	0
	6		17 (15.6 %)	20 (25.3 %)
	7		20 (18.3 %)	17 (21.5 %)
	8		72 (66.1 %)	8 (53.2 %)

Fragestellung 2a. Vergleich der Lerngelegenheiten

Evaluationsebene 1 (Rezeption)

Zwischengruppenvergleiche. Die Prüfung der Zwischengruppeneffekte (Gruppe „Aufgaben“ vs. Gruppe „Fallinventar“) auf der ersten Evaluationsebene ergab erstens keinen signifikanten Unterschied hinsichtlich der *kognitiven Bewertung der Klausur* ($p = .153$, $d = .16$). Die Relevanz und Nützlichkeit der

Übungen für die Klausur im Modul wurde übereinstimmend in beiden Gruppen durchschnittlich recht hoch bewertet ($M \geq 5.17$). Auch die Relevanz und Nützlichkeit der Übungen für die spätere Berufspraxis wurden in beiden Gruppen hoch bewertet ($M \geq 4.69$). Die *kognitiven Bewertungen für die Berufspraxis* unterschieden sich jedoch zwischen den Gruppen ($p = .001$, $d = -.49$): Die fallbasierte Übung wurde hier höher bewertet als die aufgabenbasierte Übung. Hinsichtlich der *affektiven Bewertung* zeigte sich das gleiche Bild: Auch hier bewerteten die Teilnehmer*innen die fallbasierte Übung höher ($p < .001$, $d = -.68$). Während die Studierenden bei der Bearbeitung des Fallinventars im Mittel hohe Lernfreude und Interesse empfanden, d. h. ihre eindeutige Zustimmung zur affektiven Dimensionen äußerten ($M = 4.80$), fielen diese Bewertungen in der aufgabenbasierten Gruppe geringer aus ($M = 4.11$). Die gefundenen Effekte sind als mittelgroß zu bezeichnen. Die statistischen Parameter sind zur Prüfung der Effekte sind in Tabelle 8.10 zusammengefasst.

Tabelle 8.10 Ergebnisse für den Vergleich der Lerngelegenheiten (Fragestellung 2a)

	Zwischengruppenvergleich						Innergruppenvergleich					
	Gruppe		Gruppenunterschied				Übung		Gruppenunterschied			
	Aufgaben	Fallinventar	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>	Aufgaben	Fallinventar	<i>t</i>	<i>df</i>	<i>p</i>	<i>d_z</i>
<i>Ebene 1 (Reaktion)¹</i>												
Bewertung kognitiv -- Klausur	5.32 (.86)	5.17 (1.00)	-1.03	175	.153	.16	5.53 (.62)	5.47 (.64)	-.92	76	.181	.11
Bewertung kognitiv -- Praxis	4.69 (1.12)	5.25 (1.03)	3.23	143	.001	-.50	4.90 (.99)	5.03 (.99)	1.76	76	.041	-.20
Bewertung affektiv	4.11 (1.0)	4.80 (1.03)	4.49	177	.000	-.68	4.51 (1.13)	4.77 (1.13)	2.83	76	.003	-.32
<i>Ebene 2 (Lernen)²</i>												
Kompetenz-erleben -- aktuell	3.88 (.936)	4.34 (.91)	3.29	177	.001	-.50	4.45 (.99)	4.57 (.99)	-1.32	76	.191	-.19
Kompetenz-erleben -- prognostisch	4.10 (1.04)	4.28 (.96)	1.21	177	.228	-.18	4.27 (.98)	4.41 (1.00)	-1.59	74	.117	-.12
Subjektiver Lernzuwachs	4.66 (1.02)	4.67 (.99)	0.86	178	.932	-.01	4.88 (.93)	4.74 (.93)	-2.34	76	.022	.27

Anmerkungen: Für die Variablen sind $M(SD)$ angegeben. Beim Zwischengruppenvergleich wurden die Mittelwerte der Gruppen „Aufgaben“ ($N = 75$) mit denen der Gruppe „Fallinventar“ ($N = 105$) verglichen. Beim Innergruppenvergleich wurden die mittleren Bewertungen der kombinierten Gruppe ($N = 77$) bezüglich der aufgabenbasierten Übung mit denen der fallbasierten Übung kontrastiert. Die Hypothesentestung erfolgte auf ¹Ebene 1 einseitig, ²Ebene 2 zweiseitig. Das Signifikanzniveau ($p \leq .05$) wurde pro Ebene durch Division durch die Anzahl der Vergleiche korrigiert ($p^* \leq .017$). Signifikante Mittelwertsunterschiede mit $p \leq .017$ sind fett gedruckt.

Evaluationsebene 2 (Lernen)

Zwischengruppenvergleiche. Im Zwischengruppenvergleich der aufgabenbasierten vs. fallbasierten Übung zeigte sich, dass sich die Gruppe „Fallinventar“ bei der Bearbeitung der Übung signifikant kompetenter fühlte als die Gruppe „Aufgaben“: Der Unterschied hinsichtlich des *aktuellen Kompetenzerlebens* entsprach einem mittleren Effekt ($p = .001, d = .50$). Die Studierenden in beiden Gruppen erlebten sich im Mittel „eher kompetent“ ($M \geq 3.88$). In Hinblick auf *prognostizierte Kompetenz* bei der zukünftigen Bearbeitung von Aufgaben dieser Art empfanden die Studierenden der verschiedenen Gruppe im Mittel keine Unterschiede ($p = .228$). Tendenziell empfand sich die Gruppe „Fallinventar“ hinsichtlich beider Dimensionen kompetenter als die Aufgaben-Gruppe. Hinsichtlich des *subjektiven Lernzuwachses* zeigte sich in beiden Gruppen eine vergleichbar hohe Einschätzung ($M \geq 4.66; p = .932$). Die Ergebnisse sind in Tabelle. 8.10 einsehbar.

Innergruppenvergleiche. Der Vergleich der Übungen innerhalb der kombinierten Gruppe ergab einen tendenziellen Vorteil der fallbasierten Übung in Hinblick auf die *erlebte Kompetenz bei der Übung* – aktuell und prognostisch (vgl. Tab. 8.10). Allerdings waren diese Tendenzen im Rahmen der *t*-Tests für gepaarte Stichproben nicht statistisch signifikant ($p \geq .117$). Hinsichtlich des *subjektiven Lernzuwachses* zeigte sich hingegen die Tendenz, dass diese nach der aufgabenbasierten Übung als höher empfunden wurde: Der als klein zu bezeichnende Effekt ist allerdings auf dem adjustierten Signifikanzniveau nur marginal bedeutsam ($d = .27, p > .017$).

Fragestellung 2b. Vergleich der Interventionsgruppen

Kognitive Voraussetzungen

Hinsichtlich der *Subskala deklarativen Wissens*, die sich aus Items zum Grundlagenwissen sowie den Diagnoseentscheidungen aus dem Szenariotest zusammensetzte, zeigte sich ein mittelgroßer Interventionseffekt, $F(2, 260) = 15.48, p < .001, \eta^2 = .107$. Die Post-Hoc-Vergleiche zeigten, dass hier diejenigen Studierenden profitieren konnten, die mit der aufgabenbasierten Übung gearbeitet hatten (vgl. auch Abb. 8.1): Während sowohl in der Aufgaben-Gruppe ($M = .69, SD = .19$) als auch in der kombinierten Gruppe ($M = .73, SD = .17$) durchschnittlich etwa 70 % der Items korrekt gelöst werden konnten, erreichten die Studierenden der Gruppe „Fallinventar“ hier nunmehr einen Lösungsanteil von knapp 60 % ($M = .58, SD = .21$). Die Leistung der Fallinventar-Gruppe war signifikant kleiner als die der beiden anderen Gruppen (beide $p < .001$). Die Aufgabengruppe und die kombinierte Gruppe unterschieden sich jedoch untereinander hinsichtlich des gemessenen deklarativen Wissens nicht signifikant ($p = .547$). Die Studierenden, die an keiner Übung teilgenommen hatten, konnten auf dieser Subskala diagnostischen Wissens nur ca. 30 % der Aufgaben korrekt lösen (vgl. Anhang 3.F).

Auf der *Subskala prozeduralen Wissens*, die das Ableiten von Empfehlungen im Szenariotest beinhaltet, ergab sich ebenfalls ein mittelgroßer Gruppeneffekt, $F(2, 253) = 7.90, p < .001, \eta^2 = .06$. Hier profitierten die Studierenden, die mit der fallbasierten Übung gearbeitet hatten: Studierende der Fallinventar-Gruppe ($M = .47, SD = .16$) wie auch der kombinierten Gruppe ($M = .49, SD = .13$) erreichten im Mittel knapp

50 % der möglichen Punktzahl. Die Post-Hoc-Vergleiche zeigten, dass die Leistungen dieser beiden Gruppen untereinander vergleichbar waren ($p = .938$), sich jedoch signifikant von der Aufgaben-Gruppe unterschieden ($p \leq .008$). Die Studierenden, die nur mit der aufgabenbasierten Übung gearbeitet hatten, konnten knapp 40 % der Empfehlungen korrekt vergeben (vgl. Abb. 8.1). Die Lösungsrate der Studierenden, die keine der Übungen absolvierten, lag auch in diesem Teil des Wissenstests bei lediglich rund 30 % (vgl. Anhang 3.F).

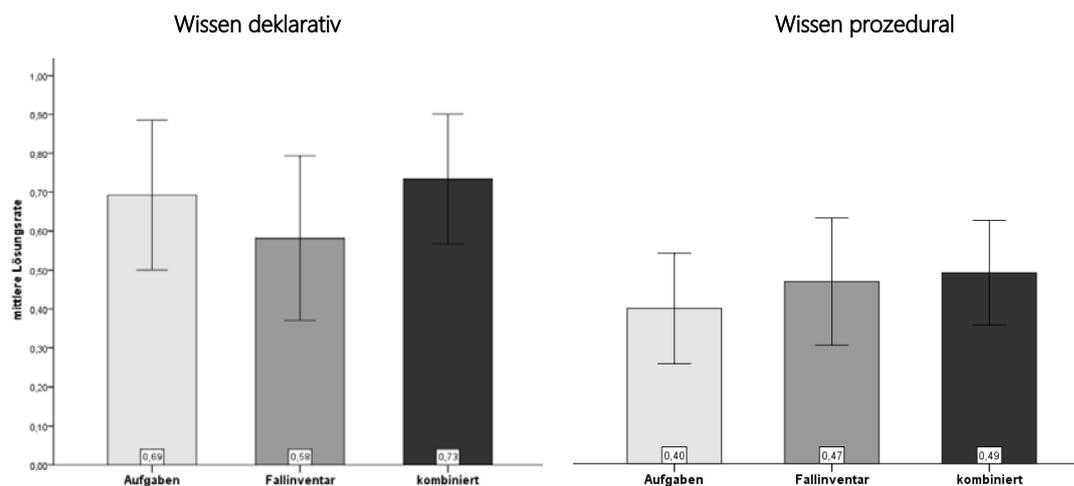


Abbildung 8.1 Gruppenunterschiede beim diagnostischen Wissen

Dargestellt sind die mittleren Lösungsraten der Studierenden \pm 1SD.
Eine Lösungsrate von 0.5 bedeutet, dass 50 % der Aufgaben gelöst wurden.

Affektiv-motivationale Voraussetzungen

Diagnostikbezogene Selbsteinschätzungen. Auf beiden Skalen diagnostikbezogener Selbsteinschätzungen zeigte sich tendenziell, dass die kombinierte Gruppe sich kompetenter einschätzte als die Studierenden in den beiden Einzelbedingungen (vgl. Abb. 8.2). Auf beiden verwendeten Skalen ergab sich ein kleiner Effekt, der allerdings nur auf der Skala selbsteingeschätzter *diagnostischer Kompetenzen* statistisch signifikant war, $F(2, 257) = 3.40, p = .035, \eta^2 = .03$. Die Post-Hoc-Gruppenvergleiche ergeben marginale Unterschiede zwischen der kombinierten Gruppe ($M = 4.12, SD = .74$) im Vergleich zur Gruppe „Fallinventar“ ($M = 3.92, SD = .59, p = .070$) sowie zur Gruppe „Aufgaben“ ($M = 3.89, SD = .59, p = .071$). Die beiden Gruppen, die an einer Übung teilgenommen hatten, unterschieden sich untereinander nicht ($p > .999$). Auf der Skala selbsteingeschätzter diagnostischer Expertise zeigte sich dieses Muster nur als Trend¹³, $F(2, 256) = 2.59, p < .10, \eta^2 = .02$ (alle Post-Hoc-Vergleiche $p \geq .132$). Die

¹³ Bei Exklusion von Item DE_8, das eine besonders hohe Streuung sowie eine geringe Trennschärfe aufweist, wären die Mittelwertunterschiede signifikant ($\text{Eta}^2 = .026, p = .034$). Zugunsten der Validität der Skala und weil eine Exklusion (siehe Tab. 3.B.2.3 in Anhang 3.B) die Reliabilität der Skala nicht substantiell erhöhen würde, wurden alle Items zur Skalenbildung einbezogen.

Skala diagnostischer Expertise bildete Fähigkeiten zur Erhebung und Nutzung diagnostischer Daten ab, während die Skala diagnostischer Kompetenz die Fähigkeiten zur korrekten Beurteilung von Leistungen erfasste. Auf beiden Skalen schätzten sich die Studierenden im Mittel als eher kompetent ein. Die Studierenden ohne Übung nahmen an, ähnlich diagnostisch kompetent zu sein wie diejenigen, die an einer der Übungen teilgenommen hatten; die Einschätzung ihrer diagnostischen Expertise fiel ähnlich hoch aus wie die der kombinierten Gruppe (vgl. Anhang 3.F).

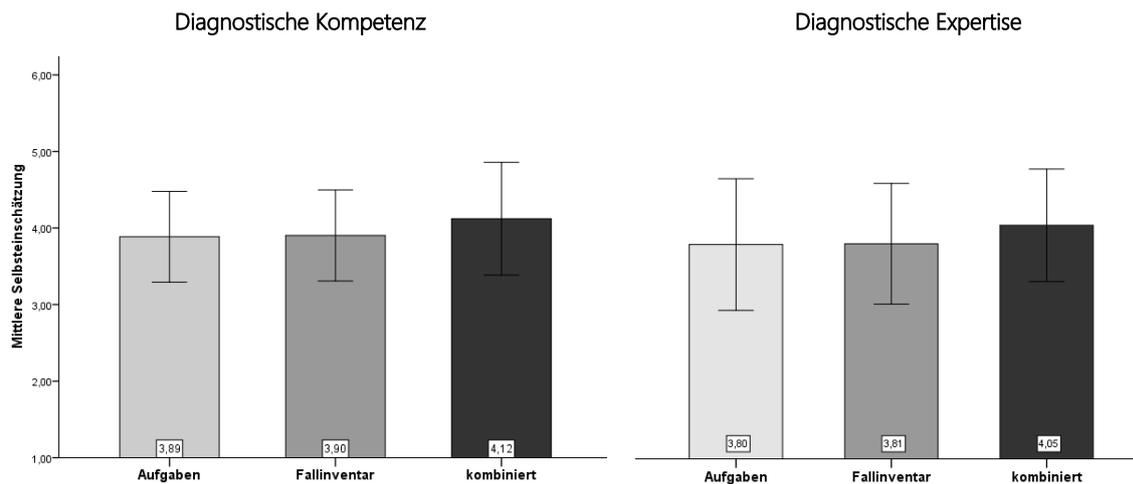


Abbildung 8.2 Gruppenunterschiede bei diagnostikbezogenen Selbsteinschätzungen

Dargestellt sind die mittleren Selbsteinschätzungen der Studierenden \pm 1SD
(1 = kann ich gar nicht, 6 = kann ich sehr gut).

Diagnostikbezogene Einstellungen. Die *Einstellung zu Diagnostik allgemein* war bei den Studierenden aller Gruppen im Mittel recht hoch ausgeprägt und unterschied sich nicht signifikant nach der Bearbeitung der unterschiedlichen Übungen, $F(2, 258) = 2.15, p = .118, \eta^2 = .02$. Auf deskriptiver Ebene zeigte sich allerdings ein kleiner Trend für eine positivere Einstellung derjenigen, die mit dem Fallinventar – einzeln oder kombiniert – gearbeitet hatten (vgl. Abb. 8.3). Die Einstellung zu Diagnostik war in diesen Gruppen als hoch zu bezeichnen ($M \geq 4.58$). Die Einstellung zu standardisierten Tests fiel in allen Gruppen etwas geringer aus, konnte im Mittel jedoch gruppenübergreifend als eher positiv eingeordnet werden $F(2, 257) = .46, p = .63, \eta^2 < .01$. Die Studierenden ohne Übung erzielten auf beiden Skalen ähnliche Einstellungswerte wie diejenigen, die eine oder beide Übungen bearbeitet hatten (vgl. Anhang 3.F).

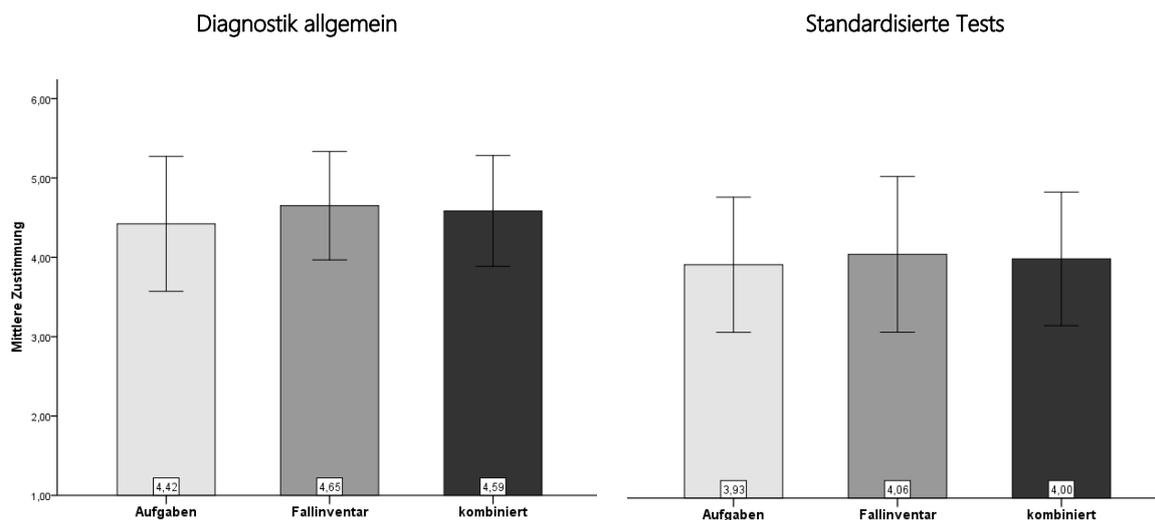


Abbildung 8.3 Gruppenunterschiede bei diagnostikbezogenen Einstellungen

Dargestellt sind die mittleren Zustimmungsraten der der Studierenden \pm 1SD
(1 = stimme gar nicht zu, 6 = stimme voll zu).

8.5 Diskussion der Ergebnisse

In der vorliegenden Studie wurde das Fallinventar als Lerngelegenheit evaluiert, die angehende Lehrkräfte auf die anspruchsvolle diagnostische Aufgabe vorbereiten soll, Schüler*innen mit Lernbesonderheiten zu diagnostizieren sowie passende Fördermaßnahmen abzuleiten. Um diese anspruchsvolle Aufgabe zu bewältigen, sind einerseits Wissen und andererseits Gelegenheiten zur Einübung des diagnostischen Vorgehens wichtig – zwei zentrale Aspekte *diagnostischer Expertise* (Hascher, 2008). Mithilfe des Fallinventars erhalten Lehramtsstudierende die Möglichkeit, ihr zuvor erworbenes Grundlagenwissen auf praxisnahe diagnostische Probleme anzuwenden. Anhand realistischer Fallvignetten lernen die Studierenden die explizite, formalisierte Vorgehensweise im diagnostischen Prozess (Jäger, 2006; Hesse & Latzko, 2017) kennen und erwerben gleichzeitig erste praktische Fähigkeiten für die Vergabe von Diagnosen sowie die Ableitung entsprechender Anschlusshandlungen. Diese erste Praxisannäherung geschieht in einem komplexitätsreduzierten Setting (Grossman et al., 2009) und soll bereits im Studium zum Aufbau prozeduralen diagnostischen Wissens beitragen.

In dieser dritten Studie der vorliegenden Arbeit war die zentrale Frage, inwiefern das Fallinventar den Aufbau prozeduralen Wissens unterstützen kann, und ob es sich eignet, eine *konventionelle aufgabenbasierte Übung* im Modul zu ersetzen. Diese aufgabenbasierte Übung wurde zur Förderung deklarativen Wissens konzipiert und seit einigen Semestern im Modul „Pädagogische Diagnostik“ eingesetzt. Die Dozierenden, die mit dieser Übung arbeiteten, hatten dabei zwar den Eindruck, dass sie die Studierenden bei der Vorbereitung der Modulabschlussprüfung unterstützt. Sie äußerten jedoch, dass die Studierenden ihrer Einschätzung nach wenig Freude bei der Bearbeitung hatten und befürchteten daher, dass die zentralen

diagnostischen Inhalte, die in dieser Übung thematisiert werden, nach der Klausur in Vergessenheit geraten und daher nicht in die Praxis übertragen werden könnten. Um den Praxisbezug zu verstärken und auch, um positive affektive Reaktionen wie Interesse der Studierenden hinsichtlich der Diagnostik von Lernbesonderheiten zu evozieren, wurde die fallbasierte Übung mit dem Fallinventar entwickelt.

Für die Evaluation dieser fallbasierten Lerngelegenheit war neben dem Aufbau diagnostischen Wissens, das hier als Zielkriterium definiert wurde, auch im Fokus, wie die Studierenden die Lerngelegenheit wahrnehmen und inwiefern sie affektiv-motivationale Aspekte für die professionelle Diagnostik von Lernbesonderheiten fördern kann. Diese Wirkungen von Interventionen für (angehende) Lehrkräfte werden in Angebot-Nutzungs-Modellen (z. B. Lipowsky, 2014) ebenfalls berücksichtigt und wurden daher in der vorliegenden Studie auf *zwei Evaluationsebenen* (Kirkpatrick & Kirkpatrick, 2006; Naugle et al., 2006) analysiert: Auf der ersten Ebene wurde untersucht, ob die fallbasierte Lerngelegenheit bei den Studierenden positive Reaktionen auslöst, die als Voraussetzung für die Bereitschaft zur Aufnahme und Verarbeitung neuer Lerninhalte gegeben sein müssen. Die Rezeption des Fallinventars wurde mit der Wahrnehmung der aufgabenbasierten Übung kontrastiert. Auf der zweiten Ebene wurde dann exploriert, ob die Lerngelegenheiten tatsächlich Lernprozesse auslösen konnten, wobei nicht nur kognitive, sondern auch affektiv-motivationale Aspekte untersucht wurden.

Evaluationsebene 1: Reaktionen auf kognitiven und affektiven Dimensionen

Beim Vergleich der aufgabenbasierten mit der fallbasierten Übung auf der *ersten Ebene (Rezeption; Lipowsky, 2014)* zeigte sich, dass grundsätzlich in beiden Lerngelegenheiten die Voraussetzungen dafür gegeben sind, dass Lernprozesse stattfinden können, denn beide Lerngelegenheiten wurden überwiegend positiv von den Studierenden wahrgenommen. Übereinstimmend mit der Annahme der Dozierenden fiel die Bewertung des Fallinventars insbesondere auf der affektiven Dimension allerdings deutlich höher aus als bei der Arbeit mit den Aufgaben: Bei der Arbeit mit den Fällen äußerten die Studierenden mehr Freude und Interesse, und zwar unabhängig davon, ob sie beide Lerngelegenheiten oder nur eine kennengelernt hatten. Diese positive Rezeption des Fallinventars war erfreulich vor dem Hintergrund, dass erste Befunde zur Arbeit mit problembasierten Fallvignetten nahelegen, dass die Lernfreude in diesen Lerngelegenheiten den aktiven, intrinsisch motivierten Wissenserwerb der Studierenden tatsächlich unterstützen kann (vgl. Wedel et al., 2020). Daneben nahmen die Studierenden auch die Praxisnähe der fallbasierten Übung positiv auf: Sie bewerteten diese Lerngelegenheit als relevant und nützlich in Bezug auf ihre spätere Berufspraxis. Die Einschätzungen zum Nutzen und zur Relevanz der Lerngelegenheit unterschieden sich allerdings weniger deutlich von der aufgabenbasierten Übung. Diese wurden auf den kognitiven Bewertungsdimensionen vor allem dann positiver bewertet, wenn im Anschluss daran das Fallinventar bearbeitet wurde (Innergruppenvergleich in der kombinierten Bedingung). Durch die Anwendung des Wissens aus der aufgabenbasierten Übung erscheint diese möglicherweise weniger praxisfern, weil hierdurch die Theorie-Praxis-Bezüge deutlicher werden.

Evaluationsebene 2: Selbsteingeschätzte Veränderungen

Eine mögliche Erklärung dafür, warum sich die Übungen auf der affektiven Dimension in der Wahrnehmung der Studierenden unterschieden, könnte darin liegen, dass die Inhalte der aufgabenbasierten Übung den Studierenden schwerer erschienen. Die Dozierenden hatten aus den früheren Semestern berichtet, dass insbesondere die statistischen und testtheoretischen Themen bei den Studierenden weniger beliebt sind und von einem Großteil als schwierig empfunden werden. Dieser Eindruck bestätigte sich beim *Vergleich der beiden Übungen auf der zweiten Evaluationsebene (Lernen; Naugle et al., 2006)*. Bei der Bearbeitung der aufgabenbasierten Übung fühlten sich die Studierenden weniger kompetent als bei der Bearbeitung der fallbasierten Übung – letztere fiel ihnen leichter. Entsprechend fiel der subjektiv empfundene Lernzuwachs in der Gruppe, die beide Übungen kennengelernt hatte, bei der aufgabenbasierten Übung retrospektiv tendenziell höher aus als beim Fallinventar. Diese Befunde überraschen vor dem Hintergrund, dass die Arbeitsaufträge der fallbasierten Übung von den befragten Expert*innen auf höheren Lernzielebenen eingeordnet wurden als die Aufgaben zum deklarativen Wissen. Aufgaben auf höheren Lernzielebenen erfordern üblicherweise komplexere kognitive Prozesse (vgl. Glessmer & Lüth, 2016; Stern et al., 2016) und müssten folglich schwerer sein. Dass die Vergabe von Diagnosen und Empfehlungen im Fallinventar als weniger schwer empfunden wurde als die Bearbeitung der aufgabenbasierten Übung, könnte einerseits darin begründet sein, dass die Studierenden ihre Fähigkeiten noch nicht ausreichend realistisch einschätzen konnten – dieser Aspekt wird weiter unten noch einmal aufgegriffen. Andererseits könnte dieses Ergebnis auch ein Zeichen dafür sein, dass die Studierenden andere kognitive Prozesse vollziehen, als die Expert*innen angenommen hatten – ein durchaus bekanntes Phänomen in der hochschuldidaktischen Praxis (Halbherr et al., 2016; Glessmer & Lüth, 2016). Ausgehend von den Ergebnissen der zweiten Studie dieser Arbeit (Kap. 7) ist denkbar, dass die Studierenden sich in ihrer Urteilsbildung nicht nur auf die formellen Informationen bezogen, deren Interpretation und Integration eine hohe kognitive Leistung erfordern (vgl. Jäger, 2006), sondern auch informelle und semiformelle Diagnoseinformationen nutzten, die leichter zu verarbeiten sind. Bei den Studierenden, die auf diese ihnen bekannten, aber unprofessionellen Verhaltensweisen zurückgriffen, könnte sich der Eindruck verstärkt haben, dass die Fälle leicht zu bearbeiten seien, sodass der subjektive Lernzuwachs hier nicht höher eingeschätzt wurde als in der aufgabenbasierten Übung. Nichtsdestotrotz haben die Studierenden in beiden Lerngelegenheiten grundsätzlich einen hohen Lernzuwachs empfunden, und kamen hinsichtlich beider Lerngelegenheiten auch zu der Einschätzung, dass sie Aufgaben dieser Art in der Zukunft recht gut bewältigen könnten.

Evaluationsebene 2: Kognitive Veränderungen

Diese Ergebnisse sind Hinweise auf tatsächliche *kognitive Veränderungen* auf der zweiten Ebene und ergänzen die Befunde zu den Wissenstests, die in dieser Untersuchung nicht im Rahmen eines Prä-Post-Designs erfasst werden konnten. Objektive Veränderungen hinsichtlich des diagnostischen Wissens wurden mit einem Interventions-Kontrollgruppen-Plan untersucht, der einen Vergleich dreier Gruppen vorsah: Studierende, die die fallbasierte Übung bearbeiteten (Interventionsgruppe 1), wurden verglichen

mit Studierenden, die die aufgabenbasierte Übung bearbeiteten (Kontrollgruppe) oder die vor dem Fallinventar zusätzlich die aufgabenbasierte Übung bearbeiteten (kombinierte Gruppe, Interventionsgruppe 2). Hierbei zeigte sich, dass durch die Arbeit mit dem Fallinventar der Aufbau prozeduralen Wissens unterstützt werden kann: Beide Gruppen, die mit dem Fallinventar gearbeitet hatten – einzeln oder in Kombination – konnten in einem Szenariotest, in denen Empfehlungen für pädagogische Anschlusshandlungen für fiktive Schüler*innen abgegeben werden, deutlich höhere Leistungen erzielen als die Aufgaben-Gruppe. Umgekehrt profitierten die Studierenden, die mit der aufgabenbasierten Übung gearbeitet hatten, beim Aufbau deklarativen diagnostischen Wissens: Sowohl die Aufgaben- als auch die kombinierte Gruppe konnten im Subtest deklarativen Wissens signifikant mehr Punkte erzielen als die Studierenden, die allein mit dem Fallinventar gearbeitet hatten. Die kombinierte Gruppe unterschied sich jeweils nicht signifikant von den jeweiligen Einzelbedingungen, obgleich sie tendenziell in beiden Wissensbereichen die höchsten Lösungsanteile aufwies. Dieses Befundmuster weist darauf hin, dass die Übungen spezifisch für die Förderung einer Wissensart geeignet sind und additiv wirken.

Diese Schlussfolgerung, dass die *Übungen spezifische Beiträge* zum Aufbau deklarativen wie prozeduralen Wissens leisten, findet teilweise Unterstützung durch die Expert*innen-Ratings, die zur Validierung der Lerngelegenheiten durchgeführt wurden. Das befragte Gremium kam zu dem Schluss, dass die Arbeitsaufträge im Fallinventar zu 100 % prozedurales Wissen fokussieren. Die Arbeitsaufträge der aufgabenbasierten Übung ordneten sie zu knapp 90 % dem deklarativen Wissen zu – ein kleiner Teil der Übung wurde jedoch auch der Förderung prozeduralen Wissens zugeschrieben. Hierbei handelte es sich um eine Übung, bei der Diagnosen sowie Förderschwerpunkte für tabellarische Fähigkeitsprofile ermittelt werden sollten. Diese Übung könnte die Studierenden zumindest teilweise auf die Items im prozeduralen Teil des Wissenstests vorbereitet haben, indem hier Fähigkeiten zur Integration diagnostischer Daten trainiert werden. Um diese Vermutung zu prüfen, wären Vergleiche der Interventionsgruppe mit einer Wartekontrollgruppe nötig. Die Implementation einer Wartekontrollgruppe war im Rahmen des Erhebungssettings zwar nicht möglich, aber erste Anhaltspunkte für die Beantwortung dieser Frage liefert eine kleine Gruppe von Studierenden, die in einer der Interventionssitzungen zufällig fehlten. Der Vergleich der Studierenden ohne Übung ermöglicht zudem eine vorsichtige Einschätzung darüber, wie groß der Wissenszuwachs durch die jeweiligen Übungen war.

Sowohl im deklarativen als auch im prozeduralen Teil des Wissenstests erreichten die Studierenden ohne Übung eine Lösungsrate von ca. 30 %. Im deklarativen Teil des Wissenstests liegt diese Leistung ca. 30 Prozentpunkte unterhalb der Leistung der Fallinventar-Gruppe sowie ca. 40 Prozentpunkte unterhalb der Leistung der Studierenden, die der Aufgaben- oder der kombinierten Gruppe angehörten. Im prozeduralen Teil des Wissenstests konnte die Aufgabengruppe im Mittel ca. zehn Prozentpunkte mehr erreichen als die Studierenden ohne Übung, und die beiden Interventionsgruppen lagen etwa 20 Prozentpunkte über der Gruppe, die an keiner der Übungen teilgenommen hatte. Diese Ergebnisse geben Hinweise darauf, dass sowohl die fallbasierte als auch die aufgabenbasierte Übung Wirkungen über die

spezifisch intendierte Wissensart hinaus haben. Diese Schlussfolgerung ist plausibel vor dem Hintergrund, dass deklaratives Wissen die Grundlage für prozedurales Wissen darstellt und dass beide Wissensarten sich im Wechselspiel miteinander entwickeln (Stern et al., 2016). Auch im Rahmen der Förderung diagnostischer Kompetenzen gibt es Hinweise darauf, dass es günstig ist, dass zunächst deklaratives Wissen erworben wird und dieses anschließend zur Anwendung kommt (z. B. Glogger-Frey & Renkl, 2017; Herppich et al., 2017a). Die fallbasierte Übung wurde daher gezielt entwickelt, um eine Anwendung deklarativen Wissens zu ermöglichen. Den vorläufigen Ergebnissen der vorliegenden Studie zufolge kann erstens geschlussfolgert werden, dass die aufgabenbasierte Übung zur Förderung deklarativen Wissens bereits teilweise den Aufbau prozeduralen Wissens begünstigt. Zweitens scheint auch die umgekehrte Schlussfolgerung zulässig: Die Anwendung deklarativen Wissens trägt nicht zur Förderung prozeduralen Wissens bei, sondern unterstützt auch die weitere Konsolidierung deklarativen Wissens.

Die Analyse der drei Übungsgruppen in Zusammenhang mit den Studierenden ohne Übung gibt außerdem zwei weitere Hinweise, die Implikationen hinsichtlich der Förderung diagnostischer Kompetenzen haben könnten. Wenn angenommen wird, dass die Ergebnisse der Studierenden ohne Übung repräsentativ sind, bedeuteten die Ergebnisse erstens, dass insbesondere hinsichtlich des deklarativen Wissens in relativ kurzer Zeit – nämlich in nur einer Seminarsitzung – ein substantieller Zuwachs erreicht werden kann. Dieser Zuwachs zeigte sich dabei nicht nur im Rahmen der aufgabenbasierten Übung, die augenscheinlich eine Steigerung um ca. 40 Prozentpunkte bewirkte, sondern auch im Rahmen der fallbasierten Übung wäre eine deutliche Steigerung um etwa 30 Prozentpunkte möglich. Für die Förderung prozeduralen Wissens ließe sich zweitens schließen, dass hier die Interventionsdosis des fallbasierten Arbeitens mit nur einer Sitzung recht knapp bemessen war und eine substantielle Förderung noch deutlich aufwändiger und zeitintensiver ausfallen müsste. Im Vergleich zur Gruppe ohne Übung konnten die Interventionsgruppen hier nur etwa 20 Prozentpunkte hinzugewinnen, und die Kontrollgruppe mit der aufgabenbasierten Übung immerhin zehn Prozentpunkte¹⁴. Trotz dieser relativ geringen Interventionsdosis ist eine erste Anbahnung prozeduralen Wissens offenbar geglückt, was das Ziel der fallbasierten Lerngelegenheit war (vgl. Kap. 6). Diese Implikationen bleiben auf Grundlage der vorliegenden Ergebnisse jedoch vorläufig und müssen im Rahmen zukünftiger Untersuchungen bestätigt werden. Jene Untersuchungen sollten bestenfalls ein Prä-Post-Design, die Implementierung einer Wartekontrollgruppe sowie eine Follow-Up-Analyse vorsehen, die die Stabilität der erzielten Effekte analysiert. Bestenfalls würden diese Untersuchungen auch praktisch tätige Lehrkräfte adressieren und

¹⁴ Es sei darauf hingewiesen, dass die Aufgaben der Übung sowie des Wissenstests, die deklaratives Wissen erfassen, auf niedrigeren Lernzielebenen angesiedelt waren als die Items zum prozeduralen Wissen. Zu dieser Bewertung kamen Expert*innen, die im Modul „Pädagogische Diagnostik“ lehrten. Möglicherweise gelang der Aufbau des deklarativen auch deshalb vergleichsweise schnell, weil die kognitiv anspruchsvollsten Prozesse hier nicht erforderlich waren. Das prozedurale Wissen wurde jedoch, den Einschätzungen der Expert*innen zufolge, auf höheren Lernzielebenen trainiert sowie getestet. Da diese komplexere kognitive Prozesse implizieren, könnte der Lernzuwachs hier geringer ausgefallen sein.

dabei analysieren, ob sich die Durchführung der Interventionen in einer Verhaltensänderung gegenüber Schüler*innen mit Lernbesonderheiten niederschlägt (Ebene 3 nach Lipowsky, 2014).

Evaluationsebene 2: affektiv-motivationale Veränderungen

Um professionelles Verhalten bei der Diagnostik von Lernbesonderheiten zu verändern, müssen neben kognitiven Voraussetzungen auch affektiv-motivationale Aspekte gegeben sein: Diese erhöhen die Wahrscheinlichkeit, dass professionelles Wissen in Verhalten überführt wird (vgl. Kompetenzmodell Blömeke et al., 2015; s. a. Kunter et al., 2011, 2013) und sollten daher bei der Evaluation von Interventionen in der Erwachsenenbildung allgemein (Kirkpatrick & Kirkpatrick, 2006) sowie in der Lehrkräftebildung im Speziellen (Lipowsky, 2014; s. a. Kittel & Rollett, 2017; Kittel et al., 2018) neben kognitiven Voraussetzungen unbedingt berücksichtigt werden. Die empirische Befundlage zu Zusammenhängen zwischen solchen Variablen sowie Verhalten ist allerdings zum gegenwärtigen Zeitpunkt schmal (vgl. z. B. Westphal et al., 2018). Auf Grundlage der bisherigen Ergebnisse kann angenommen werden, dass ein Aspekt, der zur Vorhersage professionellen diagnostischen Verhaltens herangezogen werden kann, *diagnostikbezogene Selbsteinschätzungen* sind: Diese hatten in Untersuchungen an praktisch tätigen Lehrkräften einen positiven prädiktiven Wert für die Zeit, die mit diagnostischen Tätigkeiten verbracht wird (Ohle et al., 2015) sowie die Nutzung datenbasierter Rückmeldungen zur Unterrichtsentwicklung (Westphal et al., 2018). In der vorliegenden Untersuchung zeigte tendenziell die Gruppe, die an beiden Übungen teilgenommen hatte, höhere diagnostikbezogene Selbsteinschätzungen als diejenigen, die nur eine der beiden Übungen absolvierten. Auch wenn dieser Effekt klein ist, könnte er in der Praxis von Bedeutsamkeit sein. Er deutet darauf hin, dass diejenigen, die sowohl aufgaben- als auch fallbasierte Übungsformate wahrnehmen, sich besser auf praktische diagnostische Anforderungen vorbereitet fühlen. Diese Schlussfolgerung ist allerdings als vorläufig zu bezeichnen, zum einen weil eine längerfristige Beobachtung der Entwicklung nicht stattfand, und zum anderen, weil die kleine Gruppe der Studierenden ohne Übung in etwa gleich hohe Werte zeigte wie die mit beiden Übungen.

Erfasst wurden in der vorliegenden Studie die selbsteingeschätzten Fähigkeiten zur korrekten Beurteilung von Schüler*innenleistungen (Skala diagnostischer Kompetenz; Kunter et al., 2014) sowie die selbsteingeschätzte diagnostische Expertise (nach Helmke, 2009), die prozessbezogene Fähigkeiten bei der Erhebung, Auswertung und Nutzung diagnostischer Daten abbildet. Beide Dimensionen diagnostischer Kompetenz sind zentrale Aspekte bei der Diagnostik von Lernbesonderheiten und können im Rahmen universitärer Lehrveranstaltungen gefördert werden (vgl. Thoren et al., 2020). Sie sind möglicherweise eine Voraussetzung dafür, dass die Studierenden in der Praxis nicht vor der anspruchsvollen Aufgabe der Diagnostik von Lernbesonderheiten zurückschrecken, sondern sie proaktiv angehen (zur Relevanz von Kompetenzerleben als intrinsische Handlungsmotivation allgemein vgl. Deci & Ryan, 2000). Allerdings muss diese Vermutung erst im Rahmen weitergehender Studien bestätigt werden. Tatsächliche Kompetenzen können zwar bis zu einem gewissen Grad mithilfe subjektiver Kompetenzen abgeschätzt werden (Kippers et al., 2018), allerdings müssen Selbsteinschätzungskompetenzen genauso trainiert werden wie tatsächliche Kompetenzen, und hierzu bedarf es umfangreicher eigener Lernerfahrungen (z. B.

Imme & Senkbeil, 2017). Die Ergebnisse hier müssen also mit Vorsicht interpretiert und zukünftige Studien abgewartet werden, die untersuchen, inwiefern selbsteingeschätzte Kompetenzen mit tatsächlichen diagnostischen Kompetenzen übereinstimmen (vgl. z. B. Schladitz et al., 2015).

Eine weitere affektiv-motivationale Variable, die in der vorliegenden Untersuchung analysiert wurde, waren *diagnostikbezogene Einstellungen*. Professionelle Einstellungen zur Diagnostik sind Helmke (2004) zufolge eine fundamentale Voraussetzung für ein evidenzbasiertes Vorgehen bei der Unterrichtsentwicklung: Jenes innovative und professionelle Verhalten sei nur dann zu erwarten, wenn Leistungsmessungen – zu diagnostischen oder evaluativen Zwecken – akzeptiert wären. Dementsprechend ist die Entwicklung einer professionellen Haltung ein wichtiges Ziel der Lehrkräftebildung (Helmke, ebd., 2009; s. a. Kittel et al., 2018). Bei Lehrkräften in der Praxis spielen diagnostikbezogene Einstellungen nachweislich eine Rolle bei der Vorhersage datenbasierter Unterrichtsentwicklung (Westphal et al., 2018) sowie zur Vorhersage der Zeit, die eine Lehrkraft mit diagnostischen Tätigkeiten verbringt (Ohle et al., 2015). In der vorliegenden Studie zeigten die Studierenden erfreulicherweise hohe Werte auf der Skala, die die Einstellung zu Diagnostik im Allgemeinen abbildet. Dabei ergaben sich jedoch keine signifikanten Gruppenunterschiede. Auch hinsichtlich der Einstellungen zu standardisierten Tests zeigten alle Gruppen vergleichbare Werte. Diese fielen überwiegend relativ hoch aus, erreichten jedoch geringere Zustimmungswerte als die Einstellung zu Diagnostik im Allgemeinen. Für dieses Ergebnis können zwei mögliche Gründe in Betracht gezogen werden: Auf der einen Seite konnten die Studierenden mit der fallbasierten Übung möglicherweise noch nicht gut genug von den Vorteilen eines formellen Vorgehens unter Verwendung standardisierter diagnostischer Tests überzeugt werden. Diese Annahme wäre plausibel, wenn man berücksichtigt, dass Einstellungen von Lehrkräften sich vornehmlich durch eigene Erfahrungen im Zuge von Praxiskonfrontationen ändern (für inklusionsbezogene Einstellungen vgl. Bosse & Spörer, 2014). Zwar arbeitet das Fallinventar mit problembasierten Fallvignetten, die eine intensive Auseinandersetzung anregen und auf diese Weise auch die Motivation für diagnostisches Handeln erhöhen können (Wedel & Pfetsch, 2017). Für eine Einstellungsänderung reicht diese Auseinandersetzung, die auf der ersten Stufe der Praxisannäherung anzusiedeln ist (Grossman et al., 2009), jedoch vermutlich noch nicht aus. Auf der anderen Seite könnten diese Ergebnisse auch dafür sprechen, dass Einstellungen in der universitären Ausbildungsphase zur Indikation von Interventionseffekten nicht gut geeignet sind: Wenn die Einstellungen bei Studierenden im Allgemeinen schon recht hoch ausgeprägt sind, dann kann auch durch die Teilnahme an verschiedenen Interventionen vermutlich wenig Varianz erzeugt werden. Dieser Befund deckt sich mit den Befunden aus der Studie von Julia Klug und Kollegen (2016), in der Lehrkräfte verschiedener Professionalisierungsphasen hinsichtlich ihrer diagnostischen Kompetenz untersucht wurden. Hier waren Einstellungen zwar bei Lehrkräften im Referendariat der einzige relevante Prädiktor für die Fähigkeit zur Diagnostik von Lernvoraussetzungen – bei der Gruppe der Lehramtsstudierenden hatte dieser Faktor allerdings keinen prädiktiven Wert. In dieser Bildungsphase war es ausschließlich das diagnostische Wissen, das die Kompetenzen in verschiedenen Phasen des diagnostischen Prozesses vorhersagte.

Implikationen für diagnostisches Verhalten in der Praxis

Die Vorhersage diagnostischer Kompetenz erfolgte in der Studie von Julia Klug et al. (2016) anhand eines Szenariotests, dessen Punktesumme als Kriterium diente. Dieser Szenariotest bildet verschiedene Phasen des diagnostischen Prozesses ab und ist den Autor*innen zufolge geeignet, reales diagnostisches Verhalten intern sowie ökologisch valide zu erfassen (vgl. Klug et al., 2013; aber siehe auch Praetorius et al., 2017). Für die Gruppe der Studierenden sind szenariobasierte Tests eine der wenigen Möglichkeiten, praktische bzw. prozessbezogene diagnostische Fähigkeiten zu erfassen, da diese Gruppe i. d. R. kaum Zugang zu diagnostischen Tätigkeiten in der Praxis hat. Auch in der vorliegenden Studie konnten das tatsächliche Verhalten von Lehramtsstudierenden (Evaluationsebene 3) oder gar die Auswirkungen der Interventionen auf Schüler*innen mit Lernbesonderheiten (Evaluationsebene 4 nach Lipowsky, 2014; s. a. Kirkpatrick & Kirkpatrick, 2006) nicht untersucht werden – ein bekanntes Problem bei der Beforschung von Lehramtsstudierenden (vgl. z. B. Thoren et al., 2020). Die Untersuchung dieser Zielvariablen höherer Ordnung sollte Gegenstand zukünftiger Untersuchungen sein – gerade im Kontext der Feststellung von Lernbesonderheiten: Bei diesem verbindlichen Diagnoseanlass sind die Konsequenzen des diagnostischen Verhaltens für die Betroffenen weitreichend (vgl. Karst et al., 2017; Berner, 2018; Vohrmann, 2018). Hierzu wäre die Untersuchung praktisch tätiger Lehrkräfte erforderlich, die die vorgestellten Interventionen erhalten.

Die Ergebnisse der vorliegenden Untersuchung können erste Hinweise darauf geben, wie die *pädagogisch-diagnostischen Fähigkeiten* der Studierenden zum Zeitpunkt der Messung in Bezug auf die Feststellung von Lernbesonderheiten ausgebildet sind. Der hier verwendete Szenariotest erlaubt eine Abschätzung des diagnostischen Verhaltens bei der Vergabe von Diagnosevermutungen und Empfehlungen für Fördermaßnahmen – die zwei Zielkriterien bei der Diagnostik von Lernbesonderheiten (vgl. Kap. 6). Bei der Auswertung des Szenariotests zeigte sich, dass die Fähigkeiten der Studierenden hinsichtlich beider diagnostischer Produkte die Anforderungen in der Praxis noch unzureichend erfüllen: Nach der fallbasierten Übung konnten die Studierenden etwa zwei Drittel der Diagnosen korrekt vergeben und in etwa der Hälfte der Fälle korrekte Förderempfehlungen ableiten. Verglichen mit den Lösungsraten für Diagnosen und Empfehlungen in der zweiten Studie dieser Arbeit (Kap. 7), in der das Fallinventars als szenariobasierte Umgebung eingesetzt wurde (Lösungsraten für Diagnosen: 73 %, Empfehlungen: 57 %), fällt die Leistung hier noch etwas geringer aus – zum einen vermutlich bedingt durch die Transferleistung, die im anwendungsorientierten Test erforderlich war. Zum anderen war es im Szenariotest unumgänglich, formelle Diagnoseinformationen in Form von Ergebnissen aus standardisierten Tests zu verwenden und zu integrieren. In der vorliegenden Studie zeigte sich glücklicherweise ebenfalls, dass die eingesetzten Lerngelegenheiten zu einer substantziellen Verbesserung der diagnostischen Produkte beitragen können: Diagnosen konnten von Studierenden ohne Übung in etwa 50 % der Fälle korrekt vergeben werden, nach der Bearbeitung der aufgabenbasierten Übung in ca. 60 % und durch zusätzliche Übung mit dem Fallinventar waren rund 69 % der Entscheidungen korrekt. Bei den Empfehlungen erreichten Studierende

ohne Übung nur 30 % richtige Lösungen, nach der Bearbeitung der Aufgaben rund 40 % und durch zusätzliche Arbeit mit dem Fallinventar konnten sie sich fast auf 50 % richtige Antworten steigern.

Die Analyse der Lösungsraten beider diagnostischen Produkte im Vergleich bestätigt einen weiteren zentralen Befund aus der zweiten Studie dieser Arbeit (Kap. 7), der mit den Ergebnissen des Szenariotests von Klug et al. (2016) übereinstimmt: Die *Ableitung von pädagogischen Anschlussbehandlungen* bereitet den Studierenden besondere Schwierigkeiten. In der vorliegenden Studie scheint die Formulierung von Empfehlungen sowohl vor als auch nach der Förderung deklarativen wie prozeduralen Wissens schwerer zu sein als die Vergabe von Diagnosevermutungen. Eine Erklärung für dieses Befundmuster liefern möglicherweise die Ergebnisse der Expert*innen-Befragung, die zur Validierung des diagnostischen Wissenstests durchgeführt worden war. Die Expert*innen kamen zu der Einschätzung, dass die Diagnosen im Wissenstest durch konzeptuelles Wissen zu lösen seien, wohingegen die Ableitung von Empfehlungen prozedurales Wissen erfordere. Die Begründung für diese Zuordnung, die durch eine faktorenanalytische Untersuchung bestätigt wurde, war die tabellarische Zusammenstellung der Diagnoseinformationen im Anwendungsteil des Wissenstests. Dadurch, dass dieses Raster vorgegeben sei, könnten die Diagnosen hier über einen Abgleich mit dem konzeptuellen Wissen zu Diagnosekriterien vergeben werden. Eine schrittweise Anwendung von Konzepten oder Methoden im Sinne einer Problemlösung, die für prozedurales Wissen charakteristisch wäre (vgl. Glessmer & Lüth, 2016), sei hingegen nicht erforderlich. Auf Grundlage der vorliegenden Daten kann diese Vermutung nicht abschließend geklärt werden, da hierzu eine eingehende Untersuchung der tatsächlichen Verarbeitungsprozesse der Studierenden erforderlich wäre. Die Untersuchung der kognitiven Prozessdimensionen bei der Bearbeitung des Wissenstests kann, wie auch die Verarbeitung der diagnostischen Informationen im Fallinventar, als ein Desiderat für zukünftige Forschung formuliert werden (vgl. Diskussion Kap. 7). Ein Aspekt bei der Untersuchung der Verarbeitungsprozesse könnte dabei darin bestehen, Formate mit und ohne Vorgabe tabellarischer Fähigkeitsraster zu vergleichen.

Zwischenfazit zu Studie 3: Das Fallinventar trägt zur Förderung prozeduralen Wissens bei

Die berichteten Befunde deuten darauf hin, dass die fallbasierte und die aufgabenbasierte Übung kombiniert werden sollten, um den Aufbau diagnostischer Kompetenzen im Lehramtsstudium zu fördern. Das Fallinventar kann die klassische aufgabenbasierte Übung sinnvoll ergänzen, aber nicht gänzlich ersetzen.

Die Ergebnisse der vorliegenden Evaluationsstudie zeigen, dass das diagnostische Fallinventar – wie intendiert – den Aufbau prozeduralen Wissens unterstützen kann. Damit fördert diese fallbasierte Übung die Fähigkeit der Studierenden, passende Förderempfehlungen für (fiktive) Schüler*innen mit Lernbesonderheiten abzugeben. Die Ableitung pädagogischer Anschlussbehandlungen stellt neben der Planung diagnostischer Untersuchungen eine besondere Herausforderung für angehende Lehrkräfte dar (vgl. Klug et al., 2013; 2016). Um dieses prozedurale Wissen auch sinnvoll auf unbekannte Probleme und neue Anwendungsfelder übertragen zu können, ist konzeptuelles Wissen eine Grundvoraussetzung (Stern

et al., 2016): Ein*e gute*r Diagnostiker*in muss nicht nur wissen, auf welche Weise diagnostische Fragestellungen bearbeitet werden (prozedurales Wissen), sondern muss auch verstehen, warum die Anwendung bestimmter Methoden oder die Reihenfolge bestimmter Lösungsschritte sinnvoll ist (konzeptuelles Wissen). Konzeptuelles Grundlagenwissen ließ sich im Rahmen einer klassischen aufgabenbasierten Übung effektiv fördern und dazu beitragen, die Fähigkeiten zur Abgabe korrekter Diagnosevermutungen zu verbessern. Um den Aufbau beider Wissensarten bestmöglich zu unterstützen, sollten den Ergebnissen der vorliegenden Studie zufolge beide Aufgabenformate von den Studierenden bearbeitet werden. Die Ergebnisse unterstützen jüngere Studien zur Förderung diagnostischer Kompetenz, die eine Verknüpfung deklarativen und prozeduralen Wissens proklamieren (z. B. Herppich et al., 2017a; Glogger-Frey & Renkl, 2017). Jene Ergebnisse müssen allerdings in zukünftigen Studien, die einerseits Prä-Post-Messungen und andererseits eine Wartekontrollgruppe implementieren, noch bestätigt werden.

Neben dem diagnostischen Wissen, das gute Diagnostiker*innen in der späteren Berufspraxis benötigen (vgl. Klug et al., 2016), müssen auch affektiv-motivationale Voraussetzungen geschaffen werden, um professionelles diagnostisches Handeln von Lehrkräften zu fördern (Westphal et al., 2018). In der vorliegenden Untersuchung ergaben sich Hinweise, dass durch die Bearbeitung der fallbasierten Lerngelegenheit die Motivation für diagnostisch kompetentes Handeln geschaffen werden kann. Die Studierenden erlebten bei der Bearbeitung des Fallinventars mehr Freude sowie Praxisrelevanz – mögliche Faktoren dafür, dass die Motivation für Diagnostik gegeben ist (vgl. Wedel et al., 2020). Durch die Kombination des Fallinventars mit der aufgabenbasierten Übung ließen sich zudem diagnostikbezogene Selbsteinschätzungen tendenziell erhöhen – eine Voraussetzung dafür, dass Lehrkräfte in der Praxis Zeit für diagnostische Tätigkeiten aufwenden (Westphal et al., 2018). Zur Untersuchung der Zusammenhänge zwischen Selbsteinschätzungen und tatsächlichem Verhalten sind weitere Studien nötig, nicht nur für das diagnostische Feld der Diagnostik von Lernbesonderheiten. Angesichts dieser vorläufigen Ergebnisse lässt sich jedoch erhoffen, dass die diagnostikbezogenen Selbsteinschätzungen zusammen mit den positiven Einstellungen zur Diagnostik im Allgemeinen den Aufbau diagnostischer Expertise weiter vorantreiben, deren Grundstein im Rahmen der beiden Lerngelegenheiten gelegt werden konnte.

III Zusammenfassung und Diskussion

9 Abschließende Diskussion

9.1 Zusammenfassung der empirischen Untersuchungen

Das diagnostische Fallinventar ist eine fallbasierte Lerngelegenheit, mit der angehende Lehrkräfte auf eine anspruchsvolle diagnostische Aufgabe vorbereitet werden. Die Feststellung von Lernbesonderheiten ist deshalb so anspruchsvoll, weil die Diagnosen hier oftmals mit weitreichenden Konsequenzen verbunden sind (verbindlicher Diagnoseanlass: vgl. Karst et al., 2017). Mit der Entwicklung des Fallinventars ist nicht nur eine Lerngelegenheit für diesen verbindlichen Diagnoseanlass geschaffen worden – damit steht auch ein weiteres evaluiertes Instrument zur Verfügung, das zur Förderung diagnostischer Kompetenzen eingesetzt werden kann. Die Entwicklung und Evaluation solcher Interventionen war und ist ein Desiderat der Lehrkräftebildung (Artelt & Gräsel, 2009; Schrader, 2009; 2013; Karing & Seidel, 2017; vgl. Kap. 4). Von besonderem Interesse sind solche Formate, die bereits in der universitären Phase der Lehrkräftebildung die Theorie-Praxis-Integration fördern (BMBF, 2016) – genau zu diesem Zweck soll das Fallinventar eingesetzt werden. Ziel ist es dabei, die kognitiven Voraussetzungen für kompetentes diagnostisches Handeln (Herppich et al., 2017; vgl. Kap. 3) zu fördern, d. h. anwendbares diagnostisches Wissen zu etablieren, das auf praktische diagnostische Probleme übertragen werden kann (prozedurales Wissen; vgl. Hascher, 2008). Als fallbasierte Lerngelegenheit kann es insbesondere die ersten Stufen der Theorie-Praxis-Integration fokussieren und so den weiteren Kompetenzerwerb vorbereiten (Grossman et al., 2009). Bei der Evaluation der neu entwickelten Lerngelegenheit wurden neben kognitiven Veränderungen auch affektiv-motivationale Aspekte berücksichtigt, die den Wissenserwerb unterstützen sowie den Transfer in die Praxis erleichtern können (Blömeke et al., 2015; Lipowsky, 2014). Die einzelnen Studien der Arbeit werden nachfolgend noch einmal zusammengefasst.

In der *ersten Studie* (Kap. 6) wurde die Entwicklung des Fallinventars als eine fallbasierte Lerngelegenheit vorgestellt, mit deren Hilfe Lehramtsstudierende deklaratives Wissen vertiefen, anwenden und auf diese Weise prozedurales Wissen anbahnen können. Hierfür wurden realistische, komplexitätsreduzierte Fallvignetten erstellt, in denen Schüler*innen mit Lernbesonderheiten über fähigkeitsbezogene diagnostische Informationen beschrieben werden. Die diagnostischen Informationen weisen verschiedene Formalisierungsgrade (formelle und nichtformelle diagnostische Informationen) auf und sind zur diagnostischen Entscheidungsfindung unterschiedlich relevant. Die Pilotierung zeigte, dass dieses Übungsformat für Studierende geeignet ist und die praxisvorbereitende Übung überwiegend positive Resonanz auslöste. Sowohl die tatsächlichen (kognitiver Aspekt) als auch die subjektiven Aufgabenschwierigkeiten (affektiv-motivationaler Aspekt) bei der Bearbeitung der Fallvignetten waren zufriedenstellend. Bei der Bearbeitung der prototypischen Fallvignetten entstand der Eindruck, dass die Vergabe von Diagnosevermutungen den Studierenden leichter fiel als die Ableitung pädagogischer Anschlusshandlungen, und dass dieses Muster sich in den subjektiven Aufgabenschwierigkeiten nur teilweise widerspiegelte. Dieser Eindruck wurde in der Folgestudie überprüft.

In der *zweiten Studie* (Kap. 7) wurde das Fallinventar als szenariobasierte Umgebung eingesetzt, um einerseits die Eignung der Fallvignetten für die Lerngelegenheit zu untersuchen. Zum anderen stand die Bearbeitung der Fälle hinsichtlich der Frage im Fokus, inwiefern sich Aspekte des diagnostischen Prozesses (Reihenfolge, Informationsnutzung) auf die Güte der diagnostischen Produkte (Diagnosen, Empfehlungen) auswirken. Hierzu wurden die Fallvignetten erweitert, überarbeitet sowie ihre Musterlösung validiert. Es zeigte sich erstens, dass die Lösungsraten der Fallvignetten (kognitiver Aspekt) wie auch die subjektive Sicherheit beim Treffen der pädagogisch-diagnostischen Entscheidungen (affektiv-motivationaler Bereich) überwiegend dem erwünschten mittleren Schwierigkeitsbereich zuzuordnen und daher gänzlich zum Einsatz in der Lerngelegenheit geeignet waren. Die Untersuchung des diagnostischen Prozesses bestätigte erstens, dass die Ableitung pädagogischer Anschlusshandlungen den Studierenden schwerer fällt als die Abgabe von Diagnosevermutungen. Durch experimentelle Variation der Reihenfolge dieser beiden Entscheidungen wurde zweitens offensichtlich, dass die Empfehlungsentscheidungen besonders dann schwer fallen, wenn sie – wie im normativen diagnostischen Prozess üblich – nach der Diagnose vergeben werden. Eine Schwierigkeit bei der Abgabe von Empfehlungen scheint zu sein, dass sie noch nicht von allen Studierenden zielführend mit Diagnosen verknüpft werden, eine andere, dass die Nutzung formeller diagnostischer Informationen für Empfehlungen noch nicht systematisch erfolgt. Schließlich konnten die Studierenden insbesondere bei Empfehlungen nicht realistisch einschätzen, ob ihre Entscheidung richtig war. Die Zusammenschau dieser Befunde zeigt deutlich, dass bei der Ableitung der Empfehlungen noch besonderer Förderbedarf bei den Studierenden besteht. Ob das Fallinventar die Studierenden dahingehend unterstützen kann, wurde in der letzten Studie dieser Arbeit untersucht.

In der *dritten Studie* (Kap. 8) wurden die Wirkungen des Fallinventars als Lerngelegenheit im Vergleich zu einer klassischen aufgabenbasierten Übung bzw. bei Vorgabe beider Übungen analysiert. Oberstes Ziel war die Förderung prozeduralen Wissens im Kontext der Diagnostik von Lernbesonderheiten. Die Evaluation ergab erstens, dass für beide Lerngelegenheiten die Voraussetzungen gegeben waren, um Lernprozesse anzustoßen (Ebene 1). Bei der fallbasierten Lerngelegenheit empfanden die Studierenden allerdings einen größeren subjektiven Nutzen für die Praxis (kognitiver Aspekt) sowie mehr Lernfreude (affektiv-motivationaler Aspekt). Zweitens zeigte sich, dass beide Übungen in spezifischer Weise die Entwicklung kognitiver Voraussetzungen für kompetentes diagnostischen Handeln unterstützen (Ebene 2): Während das Fallinventar maßgeblich die Entwicklung prozeduralen Wissens vorantrieb, konnte die aufgabenbasierte Übung stärker zum Aufbau deklarativen Wissens beitragen. Für die Erstellung von Diagnosen war hier offenbar deklaratives Wissen ausreichend, das im Rahmen der aufgabenbasierten Übung fokussiert wurde. Für die Ableitung von Anschlusshandlungen war hingegen prozedurales Wissen erforderlich, das im Fallinventar gefördert wurde. Es ergaben sich Hinweise darauf, dass die Förderung prozeduralen Wissens langsamer und aufwändiger ist als der Aufbau deklarativen Wissens, und dementsprechend mehr Förderung nötig ist, um Empfehlungen zu optimieren. Außerdem deutete sich an, dass sich die beiden Wissensarten im Wechselspiel miteinander entwickelten, dass also die aufgabenbasierte Übung die Entwicklung prozeduralen Wissens begünstigte und das Fallinventar die Konsolidierung deklarativen Wissens unterstützte. Die Kombination beider Übungen hatte nunmehr

einen kleinen Vorteil für die Entwicklung affektiv-motivationaler Voraussetzungen, konnte insgesamt jedoch einen wichtigen Grundstein für die Entwicklung diagnostischer Expertise legen – und zwar bereits im Lehramtsstudium.

9.2 Einordnung der Ergebnisse in den Forschungskontext

Die Ergebnisse der verschiedenen Studien zum Fallinventar geben einerseits wichtige Hinweise zur Förderung und Entwicklung diagnostischer Expertise, andererseits können sie die Befundlage zur Grundlagenforschung zu diagnostischer Kompetenz erweitern. Im Folgenden werden beide Aspekte genauer erläutert.

9.2.1 Hinweise zur Förderung diagnostischer Kompetenzen

Eine erste Schlussfolgerung, die auf Basis der vorliegenden Daten gezogen werden kann, ist dass die *Förderung prozeduralen Wissens* bereits im Lehramtsstudium möglich ist. Prozedurales Wissen ist ein essenzieller Bestandteil diagnostischer Expertise (Hascher, 2008), und da ihr Erwerb langwierig und aufwändig ist (Schrader, 2017), sollte er konsequenterweise möglichst früh einsetzen (vgl. Kap. 2.1.2). Auch wenn erste jüngere Ansätze die Förderung prozeduralen Wissens (z. B. Herppich et al., 2017a; Glogger-Frey & Renkl, 2017) oder die Verbesserung der Fähigkeiten zur Durchführung des diagnostischen Prozesses (z. B. Hetmanek & Van Gog, 2017; Klug, 2017) fokussieren, liegen aktuell kaum evaluierte Interventionen zur Förderung diagnostischer Kompetenzen von Lehramtsstudierenden vor (vgl. auch Karing & Seidel, 2017). Erfreulicherweise werden zunehmend fallbasierte Lerngelegenheiten entwickelt und erprobt (z. B. Wedel et al., 2020), die auch bereits im Studium eine erste Praxisannäherung ermöglichen und so auf diagnostisch kompetente Vorgehensweisen in der Praxis vorbereiten. Mit dem Fallinventar konnte somit eine weitere fallbasierte Lerngelegenheit entwickelt werden, die im Lehramtsstudium zur Entwicklung diagnostischer Kompetenzen beitragen kann. Die Evaluation des Fallinventars zeigte, dass es den Aufbau prozeduralen Wissens unterstützen kann und erweitert damit auch die Befundlage zur Fördermöglichkeiten prozeduralen diagnostischen Wissens in der Lehrkräftebildung – ein bislang wenig erforschtes Thema, obwohl die Relevanz prozeduralen Wissens zur Entwicklung von diagnostischer Expertise immer wieder betont wird.

Die Replikation der Ergebnisse unter Verwendung von Messwiederholungsdesigns mit Wartekontrollgruppe ist zweifelsohne ein Desiderat für die zukünftige Forschung. Die ersten hier vorliegenden Ergebnisse liefern dennoch einen wichtigen Beitrag zur Interventionsforschung im Bereich der ersten Phase der Lehrkräftebildung im Allgemeinen sowie zur Förderung der kognitiven Voraussetzungen diagnostischer Kompetenzen im Speziellen. In der dritten Studie dieser Arbeit (Kap. 8) zeigte sich, dass deklaratives Wissen im universitären Rahmen gut und schnell gefördert werden kann, und dass dieses Wissen die *Abgabe von Diagnosevermutungen* substanziell unterstützt. Dass Lehrkräfte in ihrer täglichen Unterrichtspraxis Vermutungen über Diagnosen aufstellen, sogenannte informelle Diagnosen (Schrader, 2010), ist wichtig, damit weitere professionelle diagnostische Schritte eingeleitet werden. Diese diagnostischen Hypothesen sind Ausgangspunkt eines diagnostischen Prozesses, in dem mithilfe

evidenzbasierter Verfahren auf systematische Weise eine sogenannte formelle Diagnose (Schrader, ebd.) erstellt wird. Diese manifeste Diagnose ist im schulischen System ein Ausgangspunkt dafür, dass Schüler*innen mit Lernbesonderheiten, die in ihrer seelischen Integrität und gesellschaftlichen Teilhabe bedroht oder bereits sogar so stark eingeschränkt sind, dass es einer Behinderung gleichkommt, Unterstützung z. B. in Form einer integrativen Lerntherapie (Eingliederungshilfe nach SGB VIII), erhalten. Aber nicht nur für Schüler*innen mit massiven Verzögerungen im Fähigkeitserwerb, sondern auch für jene mit einer kognitiven Hochbegabung ist es essenziell, dass ihre Problematik richtig erkannt wird und dass sie Unterstützung bekommen, um ihre Potentiale wirklich zu entfalten (vgl. z. B. Vohrmann, 2018). Insofern waren die Ergebnisse der dritten Studie (Kap. 8) erfreulich, denen zufolge vor allem die aufgabenbasierte Übung, zu einem geringeren Teil auch die fallbasierte Übung, die Fähigkeiten der Studierenden erhöhen konnten, korrekte Diagnosevermutungen abzugeben. Zukünftige Studien müssen hierbei noch zeigen, wie stabil dieser Effekt ist, und inwiefern der Transfer auf die tatsächliche schulische Praxis erfolgen kann. Die gezogenen Schlussfolgerungen dieser Studie beziehen sich auf einen Szenariotest, der zwar eine gewisse Praxisnähe erreicht, aber noch immer komplexitätsreduziert gestaltet war und außerdem ausschließlich relevante diagnostische Informationen – und das sogar in tabellarischer Form – bereitstellte. Das deklarative Wissen, das zum Abgeben zutreffender Diagnosevermutungen notwendig ist, muss dementsprechend im weiteren Verlauf der Professionalisierung weiter gefördert werden, auch deshalb, weil das Niveau der Trefferquoten im Szenariotest noch nicht zufriedenstellend war.

Noch mehr Unterstützungsbedarf besteht – auch nach der Intervention – für die *Ableitung pädagogischer Anschlusshandlungen*. Dieser Schritt im Übergangsbereich zwischen diagnostischem und pädagogischem Handeln fällt den Studierenden den Ergebnissen der zweiten Studie (Kap. 7) zufolge besonders schwer und repliziert erste Befunde zur Entwicklung prozessbezogener diagnostischer Fähigkeiten im Professionsverlauf (Klug et al., 2016). Bei der Evaluation des Fallinventars (Kap. 8) zeigte sich, dass diese fallbasierte Lerngelegenheit die Entwicklung der Fähigkeiten erhöhen kann, die die Ableitung passender Anschlusshandlungen in einem Szenariotest ermöglichen. Diese Ergebnisse unterstützen auf der einen Seite erste Befunde zur Förderung kognitiver Voraussetzungen diagnostischer Kompetenz (Herppich et al., 2017a; Glogger-Frey-Renkl, 2017; vgl. Kap. 4.3) die betonen, dass prozedurales Wissen auf der Grundlage einer soliden Basis deklarativen Wissens effektiv aufgebaut werden kann. Auch in der vorliegenden Arbeit zeigte sich, dass zur spezifischen Förderung beider Wissensarten aufgaben- und fallbasierte Übungen kombiniert werden sollten. Die Kombination beider Wissensarten ist für das diagnostische Handeln deshalb wichtig, weil eine kompetente Lehrkraft nicht nur wissen muss, wie der diagnostische Prozess funktioniert und welche Methoden eingesetzt werden können (prozedurales Wissen; vgl. Glessmer & Lüth, 2016). Das deklarative Wissen liefert einerseits die Wissensbasis hierfür (Faktenwissen), und andererseits das tiefgehende Verständnis für die Prozeduren und Methoden (konzeptuelles Wissen). Erst durch konzeptuelles Verständnis können Techniken und Vorgehensweisen sinnvoll verwendet und auch auf unbekannte Aufgaben angewendet werden (vgl. Kap. 3.2). Diagnostische Kompetenz zeigt sich im Speziellen auch darüber, ob eine Lehrkraft für den jeweiligen Anlass die richtige

Prozedur auswählt und durchführt, und die kann sich durchaus je nach Anlass unterscheiden (vgl. Brühweiler, 2017; Herppich et al., 2018; Hetmanek & Van Gog, 2017). Um sich für eine angemessene diagnostische Prozedur zu entscheiden, braucht die Lehrkraft also ein konzeptuelles Verständnis davon, warum bestimmte diagnostische Schritte in einer bestimmten Reihenfolge stattfinden und unter welchen Bedingungen man ggf. von der normativen Reihenfolge des diagnostischen Prozesses (Jäger, 2006) abweichen kann. Fallbasierte Lerngelegenheiten bieten vermutlich einen guten Rahmen dafür, deklaratives Wissen anzuwenden und somit sinnhaftes, prozedurales Wissen zu etablieren. Die Befunde der dritten Studie liefern jedenfalls erste Hinweise darauf, dass das Fallinventar auch zur Konsolidierung deklarativen Wissens beigetragen konnte. Die Verknüpfung beider Wissensarten könnte langfristig anhaltende Lerneffekte bewirken (vgl. Stern et al., 2016), die sich in der Berufspraxis niederschlagen. Zukünftige Studien müssen allerdings erst untersuchen, inwiefern die Bearbeitung beider Übungen Langzeiteffekte für beide Wissensarten nach sich zieht.

Auf der anderen Seite implizieren die Ergebnisse der Evaluation, dass die *Förderung prozeduralen diagnostischen Wissens* im Rahmen des weiteren Professionalisierungsprozesses unbedingt noch weiter ausgebaut werden muss. Zwar konnte die fallbasierte Übung zum Aufbau prozeduralen Wissens beitragen, aber das Niveau dieses Wissens war noch nicht zufriedenstellend. Da das Fallinventar auf der ersten Stufe der schrittweisen Annäherung an die Praxis (Grossman et al., 2009) angesiedelt ist, kann diese fallbasierte Übung nicht mehr als ein Anfang im Erwerbsprozess diagnostischer Expertise sein. Auch wenn sich durch die eingesetzten Übungen in diesem kurzen Zeitrahmen erste Verbesserungen prozeduralen Wissens erwirken ließen, erfordert die umfassende Förderung prozeduralen Wissens – als ein elementarer Bestandteil diagnostischer Expertise (Hascher, 2008) – umfangreicher praktischer Erfahrungen (Schrader, 2017). Allerdings bietet die erste Phase der Lehrkräftebildung vermutlich wenig Spielraum für weitere praxisvorbereitende oder praktische Lerngelegenheiten. Im Rahmen der Lehrkräftebildung an der Freien Universität Berlin beispielsweise können die Studierenden ihre prozessbezogenen Fähigkeiten zur Diagnostik von Lernbesonderheiten nach Absolvieren des Moduls „Pädagogische Diagnostik“ nur dann erweitern, wenn sie an einem diagnostischen Lernforschungsprojekt im Praxissemester teilnehmen (Thoren et al., 2020; Wißmann & Pöhler, 2020) – eine Option, die nur etwa einem Bruchteil der Studierenden zur Verfügung steht. Die Verbesserung der prozessbezogenen Diagnosefähigkeiten muss entsprechend auch in späteren Phasen der Lehrkräftebildung fokussiert werden.

Eine weitere Möglichkeit, um den Aufbau prozeduralen Wissen noch innerhalb des Lehramtsstudiums stärker zu fördern, liegt in einer *Überarbeitung des Fallinventars* selbst: Denkbar wäre, dass höhere Zuwächse hinsichtlich des prozeduralen Wissens erreicht werden könnten, wenn die Vergabe von Diagnosen und die Ableitung von Empfehlungen klarer strukturiert und instruiert würden. Wie die Befunde der zweiten Studie (Kap. 7) offenlegten, haben die Studierenden ihre Empfehlungen für Anschlusshandlungen noch nicht ausreichend mit den Diagnosevermutungen verknüpft und die formellen Diagnoseinformationen noch nicht systematisch genug genutzt. Zwar wurde bei der Durchführung der Lerngelegenheit im Rahmen der Evaluation darauf hingewiesen, dass die Informationen zu den schulischen und kognitiven

Fähigkeiten, die in den Vignetten dargestellt waren, sich hinsichtlich ihrer Objektivität und Reliabilität unterscheiden. Wie die verschiedenen, teilweise widersprüchlichen Informationen jedoch ausgewählt und integriert werden können, wurde nicht explizit thematisiert. Zur Unterstützung der Informationssammlung und -nutzung könnte möglicherweise ein bewusster Zwischenschritt eingeführt werden, bei dem die formellen Diagnoseinformationen in ein Fähigkeitsraster überführt werden, ähnlich wie es bei der Konzeptualisierung des Fallinventars verwendet wurde (vgl. Tab. 6.4 prototypische Fähigkeitsraster). Auf Grundlage dieses Fähigkeitsrasters würden dann sowohl die Diagnosen als auch die Förderempfehlungen abgeleitet werden, und auf diese Weise das Prinzip der symptomorientierten Förderung (vgl. Heine et al., 2012) trainiert werden. Die Anfertigung eines Fähigkeitsrasters wäre entsprechend eine Möglichkeit des Scaffoldings während der Bearbeitung des Fallinventars, um die Informationsnutzung und die Verschränkung von Diagnosen und Empfehlungen zu optimieren. Wenn Studierende diesen Zwischenschritt bei neuen diagnostischen Fällen selbstständig erinnerten und einsetzten, könnten sie über die verbesserte Datennutzung die Güte ihrer Diagnosen sowie Empfehlungen erhöhen (Behrmann & Van Ophuysen, 2017). Eine andere Möglichkeit, um dieses Ziel zu erreichen, wäre die gezielte Förderung der Reflektion des eigenen diagnostischen Vorgehens und eigener Urteilsverzerrungen (vgl. Hesse & Latzko, 2017; Kap. 4.2). Damit Studierende die Fallarbeit gezielt reflektieren, erscheint der Einsatz der Methode der strukturierten Reflexion vielversprechend. Der Ansatz der strukturierten Reflexion, der ursprünglich aus der medizinischen Diagnostikausbildung stammt (vgl. Hetmanek & Van Gog, 2017; Kap. 4), ermöglicht eine Bewusstmachung des eigenen diagnostischen Vorgehens. Er stellt eine Struktur für die Sammlung und Integration diagnostischer Informationen bereit, die auf jeden beliebigen diagnostischen Fall – auch außerhalb szenariobasierter Umgebungen – übertragbar ist. Dieses Vorgehen bereitet insbesondere die Vergabe einer qualitativ hochwertigen Diagnose vor, und müsste vermutlich noch durch Maßnahmen ergänzt werden, die die „Übersetzung“ der Diagnosen in entsprechende Anschlusshandlungen gezielt unterstützen. Um (angehende) Lehrkräfte in die Lage zu versetzen, zielführende Anschlusshandlungen abzuleiten, reicht die Vermittlung von Interventionsprinzipien nicht aus, sondern es sind umfassende Schulungsmaßnahmen erforderlich (vgl. Förster & Souvignier, 2014, 2015, 2017; Kap. 4.2.2).

Insbesondere die Ergebnisse der zweiten Studie (Kap. 7) legen nahe, dass die *Übersetzung von Diagnosen in Anschlusshandlungen* nicht automatisch passiert, sondern dass es hierzu gezielter Interventionen bedarf (Studie 2, Kap. 7). Bei der Bearbeitung des Fallinventars fielen zwar Empfehlungen bedingungsübergreifend schwerer als Diagnosen, aber besondere Schwierigkeiten hatten die Studierenden, wenn sie diesen Empfehlungen die Vergabe einer Diagnose vorausging. Diese normative Reihenfolge sollte zu den besten Ergebnissen führen, was bei den Studierenden offenbar nicht der Fall war. Dass es auch praktisch tätigen Lehrkräften generell schwerfällt, diagnostische Daten in Anschlusshandlungen zu überführen, zeigen beispielsweise Studien zur Nutzung von Ergebnissen aus Vergleichsarbeiten (vgl. z. B. Hosenfeld & Groß-Ophoff, 2007). Modelle zur evidenzbasierten Unterrichtsentwicklung sehen entsprechend nach der Rezeption der Daten eine Phase der Reflektion vor, die der Ausführung von Handlungen vorausgeht (Helmke, 2004). In dieser Reflektionsphase sollen mögliche Ursachen gefunden

werden, die das Zustandekommen der Daten erklären könnten. Sogenannte Erklärungshypothesen sind im Modell des pädagogisch-diagnostischen Prozesses nach Ingrid Hesse und Birgit Latzko (2017) neben Entstehungshypothesen ebenfalls verankert (vgl. Kap. 2). Wie diese Reflektion allerdings gefördert werden kann, wird in diesen Modellen nicht formuliert. Möglicherweise ist eine Bewusstmachung eigener evtl. vorschneller Schlussfolgerungen bereits ausreichend (De-Automatisierung von Typ1-Prozessen, vgl. Hetmanek & Van Gog, 2017). Andernfalls ist das Formulieren von alternativen Entstehungshypothesen und darauf basierend die Formulierung alternativer Anschlusshandlungen (vgl. FIWE-Methode nach Beywl & Schepp-Winter, 2000) womöglich ein zielführender Ansatzpunkt. Diese systematische Reflektion kann jedoch nicht im Rahmen der Sitzung mit dem Fallinventar eingeübt werden, hierfür bedarf es weiterer Interventionen. Aktuell wird im Modul „Pädagogische Diagnostik“ eine Lerngelegenheit zur Nutzung von VERA-Ergebnissen erprobt, die die Kompetenzen zur Nutzung diagnostischer Daten aus Vergleichsarbeiten fördern soll. Die Ergebnisse der begleitenden Studien, die im Zuge der zweiten Förderphase der Qualitätsoffensive Lehrerbildung durchgeführt werden, bleiben abzuwarten.

9.2.2 Implikationen für die Modellierung diagnostischer Kompetenz

Obwohl die Ableitung von pädagogischen Anschlusshandlungen ein zentraler Aspekt des diagnostischen Prozesses im pädagogischen Kontext sein sollte (vgl. Hesse & Latzko, 2017; Kap. 2) und dieser Schritt angehenden wie erfahrenen Lehrkräften schwer fällt, spielt diese sogenannte „postaktionale Phase“ (Klug et al., 2013, 2016) in der Modellierung diagnostischer Kompetenzen von Lehrkräften bislang allenfalls eine untergeordnete Rolle. So wurden in zahlreichen Studien Faktoren gesucht und gefunden, die die Genauigkeit von diagnostischen Urteilen beeinflussen (z. B. Südkamp et al., 2012). Ob diese Faktoren aber für die Korrektheit von Anschlusshandlungen genauso ausschlaggebend sind wie die von Diagnosen, ist m. W. bislang nicht systematisch untersucht worden. Aktuelle Modelle diagnostischer Kompetenz (z. B. Herppich et al., 2018) fokussieren die Diagnose als Endpunkt des diagnostischen Prozesses, und die Anschlusshandlungen stehen außerhalb des Augenmerks. Auch wenn diskutiert werden kann, dass es sich um klar abgrenzbare Kompetenzbereiche handelt – nämlich pädagogische vs. diagnostische Kompetenz (vgl. Kap. 1.2.2 und 2.2.3) – gehen beide Handlungsaspekte bei Lehrkräften stets Hand in Hand (Hesse & Latzko, ebd.; Jürgens & Lissmann, 2015) und müssen daher gemeinsam gedacht werden (Verweis Theoriekap. 2). Genau diese Schnittstelle stellt die Pädagog*innen vor große Herausforderungen und bedarf daher besonderer Aufmerksamkeit. Im Zentrum steht dabei die gegenwärtig offene Frage, warum die Ableitung von Anschlusshandlungen derart schwer fällt und welche Faktoren demnach die Güte der Empfehlungen beeinflussen.

Eine mögliche Antwort auf diese Frage können die Ergebnisse der zweiten Studie geben. Dort war eine Erklärung dafür, dass die Trefferquoten bei Empfehlungen geringer waren als bei Diagnosevermutungen, dass die Verarbeitung der diagnostischen Informationen nicht systematisch erfolgte: Je nach Entscheidung und Reihenfolge zogen die Studierenden unterschiedliche Informationen heran, obwohl es hierfür keinen äußeren Grund gab. Die Beweggründe für diese Unterschiede in der Verarbeitung diagnostischer Informationen können mit den vorliegenden Daten nicht ermittelt werden. Sie sind jedoch ein deutlicher

Beleg dafür, dass nicht nur die Güte von Diagnosen, sondern auch die der Empfehlungen durch die Informationsauswahl beeinflusst werden – eine Erweiterung des Vier-Komponenten-Modells der Diagnosequalität (Behrmann & Van Ophuysen, 2017). Ferner weisen sie darauf hin, dass den Studierenden vermutlich nicht klar war, welche diagnostischen Informationen in welcher Weise zu nutzen seien – ein Aspekt prozeduralen Wissens (vgl. Hascher, 2008) und ein möglicher Hinweis darauf, dass sich die *kognitiven Prozessdimensionen für Diagnosen und Empfehlungen* unterscheiden. Die Ergebnisse der dritten Studie dieser Arbeit lassen diese zweite mögliche Antwort auf die Frage zu: Der Grund dafür, warum die Ableitung von Anschlusshandlungen schwerer fällt, könnte in der Komplexität der kognitiven Prozesse liegen, die für die jeweilige Entscheidung erforderlich sind. Zur Lösung der Fälle im Szenario basierten Wissenstest, der zur Evaluation kognitiver Veränderungen eingesetzt wurde, war zur Abgabe von Diagnosen offenbar konzeptuelles Wissen ausreichend, während für Empfehlungen prozedurales Wissen erforderlich schien. Dieses zunächst überraschende Ergebnis zeigte sich bei der Validierung des selbstentwickelten Tests, bei dem ursprünglich davon ausgegangen war, dass auch die Vergabe von Diagnosen prozedurales Wissen voraussetze. Bei der Einschätzung der kognitiven Prozessdimensionen beim Fallinventar wiederum hatten die Expert*innen sowohl Diagnose- als auch Empfehlungsentscheidungen dem prozeduralen Wissen zugeschrieben. Empfehlungen wurden dabei auf einer höheren Lernzielebene eingeordnet als Diagnosen, was dafür spricht, dass die Expert*innen auch hier komplexere kognitive Prozesse für Empfehlungen annahmen. Dass die Diagnosen dabei dem prozeduralen Wissen zugeordnet wurden, lag maßgeblich daran, dass dort verschiedene, teilweise widersprüchliche diagnostische Informationen vorlagen. Bei den Falldarstellungen im Szenariotest hingegen waren ausschließlich relevante Informationen gelistet, die außerdem in tabellarischer Form zusammengestellt waren, sodass die Expert*innen von einer Erleichterung der Informationsverarbeitung ausgingen.

Auch die Unschärfe des Begriffs des prozeduralen Wissens trägt sicher dazu bei, dass eine sichere Zuordnung der Diagnoseentscheidungen zu den Wissensarten nicht leicht fiel (vgl. Kap. 3.2.2.). Prozedurales Wissen wird gemeinhin als handlungsbezogenes Wissen definiert, welches im praktischen Kontext häufig automatisiert vorliegt (vgl. Hetmanek & Van Gog, 2017; Stern et al., 2016). Im hochschuldidaktischen Rahmen wird es jedoch eher als Wissen über fachspezifische Methoden und Prozeduren verstanden (Glessmer & Lüth, 2016), das durchaus bewusst sein kann und weniger Wiederholungen benötigt als automatisiertes prozedurales Wissen. Außerdem muss bedacht werden, dass sich der diagnostische Prozess in szenariobasierten Umgebungen durchaus vom diagnostischen Vorgehen in der Praxis unterscheidet: Es werden hier keine Erhebungen geplant, durchgeführt und ausgewertet, deren Verarbeitung die Qualität der diagnostischen Daten bedingt, die wiederum in die Diagnose eingeht (vgl. Vier-Komponentenmodell der Diagnosequalität; Behrmann & Van Ophuysen, 2017). Vielmehr ist es hier so, dass Daten unterschiedlicher Güte bereits vorliegen, und diese müssen ausgewählt und verarbeitet werden und so ergibt sich die Qualität der Diagnose. So kommt es möglicherweise „künstlich“ zustande, dass die Diagnosevermutungen im Szenariotest kein prozedurales Wissen erfordern. Angesichts dieser Unterschiede zwischen dem Vorgehen in szenariobasierten Umgebungen und der realen Praxis ist nicht

auszuschließen, dass auch beim Fallinventar für *Diagnosen kein prozedurales Wissen* erforderlich ist. Die Expert*innen-Kommission hatte hierzu rege diskutiert, und sich schließlich darauf geeinigt, dass wohl prozedurales Wissen erforderlich sei – in der Annahme, dass die Studierenden die verschiedenen Diagnoseinformationen selektieren und integrieren. Dass das nicht bei allen Studierenden der Fall war, lässt sich anhand der Ergebnisse der zweiten Studie annehmen. Auf der anderen Seite ist bei den aktuellen Daten nicht auszuschließen, dass die Diagnosen ökologisch valide die erforderliche Wissensart abbilden, und dass Diagnosen entsprechend konzeptuelles Wissen erfordern, während nur Empfehlungen prozedurales Wissen erfordern. Sämtliche Deutungen der kognitiven Prozessdimensionen müssen zum aktuellen Zeitpunkt jedoch Spekulation bleiben und durch eine differenziertere Untersuchung der Verarbeitungsprozesse bei den Studierenden selbst erst bestätigt werden – und zwar für das Fallinventar sowie den Wissenstest. Hierzu kämen insbesondere die Methode des lauten Denkens, die genauere Auskunft über das Vorgehen und die Beweggründe der Studierenden geben könnte, oder Eyetracking gestützte Untersuchungen in Betracht, bei denen u. a. die Verweildauer auf einzelnen diagnostischen Informationen sowie der Grad kognitiver Beanspruchung („cognitive load“) objektiv gemessen werden können.

Ein weiterer Aspekt, der bei der Modellierung wie Förderung diagnostischer Kompetenzen – bislang – keine zentrale Rolle spielt, sind *affektiv-motivationale Aspekte*. Diese finden sich zwar in allgemeinen Definitionen von Kompetenz (Blömeke et al., 2015; Klieme & Leutner, 2006; Weinert, 2001), werden jedoch in aktuellen Arbeitsmodellen zur Erforschung diagnostischer Kompetenz (z. B. Herppich et al., 2018) oder auch in Strukturierungen diagnostischer Förderansätze (vgl. Karing & Seidel, 2017) allenfalls am Rande erwähnt. Dabei könnten sie einen ganz entscheidenden Einfluss darauf haben, ob die erworbenen Kompetenzen (z. B. Wissen im Sinne einer Disposition) auch in Verhalten umgesetzt werden (manifeste Kompetenz). Im Rahmen von Fortbildungen für praktisch tätige Lehrkräfte sollten sie daher immer berücksichtigt werden (vgl. Lipowsky, 2014). In der vorliegenden Arbeit wurden affektiv-motivationale Aspekte neben den kognitiven Dimensionen, die auch hier zweifelsohne im Zentrum standen, ergänzend berücksichtigt. Hierzu wurden in der Evaluationsstudie (Kap. 8) diagnostikbezogene Selbsteinschätzungen sowie Einstellungen herangezogen. In der vorliegenden Arbeit zeigten sich dabei hinsichtlich beider Variablen kaum Effekte, die mit den spezifischen Lerngelegenheiten im Zusammenhang standen. Ein Grund hierfür könnte sein, dass die Intensität der Auseinandersetzung in den Lerngelegenheiten nicht ausreichend war, um Änderungen in den erfassten Konzepten zu bewirken. In Bezug auf *diagnostikbezogene Einstellungen* wäre denkbar, dass diese im Rahmen universitärer Lehrveranstaltungen möglicherweise schwerer zu verändern sind (vgl. Thoren et al., 2020) als über praktische Erfahrungen und Konfrontationen (z. B. Bosse & Spörer, 2014). Zwar sind Einstellungsänderungen bspw. im sonderpädagogischen Kontext über die Sensibilisierung für Inklusion im Rahmen von Lehrveranstaltungen möglich (vgl. auch Opalinski & Scharenberg, 2018), hier wären aber vermutlich noch stärkere kognitive (z. B. Reflexion, vgl. Richardson, 1996) oder affektive (z. B. persönliche Begegnungen) Reize als auslösende Momente notwendig gewesen.

Ein anderer Grund liegt möglicherweise darin, dass diese Konstrukte zwar für Lehrkräfte in der Praxis einen prädiktiven Wert aufweisen (vgl. Klug et al., 2016; Ohle et al., 2015; Westphal et al., 2018), für Lehramtsstudierende aber weniger geeignet sind, um kompetentes Verhalten – in diesem Fall Testverhalten – vorherzusagen. Hinsichtlich der Förderung *diagnostikbezogenen Selbsteinschätzungen* muss außerdem bedacht werden, dass eine Erhöhung derselben nicht uneingeschränkt angestrebt werden sollte: Einerseits sollten Lehrkräfte über ein Mindestmaß an Selbstwirksamkeit verfügen, damit sie sich diagnostischer Aufgaben überhaupt annehmen (z. B. Ohle et al., 2015). Auf der anderen Seite sollten selbsteingeschätzte Kompetenzen realistisch sein, was bei Lehramtsstudierenden nicht unbedingt der Fall ist (z. B. Müller, 2010). Konkret für die Erstellung korrekter Diagnosen und Empfehlungen könnten realistische Selbsteinschätzungen besonders wichtig sein, um bspw. bei eigener Inkompetenz an kompetente Dritte zu verweisen (vgl. Jäger, 2006). Realistische Selbsteinschätzungen setzen einerseits vermutlich umfangreiche Praxiserfahrungen voraus (Ihme & Senkbeil, 2017; Hellmich & Görel, 2014), die im Rahmen des Studiums nicht leicht zu erwerben sind. Außerdem sind sie möglicherweise auch mit Persönlichkeitseigenschaften konfundiert, die im Rahmen universitärer Ausbildungsformate schwer veränderbar sind. Differenzierte Analysen zu den Zusammenhängen allgemeiner sowie spezifischer Voraussetzungen der Lehrkräfte sind daher ebenso ein Desiderat für zukünftige Forschung wie die Untersuchung von Zusammenhängen zwischen kognitiven sowie affektiv-motivationalen Voraussetzungen (zum Zusammenhang von Wissen, Einstellungen und Selbstwirksamkeit bei Lehrkräften vgl. Ernest, 1989; Richards, 1996). Zur Untersuchung dieser Zusammenhänge speziell bei diagnostischen Kompetenzen und speziell bei Lehramtsstudierenden müssen außerdem zunächst geeignete affektiv-motivationale Variablen identifiziert werden, die in Zusammenhang mit Fähigkeiten zur Vergabe korrekter Diagnosen und Empfehlungen stehen sowie ausreichend änderungssensitiv sind.

9.3 Ausblick und Vorschlag eines erweiterten Rahmenmodells für die Erforschung diagnostischer Kompetenzen

Zur Förderung diagnostischer Kompetenzen stehen gegenwärtig verschiedene Ansätze unverbunden nebeneinander, deren Wirksamkeitsnachweise teilweise ausstehen. Um diese Interventionen für Lehrkräfte verschiedener Professionalisierungsstufen zu systematisieren, wäre ein Modell hilfreich, das auch für die Entwicklung und Evaluation zukünftiger Interventionen Orientierung bieten könnte.

Zukünftige Studien zur Erforschung diagnostischer Kompetenzen sollten nicht nur korrekte Diagnosen als Kriterium fokussieren, sondern außerdem sowohl die pädagogischen Anschlusshandlungen als auch die Qualität der diagnostischen Daten als „diagnostische Produkte“ berücksichtigen. Eine offene Frage ist derzeit, inwiefern die Qualität dieser Produkte durch situationsspezifische Fähigkeiten im diagnostischen Prozess beeinflusst wird. Eine andere, inwiefern diese situationsspezifischen Fähigkeiten durch einerseits kognitive (deklaratives und prozedurales Wissen) und andererseits affektiv-motivationale Voraussetzungen moduliert werden. Um diese Wirkzusammenhänge weiter zu untersuchen, wird folgendes Rahmenmodell zur Strukturierung dieser Forschungsbemühungen vorgeschlagen (vgl. Abb. 9.1).

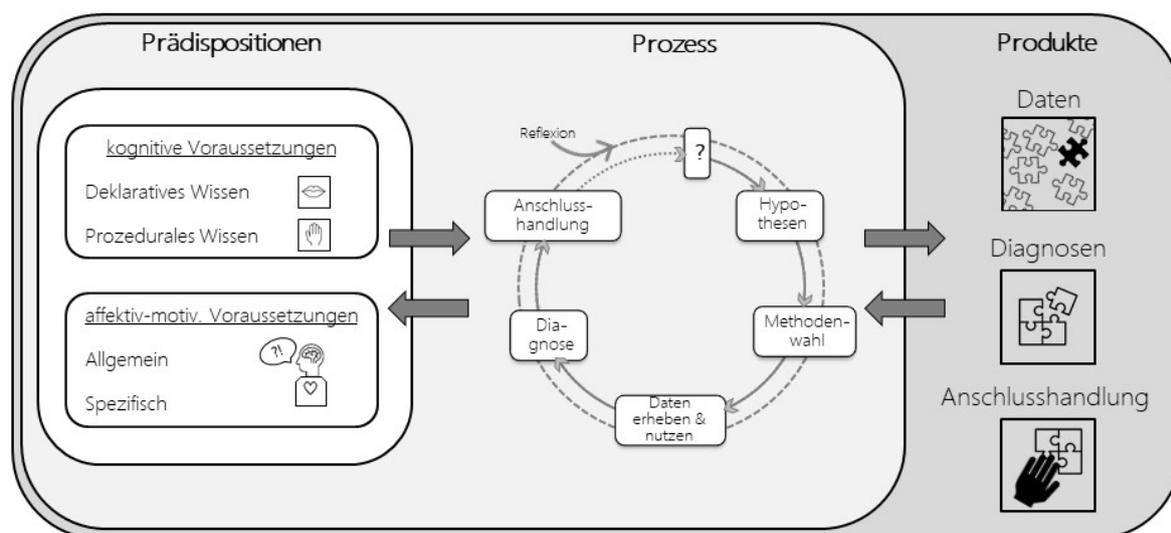


Abbildung 9.1 Vorschlag eines erweiterten Rahmenmodells diagnostischer Kompetenz

Das vorgeschlagene Modell integriert aktuelle Kompetenzmodelle im Kontext der Diagnostik.

Grundlegend ist die Vorstellung von Kompetenz, dass sich diese anhand der Voraussetzungen sowie der Handlungen messen lässt (Blömeke et al., 2015). Kompetentes diagnostisches Verhalten wird in Form von Diagnosen oder diagnostischen Daten sichtbar (Behrmann & Van Ophuysen, 2017). Im Kontext der pädagogischen Diagnostik sollte daneben auch die Ableitung von Anschluss-handlungen als Aspekt kompetenten Verhaltens bedacht werden (Klug et al., 2013, 2016). Zu den Voraussetzungen zählen sowohl kognitive als auch affektiv-motivationale Aspekte (Blömeke et al., 2015). Diese spezifischen wie unspezifischen Voraussetzungen haben Einfluss auf den diagnostischen Prozess, deren Ablauf sich je nach diagnostischer Situation unterschiedlich gestalten kann und auf diese Weise die Qualität der diagnostischen Produkte maßgeblich beeinflusst (Herppich et al., 2018). Ein Faktor mit besonderer Bedeutung für die Qualität der diagnostischen Produkte im Prozess ist dabei vermutlich die Reflexion über die Auswahl und Durchführung der diagnostischen Schritte (Hesse & Latzko, 2017; Hetmanek & Van Gog, 2017; Förster & Souvignier, 2017).

Angesichts der verschiedenen Ansätze, die aktuell im Rahmen verschiedener Projekte der Lehrkräftebildung entwickelt und evaluiert werden, könnte ein solches Modell die Forschungsbemühungen verschiedener Arbeitsgruppen auf einen gemeinsamen Nenner bringen.

9.4 Fazit

Fallbasierte Lerngelegenheiten sind vielversprechende Ansätze zur Förderung diagnostischer Kompetenzen in der universitären Ausbildungsphase von Lehrkräften. Durch den Einsatz des Fallinventars können Lehramtsstudierende zumindest in Anfängen auf die anspruchsvolle Aufgabe der Diagnostik von Lernbesonderheiten vorbereitet werden. Dabei lernen sie das Vorgehen eines regelgeleiteten diagnostischen Prozesses nicht nur kennen, sondern wenden dieses selbst auf virtuelle Schüler*innen an. Diese Erfahrung könnte dazu beitragen, dass auch die realen Lernbesonderheiten von Schüler*innen in der späteren Berufspraxis besser erkannt werden. Damit könnte ein wichtiger Schritt dafür getan sein, um Schüler*innen wie Sarah oder Ricardo, die eingangs vorgestellt wurden, das Lernen und Leben zu erleichtern.

Literatur

- Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57, 175–193. <https://doi.org/10.2378/peu2010.art13d>
- Anderson, L. W., Bloom, B. S., & Krathwohl, D. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological review*, 64(6p1), 359-372. <https://doi.org/10.1037/h0043445>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
- Bandura, A. (2010). Self-efficacy. In I. B. Weiner & W. E. Craighead (Hrsg.), *The Corsini encyclopedia of psychology* (Bd. 4, S. 1534–1536). John Wiley & Sons.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Behrmann, L., & van Ophuysen, S. (2017). Das Vier-Komponenten-Modell der Diagnosequalität. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (S. 38–42). Waxmann.
- Berner, S. (2018). Spirale des Scheiterns. *Grundschule*, 3, 10–11.
- Beywl, W., & Schepp-Winter, E. (2000). Zielgeführte Evaluation von Programmen. *Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg.): Materialien zur Qualitätssicherung in der Kinder- und Jugendhilfe. Qs*, 29.
- 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>
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company.
- BMBF - Bundesministerium für Bildung und Forschung (2016). *Neue Wege in der Lehrerbildung*. Die Qualitätsoffensive Lehrerbildung. https://www.qualitaetsoffensive-lehrerbildung.de/lehrerbildung/shareddocs/downloads/files/bmbf-neue_wege_in_der_lehrerbildung_barrierefrei.pdf?__blob=publicationFile&v=1

- Böhmer, I., Hörstermann, T., Gräsel, C., Krolak-Schwerdt, S., & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? *Journal for Educational Research Online*, 7(2), 59–81. <https://doi.org/10.25656/01:11490>
- Böhmer, M., Englich, B., & Böhmer, I. (2017). Schülerbeurteilungen aus der Perspektive dualer Prozessmodelle der sozialen Urteilsbildung. In Anna Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 50–54). Waxman.
- Bohner, G. (2003). Einstellungen. In W. Stroebe, K. Jonas, & M. Hewstone (Hrsg.), *Sozialpsychologie. Eine Einführung* (3. Aufl., S. 265–315). Springer.
- Bosse, S., & Spörer, N. (2014). Erfassung der Einstellung und der Selbstwirksamkeit von Lehramtsstudierenden zum inklusiven Unterricht. *Empirische Sonderpädagogik*, 6(4), 279–299. <https://doi.org/10.25656/01:10019>
- Bromme, R. (2008). Lehrerexpertise. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 159–167). Hogrefe.
- Bruder, S., Klug, J., Hertel, S., & Schmitz, B. (2010). Modellierung der Beratungskompetenz von Lehrkräften. Projekt Beratungskompetenz. In E. Klieme, D. Leutner, & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes (Zeitschrift für Pädagogik, Beiheft; 56)* (S. 274–285). Beltz.
- Cohen, J. W. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). L. Erlbaum Associates.
- Cohen, N. J., Eichenbaum, H., Deacedo, B. S., & Corkin, S. (1985). Different memory systems underlying acquisition of procedural and declarative knowledge. *Annals of the New York Academy of Sciences*, 444(1), 54–71. <https://doi.org/10.1111/j.1749-6632.1985.tb37579.x>
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- DGKJP - Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie, Psychosomatik und Psychotherapie e.V. (2015). *S3-Leitlinie: Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese- und/oder Rechtschreibstörung*. https://www.awmf.org/uploads/tx_szleitlinien/028-044L_S3_Lese-Rechtschreibst%C3%B6rungen_Kinder_Jugendliche_2015-06-abgelaufen.pdfDGKJP - Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie, Psychosomatik und Psychotherapie e V. (DGKJP). (2018). *S3-Leitlinie: Diagnostik und Behandlung der Rechenstörung*. https://www.awmf.org/uploads/tx_szleitlinien/028-046L_S3_Rechenst%C3%B6rung-2018-03_1.pdf

- DIMDI - Deutsches Institut für Medizinische Dokumentation und Information (2018). *Kapitel V. Psychische und Verhaltensstörungen (F00-F99). Entwicklungsstörungen (F80-F89)*.
<https://www.dimdi.de/static/de/klassifikationen/icd/icd-10-gm/kode-suche/htmlgm2018/block-f80-f89.htm>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humannwissenschaften* (5. Aufl.). Springer.
- Dünnebier, K., Gräsel, C., & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung: Eine experimentelle Studie zu Ankereffekten. *Zeitschrift für Pädagogische Psychologie*, 23(34), 187–195. <https://doi.org/10.1024/1010-0652.23.34.187>
- Egener, L. (2018). *Einstellungen und Selbstwirksamkeitserwartungen von Studierenden des Lehramts für sonderpädagogische Förderung zur schulischen Inklusion*. Universität Paderborn.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und Forschungsmethoden*. Beltz.
- Faller, H. (2005). Sensitivität, Spezifität, positiver und negativer Vorhersagewert. *Die Rehabilitation*, 22(1), 44–49. <https://doi.org/10.1055/s-2004-834719>
- Forlin, C., Cedillo, I. G., Romero-Contreras, S., Fletcher, T., & Rodriguez Hernández, H. J. (2010). Inclusion in Mexico: Ensuring supportive attitudes by newly graduated teachers. *International Journal of inclusive education*, 14(7), 723–739. <https://doi.org/10.1080/13603111003778569>
- Förster, N., & Souvignier, E. (2015). Effects of providing teachers with information about their students' reading progress. *School Psychology Review*, 44(1), 60-75. <https://doi.org/10.1080/13603111003778569>
- Förster, N., & Souvignier, E. (2017). Förderung diagnostischer Kompetenz durch Bereitstellung formativer Diagnostik. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 231–239). Waxman.
- Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction*, 32, 91–100. <https://doi.org/10.1016/j.learninstruc.2014.02.002>
- Glessmer, M. S., & Lüth, T. (2016). Lernzieltaxonomische Klassifizierung und gezielte Gestaltung von Fragen. *Zeitschrift für Hochschulentwicklung*, 11(5). <https://doi.org/10.3217/zfhe-11-05/12>
- Glogger-Frey, I., & Renkl, A. (2017). Diagnostische Kompetenz fördern – Vorwissen aufgreifende Methoden in Kombination mit beispielbasiertem Kurztraining. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (S. 217–222). Waxmann.
- Goldman, S. R., & Hasselbring, T. S. (1997). Achieving meaningful mathematics literacy for students with learning disabilities. *Journal of learning disabilities*, 30(2), 198–208. <https://doi.org/10.1177/002221949703000207>

- Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: theory and practice*, 15(2), 273–289. <https://doi.org/10.1080/13540600902875340>
- Grünke, M., & Grosche, M. (2014). Lernbehinderung. In G. W. Lauth, M. Grünke, & J. C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen: Förderung, Training und Therapie in der Praxis* (2. Aufl., S. 76–89). Hogrefe.
- GsVO. (2005, 2019). *Verordnung über den Bildungsgang der Grundschule (Grundschulverordnung—GsVO) Vom 19. Januar 2005 i. D. F. v. 20.09.2019*.
<http://gesetze.berlin.de/jportal/?quelle=jlink&query=GrSchulV+BE&psml=bsbeprod.psml&max=true&caiz=true>
- Gummer, E. S., & Mandinach, E. B. (2015). Building a Conceptual Framework for Data Literacy. *Teachers College Record*, 117, 1-22. <https://doi.org/10.1177/016146811511700401>
- Halbherr, T., Dittmann-Domenichini, N., Piendl, N., & Schlienger, C. (2016). Authentische, kompetenzorientierte Online-Prüfungen an der ETH Zürich. *Zeitschrift für Hochschulentwicklung*, 11(2), 247–269.
<https://doi.org/10.3217/zfhe-11-02/15>
- Hartinger, A., Kleickmann, T., & Hawelka, B. (2006). Der Einfluss von Lehrervorstellungen zum Lernen und Lehren auf die Gestaltung des Unterrichts und auf motivationale Schülervariablen. *Zeitschrift für Erziehungswissenschaft*, 9(1), 110–126. <https://doi.org/10.1007/s11618-006-0008-1>
- Hascher, T. (2008). Diagnostische Kompetenz im Lehrberuf. In C. Kraler, *Wissen erwerben, Kompetenzen entwickeln. Modelle zur kompetenzorientierten Lehrerbildung* (S. 71–86). Waxmann.
- Hasselhorn, M., Schneider, W., & Trautwein, U. (2014). *Lernverlaufsdiagnostik* (Bd. 12). Hogrefe.
- Hasselhorn, M., & Schuchardt, K. (2006). Lernstörungen. *Kindheit und Entwicklung*, 15(4), 208–215.
<https://doi.org/10.1026/0942-5403.15.4.208>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81–112.
<https://doi.org/10.3102/003465430298487>
- Hattie, J. (2003). *Teachers Make a Difference, What is the research evidence?* ACER Research Conference.
- Heine, A., Engl, V., Thaler, V. M., Fussenegger, B., & Jacobs, A. M. (2012). *Neuropsychologie von Entwicklungsstörungen schulischer Fertigkeiten (Fortschritte der Neuropsychologie)*. Hogrefe.
- Heine, A., & Jacobs, A. M. (2011). *Lehr-Lern-Forschung unter neurowissenschaftlicher Perspektive. Ergebnisse der zweiten Förderphase des Programms NIL: Neurowissenschaften—Instruktion—Lernen*. Waxmann Verlag.
- Hellmich, F., & Görel, G. (2014). Erklärungsfaktoren für Einstellungen von Lehrerinnen und Lehrern zum inklusiven Unterricht in der Grundschule. *Zeitschrift für Bildungsforschung*, 4(3), 227–240.
<https://doi.org/10.1007/s35834-014-0102-z>
- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Kallmeyersche Verlagsbuchhandlung.

- Helmke, A. (2004). Von der Evaluation zur Innovation: Pädagogische Nutzbarmachung von Vergleichsarbeiten in der Grundschule. *Das Seminar*, 2, 90–112.
- Helmke, A. (2007). Aktive Lernzeit optimieren: Was wissen wir über effiziente Klassenführung? *Pädagogik*, 59(5), 44–49.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Klett-Kallmeyer.
- Helmke, A., & Hosenfeld, I. (2005). Standardbezogene Unterrichtsevaluation. In G. Brägger, B. Bucher, & N. Landwehr, *Schlüsselfragen zur externen Schulevaluation* (S. 127–151). hep.
- Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2004). Schulleitung und Schulentwicklung: Voraussetzungen, Bedingungen, Erfahrungen. In R. Arnold & C. Griese (Hrsg.), *Schulmanagement und Schulentwicklung*. (S. 105–119). Schneider-Verlag.
- Helmke, A., & Lenske, G. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beiträge zur Lehrerinnen-und Lehrerbildung*, 31(2), 214–233. <https://doi.org/10.25656/01:13848>
- Hemmerich, W. (2015). *StatistikGuru: Cohen's d berechnen*. Abgerufen am 29.03.2022 von <https://statistikguru.de/rechner/cohens-d.html>
- Herppich, S., Altmann, A., Wittwer, J., & Nückles, M. (2017a). Förderung von Instruktionsstrategien zum verbesserten Diagnostizieren im Eins-zu-Eins-Tutoring. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 203–208). Waxman.
- Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2013). Does it make a difference? Investigating the assessment accuracy of teacher tutors and student tutors. *The Journal of Experimental Education*, 81(2), 242–260. <https://doi.org/10.1080/00220973.2012.699900>
- Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>
- Herppich, S., Praetorius, A.-K., Hetmanek, A., Glogger-Frey, I., Ufer, S., Leutner, D., Behrmann, L., Böhmer, I., Böhmer, M., & Förster, N. (2017b). Ein Arbeitsmodell für die empirische Erforschung der diagnostischen Kompetenz von Lehrkräften. *Pädagogische Psychologie und Entwicklungspsychologie*, 94, 75–93.
- Hesse, I., & Latzko, B. (2017). *Diagnostik für Lehrkräfte*. Verlag Barbara Budrich, UTB.
- Hetmanek, A., & Van Gog, T. (2017). Förderung von diagnostischer Kompetenz: Potenziale von Ansätzen aus der medizinischen Ausbildung. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 209–216). Waxmann.

- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of educational research*, 59(3), 297-313. <https://doi.org/10.2307/1170184>
- Hosenfeld, I., & Groß Ophoff, J. (2007). *Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht*. Verlag Empirische Pädagogik.
- Hosenfeld, I., Schrader, F.-W., & Tuyet, H. (2006). Von der Rezeption zur Ergebnisrückmeldung: Leistungsevaluation im Spannungsfeld von System-Monitoring und Schulentwicklung. In I. Hosenfeld & F.-W. Schrader (Hrsg.), *Schulische Leistung. Grundlagen, Bedingunge, Perspektiven* (S. 289–315). Waxman.
- Ihme, J. M., & Senkbeil, M. (2017). Warum können Jugendliche ihre eigenen computerbezogenen Kompetenzen nicht realistisch einschätzen? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 49(1), 24–37. <https://doi.org/10.1026/0049-8637/a000164>
- Ingenkamp, K. (1985). Erfassung und Rückmeldung des Lernerfolgs. In D. Lenzen (Hrsg.), *Enzyklopädie Erziehungswissenschaft* (Bd. 4, S. 173–204). Klett.
- Ingenkamp, K. H., & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6. Aufl.). Beltz.
- IQB, Institut zur Qualitätssicherung im Bildungswesen (2018). VERA - Ein Überblick. Abgerufen am 29.03.2022 von <https://www.iqb.hu-berlin.de/vera>
- Ise, E., Engel, R. R., & Schulte-Körne, G. (2012). Was hilft bei der Lese-Rechtschreibstörung? *Kindheit und Entwicklung*, 21, 122–136. <https://doi.org/10.1026/0942-5403/a000077>
- Jäger, R. S. (2006). Diagnostischer Prozess. In F. Petermann & M. Eid, *Handbuch der psychologischen Diagnostik* (S. 89–96). Hogrefe.
- Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal processes. In R. Schwarzer (Hrsg.), *Self-efficacy: Thought control of action* (S. 195–216). Routledge.
- Jürgens, E., & Lissmann, U. (2015). *Pädagogische Diagnostik*. Beltz.
- K2teach - Know how to teach (2018). *K2teach - Know how to teach*. Abgerufen am 29.03.2022 von <https://www.fu-berlin.de/sites/k2teach/index.html>
- K2teach. (2019, Februar 13). *Infopoint*. Abgerufen am 29.03.2022 von <https://www.fu-berlin.de/sites/k2teach/projektergebnisse/index.html>
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kaiser, J., Helm, F., Retelsdorf, J., Südkamp, A. & Möller, J. (2012). Zum Zusammenhang von Intelligenz und Urteilsgenauigkeit bei der Beurteilung von Schülerleistungen im Simulierten Klassenraum. *Zeitschrift für Pädagogische Psychologie*, 26(4), 251–261. <https://doi.org/10.1024/1010-0652/a000076>

- Kaiser, J., Möller, J., Helm, F., & Kunter, M. (2015). Das Schülerinventar: Welche Schülermerkmale die Leistungsurteile von Lehrkräften beeinflussen. *Zeitschrift für Erziehungswissenschaft*, 18(2), 279–302. <https://doi.org/10.1007/s11618-015-0619-5>
- Kaiser, J., Praetorius, A. K., Südkamp, A., & Ufer, S. (2017). Die enge Verwobenheit von diagnostischem und pädagogischem Handeln als Herausforderung bei der Erfassung diagnostischer Kompetenz. In Anna Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (S. 114–123).
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(34), 197–209. <https://doi.org/10.1024/1010-0652.23.34.197>
- Karing, C., & Seidel, T. (2017). Förderung diagnostischer Kompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 201–202). Waxmann.
- Karst, K., & Förster, N. (2017). Ansätze zur Modellierung diagnostischer Kompetenz. In Anna Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 19–21). Waxman.
- Karst, K., Klug, J., & Ufer, S. (2017). Strukturierung diagnostischer Situationen im inner- und außerunterrichtlichen Handeln von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 102–103). Waxmann.
- Khan, S. B., & Weiss, J. (1973). The teaching of affective responses. In R. M. W. Travers (Hrsg.), *Second handbook of research on teaching* (S. 759–804). Rand McNally.
- Kippers, W. B., Poortman, C. L., Schildkamp, K., & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? *Studies in educational evaluation*, 56, 21–31. <https://doi.org/10.1016/j.stueduc.2017.11.001>
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating Training Programs: The four Levels* (3. Aufl.). Berrett-Koehler Publishers.
- Kittel, D., & Rollett, W. (2017). Entwickelt sich die professionelle Reflexionskompetenz von Lehrkräften durch die Teilnahme an einem berufsbegleitenden Weiterbildungsstudiengang? *Zeitschrift für Hochschule und Weiterbildung*, (2), 45-51. <https://doi.org/10.25656/01:16603>
- Kittel, D., Rollett, W., & Groß Ophoff, J. (2017). Profitieren berufstätige Lehrkräfte durch ein berufsbegleitendes weiterbildendes Studium in ihren Forschungskompetenzen? *Bildung und Erziehung*, 70(4), 437–452.
- Klempin, C., Rehfeldt, D., Seibert, D., Brämer, M., Köster, H., Lücke, M., ... & Sambanis, M. (2019). Stabilisierung der Selbstwirksamkeitserwartung über Komplexitätsreduktion. *Unterrichtswissenschaft*, 1-27.

- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903.
- Klug, J. (2017). Tagebücher zum Diagnostizieren und zur Förderung von Diagnostik und Lernerfolg. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 223–231). Waxman.
- Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education*, 30, 38–46. <https://doi.org/10.1016/j.tate.2012.10.004>
- Klug, J., Bruder, S., & Schmitz, B. (2016). Which variables predict teachers diagnostic competence when diagnosing students' learning behavior at different stages of a teacher's career? *Teachers and Teaching*, 22(4), 461–484. <https://doi.org/10.1080/13540602.2015.1082729>
- KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004 (i.d.F vom 12.06.2014)*. http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Köller, M., Köller, O., & Baumert, J. (2016). Professionelle Kompetenz von Lehrkräften. In J. Möller, M. Köller, & T. Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung: Schule und Unterricht—Lehren und Lernen* (S. 9–22). Klett Kallmeyer.
- Köller, O., Baumert, J., & Neubrand, J. (2000). Epistemologische Überzeugungen und Fachverständnis im Mathematik- und Physikunterricht. In J. Baumert, W. Bos, & R. Lehmann (Hrsg.), *Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 229–269). Leske & Budrich.
- Krapp, A. (1979). *Prognose und Entscheidung: Zur theoretischen Begründung und Differenzierung der pädagogisch-psychologischen Prognose*. Beltz.
- Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L. W. Anderson & L. A. Sosniak (Hrsg.), *Bloom's taxonomy: A forty-year retrospective: Ninety-third yearbook of the National Society for the Study of Education* (S. 64–81). NSSE.
- Kretschmann, R. (2015). Vorwort. In H. Schäfer & C. Rittmeyer (Hrsg.), *Handbuch Inklusive Diagnostik* (S. 9–11). Beltz Verlag.
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess: Der Lehrer als «flexibler Denker». *Zeitschrift für Pädagogische Psychologie*, 23(34), 175–186. <https://doi.org/10.1024/1010-0652.23.34.175>

- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: A social cognitive analysis. *Social Psychology of Education, 16*(2), 215–239. <https://doi.org/10.1007/s11218-013-9215-9>
- Kube, J. I. (2009). *Vornamensforschung: Fragebogenuntersuchung bei Lehrerinnen und Lehrern, ob Vorurteile bezüglich spezifischer Vornamen von Grundschulern und davon abgeleitete erwartete spezifische Persönlichkeitsmerkmale vorliegen*. Universität Oldenburg.
- Kunina-Habenicht, O., Lohse-Bossenz, H., Kunter, M., Dicke, T., Förster, D., Gößling, J., Schulze-Stocker, F., Schmeck, A., Baumert, J., Leutner, D., & Terhart, E. (2012). Welche bildungswissenschaftlichen Inhalte sind wichtig in der Lehrerbildung? Ergebnisse einer Delphi-Studie. *Zeitschrift für Erziehungswissenschaft, 15*, 649–682. <https://doi.org/10.1007/s11618-012-0324-6>
- Kunter, M. (2011). Forschung zur Lehrermotivation. In E. Terhart, H. Bennewitz, & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl.). Waxman.
- Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 55–68). Waxmann.
- Kunter, M., Leutner, D., Seidel, T., & Terhart, E. (2014). *Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung (BilWiss). Dokumentation der Erhebungsinstrumente für den ersten und zweiten Messzeitpunkt*. Frankfurt. https://www.iqb.hu-berlin.de/fdz/studies/BilWiss/BilWiss_Skalenha.pdf
- Kunter, M. (2014). Forschung zur Lehrermotivation. In E. Terhart, H. Bennewitz, & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 527–539). Waxmann.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805. <https://doi.org/10.1037/a0032583>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*(863). <https://doi.org/10.3389/fpsyg.2013.00863>
- Landerl, K., Kaufmann, L., & Vogel, S. (2017). *Dyskalkulie: Modelle, Diagnostik, Intervention* (Bd. 3066). UTB.
- Langfeld, H.-P. (2017). *Diagnostischer Prozess*. Dorsch Lexikon Psychologie, Hogrefe. Abgerufen am 29.03.2022 von <https://m.portal.hogrefe.com/dorsch/diagnostischer-prozess/>
- Latzko, B., & Hesse, I. (2009). *Diagnostische Kompetenzen als Grundlage für die Gestaltung inklusiven Unterrichts*. Vortrag abgerufen von https://uol.de/fileadmin/user_upload/diz/download/Veranstaltungen/Schulische_Inklusion/Diagnostische_Kompetenzen_inklusive_Unterricht.pdf

- Lenske, G., Thillmann, H., Wirth, J., Dicke, T., & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests. *Zeitschrift für Erziehungswissenschaft*, *18*, 225–245. <https://doi.org/10.1007/s11618-015-0627-5>
- Linninger, C., Dicke, T., Kunina-Habenicht, O., Schulze-Stocker, F., Leutner, D., Terhart, E., Emmenlauer, S., Seidel, T., & Kunter, M. (2015). Assessing Teachers' Educational Knowledge. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *47*(2), 72–83. <https://doi.org/10.1026/0049-8637/a000126>
- Lipowsky, F. (2014). Theoretische Perspektiven und empirische Befunde zur Wirksamkeit von Lehrerfort- und -weiterbildung. In E. Terhart, H. Bennewitz, & M. Rothland, *Handbuch der Forschung zum Lehrerberuf* (S. 398–417). Waxmann.
- Lorenz, C., & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 211–222. <https://doi.org/10.1024/1010-0652.23.34.211>
- LISUM - Landesinstitut für Schule und Medien Berlin-Brandenburg (2018). *ILeA – Individuelle Lernstandsanalysen*. Abgerufen am 22.03.2022 von <https://bildungsserver.berlin-brandenburg.de/ilea/>
- Lukesch, N. (1998). *Einführung in die pädagogisch-psychologische Diagnostik* (2. Aufl.). Roderer.
- Maier, U. (2008). Rezeption und Nutzung von Vergleichsarbeiten aus der Perspektive von Lehrkräften. *Zeitschrift für Pädagogik*, *54*(1), 95–117. <https://doi.org/10.25656/01:4338>
- Mamede, S., Schmidt, H. G., & Penaforte, J. C. (2008). Effects of reflective practice on the accuracy of medical diagnosis. *Medical Education*, *42*, 468–475. <https://doi.org/10.1111/j.1365-2923.2008.03030.x>
- Mamede, S., Van Gog, T., Moura, A. S., de Faria, R. M., Peixoto, J. M., Rikers, R. M., & Schmidt, H. G. (2012). Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Medical education*, *46*(5), 464–472. <https://doi.org/10.1111/j.1365-2923.2012.04217.x>
- Mamede, S., Van Gog, T., Moura, A. S., de Faria, R. M., Peixoto, J. M., & Schmidt, H. G. (2014). How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Academic Medicine*, *89*(1), 121–127. <https://doi.org/10.1097/ACM.0000000000000076>
- Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376. <https://doi.org/10.1016/j.tate.2016.07.011>
- May, P. (2012). *Hamburger Schreibporbe (HSP 1—10). Manual/ Handbuch Diagnose orthografischer Kompetenz. Manual/ Handbuch Diagnose orthografischer Kompetenz*. vpm.

- Miller, S. P. & Hudson, P. J. (2007). Using Evidence-Based Practices to Build Mathematics Competence Related to Conceptual, Procedural, and Declarative Knowledge. *Learning Disabilities Research & Practice*, 22(1), 47–57. <https://doi.org/10.1111/j.1540-5826.2007.00230.x>
- Müller, K. (2010). *Das Praxisjahr in der Lehrerbildung: Empirische Befunde zur Wirksamkeit studienintegrierter Langzeitpraktika*. Julius Klinkhardt.
- Naugle, K. A., Naugle, L. B. & Naugle, R. J. (2000). Kirkpatrick's evaluation model as a means of evaluating teacher performance. *Education*, 121(1), 135–144.
- Naugle, K. A., Naugle, L. B., & Naugle, R. J. (2006). Kirkpatrick's Evaluation model as a means of evaluating teacher performance. In D. L. Kirkpatrick & J. D. Kirkpatrick, *Evaluating Training Programs: The four Levels* (3. Aufl., S. 135–144). Berrett-Koehler Publishers.
- Ohle, A., McElvany, N., Horz, H., & Ullrich, M. (2015). Text-picture integration - Teachers' attitudes, motivation and self-related cognitions in diagnostics. *Journal for educational research online*, 7(2), 11–33. <https://doi.org/10.25656/01:11488>
- Opalinski, S., & Scharenberg, K. (2018). Veränderung inklusionsbezogener Überzeugungen bei Studierenden durch diversitätssensible Lehrveranstaltungen. *Bildung und Erziehung*, 71(4), 449–464. <https://doi.org/10.13109/buer.2018.71.4.449>
- Oser, F. (1997). Standards in der Lehrerbildung. Teil 1: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerinnen-und Lehrerbildung*, 15(1), 26–37. <https://doi.org/10.25656/01:13343>
- Oser, F. (2001). Standards: Kompetenzen von Lehrpersonen. In J. Oelkers & F. Oser, *Die Wirksamkeit der Lehrerbildungssysteme*. Rüegger.
- Pawlik, K. (2006). *Handbuch Psychologie: Wissenschaft-Anwendung-Berufsfelder; mit 46 Tabellen*. Springer.
- Pezzulo, G. (2011). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind & Language*, 28(1), 78–114. <https://doi.org/10.1111/j.1468-0017.2010.01411.x>
- Praetorius, A.-K., Hetmanek, A., Herppich, S., & Ufer, S. (2017). Herausforderungen bei der empirischen Erforschung diagnostischer Kompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 95–102). Waxmann.
- Renkl, A. (2011). Aktives Lernen: Von sinnvollen und weniger sinnvollen theoretischen Perspektiven zu einem schillernden Konstrukt. *Unterrichtswissenschaft*, 39(3), 197–212.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive science*, 38(1), 1–37. <https://doi.org/10.1111/cogs.12086>
- Retelsdorf, J., & Möller, J. (2016). Diagnostik und Leistungsbeurteilung. In J. Möller, M. Möller, & T. Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung: Schule und Unterricht-Lehren und Lernen* (S. 23–38). Kallmeyer.

- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Hrsg.), *Handbook of research on teacher education* (2. Aufl., S. 102–119). Macmillan.
- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*(1), 175. <https://doi.org/10.1037/0022-0663.91.1.175>
- Ritts, V., Patterson, M. L., & Tubbs, M. E. (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research*, *62*(4), 413–426. <https://doi.org/10.2307/1170486>
- Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Hrsg.), *Attitude organization and change: An analysis of consistency among attitude components. (Yales studies in attitude and communication)* (Bd. 3). Yale University Press.
- Rost, D. H., & Buch, S. (2010). Hochbegabung. In *Handwörterbuch pädagogische Psychologie* (S. 257ff). Beltz.
- Rost, D. H., & Hanses, P. (1998). Wer nichts leistet, ist nicht begabt? Zur Identifikation hochbegabter Underachiever durch Lehrkräfte. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *29*(2), 167–177.
- Rothland, M. (2019, Mai). *Lehrerbildung zwischen Wissenschaft und Berufspraxis: Wächst zusammen, was (nicht) zusammen gehört? Profildokumentation Lebramt – Konturen für Studium, Fächer, Universität*. Programmworkshop der Qualitätsoffensive Lehrerbildung an der Christian-Albrechts-Universität zu Kiel.
- Rudolph, U., Böhm, R., & Lummer, M. (2007). Ein Vorname sagt mehr als 1000 Worte. Zur sozialen Wahrnehmung von Vornamen. *Zeitschrift für Sozialpsychologie*, *38*(1), 17–31. <https://doi.org/10.1024/0044-3514.38.1.17>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68. <https://doi.org/10.1037/0003-066X.55.1.68>
- Sacher, W. (2009). *Leistungen entwickeln, überprüfen und beurteilen. Bewährte und neue Wege für die Primar- und Sekundarstufe* (5. Aufl.). Klinkhardt.
- Schäfer, H., & Rittmeyer, C. (2015). *Handbuch Inklusive Diagnostik*. Beltz Verlag.
- Schildkamp, K., Lai, M. K., & Earl, L. (2013). *Studies in educational leadership. Volume 17: Data-based decision making in education. Challenges and opportunities*.
- Schladitz, S., Groß Ophoff, J., & Wirtz, M. (2015). Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz. In Sigrid Blömeke & O. Zlatkin-Troitschanskaia, *Zeitschrift für Pädagogik*, *61*. Beibef: *Kompetenzen von Studierenden* (S. 167–184).
- Schnotz, W. (2006). Conceptual Change. In Detlef H. Rost, J. R. Sparfeldt, & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 77–82). Beltz.

- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Peter Lang.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 237–245. <https://doi.org/10.1024/1010-0652.23.34.237>
- Schrader, F.-W. (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost, *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 102–109). Beltz.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 154–165. <https://doi.org/10.25656/01:13843>
- Schrader, F.-W. (2017). Diagnostische Kompetenz von Lehrkräften – Anmerkungen zur Weiterentwicklung des Konstrukts. In Anna Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 247–255). Waxman.
- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.
- Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Freie Universität Berlin.
- Sedlmeier, P., & Renkewitz, F. (2008). *Forschungsmethoden und Statistik in der Psychologie*. Pearson Studium.
- SenBJF - Senatsverwaltung für Bildung, Jugend und Familie. (2018). *Fördermaßnahmen konkret! Eine Handreichung für pädagogische Fachkräfte zur Entwicklung von Fördermaßnahmen*. https://www.berlin.de/sen/bildung/schule/foerderung/sonderpaedagogische-foerderung/fachinfo/foerdermassnahmen_konkret.pdf
- SenBJF - Senatsverwaltung für Bildung, Jugend und Familie. (2019). *Schwierigkeiten im Lesen, Rechtschreiben und Rechnen. Leitfaden zur Diagnostik mit Hinweisen zum Nachteilsausgleich und Notenschutz*. https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/faecher/sprachen/deutsch/LRS/Schwierigkeiten_im_Lesen_Rechtschreiben_und_Rechnen_-_Leitfaden_Nachteilsausgleich_und_Notenschutz_Dezember_2019.pdf
- SenBJW - Senatsverwaltung für Bildung, Jugend und Wissenschaft. (2012). *Leitfaden zur Feststellung sonderpädagogischen Förderbedarfs an Berliner Schulen, Förderschwerpunkt Lernen*. https://www.berlin.de/sen/bildung/schule/foerderung/sonderpaedagogische-foerderung/fachinfo/leitfaden_foerderbedarf-2017.pdf
- Shavelson, R. (2006). On the integration of formative assessment in teaching and learning: Implications for new pathways in teacher education. In F. K. Oser, F. Achtenhagen, & U. Renold (Hrsg.), *Competence Oriented Teacher Training: Old Research Demands and New Pathways* (S. 61–78). Brill, Sense.

- Souvignier, E., & Förster, N. (2011). Effekte prozessorientierter Diagnostik auf die Entwicklung der Lesekompetenz leseschwacher Viertklässler. *Empirische Sonderpädagogik*, 3(3), 243–255.
<https://doi.org/10.25656/01:9326>
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95.
<https://doi.org/10.1024/1010-0652.19.12.85>
- Sprietsma, M. (2009). *Discrimination in Grading? Experimental Evidence from Primary School* (SSRN Scholarly Paper ID 1520546). Social Science Research Network. <https://papers.ssrn.com/abstract=1520546>
- Stangl, W. (2018a). *Deklaratives Gedächtnis*. Online Lexikon für Psychologie und Pädagogik. Abgerufen am 29.03.2022 von <https://lexikon.stangl.eu/7511/deklaratives-gedaechtnis/>
- Stangl, W. (2018b). *Prozedurales Gedächtnis*. Online Lexikon für Psychologie und Pädagogik. Abgerufen am 29.03.2022 von <https://lexikon.stangl.eu/7415/prozedurales-gedaechtnis/>
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795–819.
<https://doi.org/10.1002/pits.20113>
- Stern, E., Schalk, L., & Schumacher, R. (2016). Lernen. In Jens Möller, M. Köller, & T. Riecke-Baulecke, *Basiswissen Lehrerbildung: Schule und Unterricht – Lehren und Lernen* (S. 106–117). Klett-Kallmeyer.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' Judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>
- Südkamp, A., & Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum: Direkte und indirekte Einschätzungen von Schülerleistungen. *Zeitschrift für Pädagogische Psychologie*, 23(34), 161–174.
<https://doi.org/10.1024/1010-0652.23.34.161>
- Südkamp, A., Möller, J. & Pohlmann, B.. (2008). Der Simulierte Klassenraum Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22(3–4), 261–276.
<https://doi.org/10.1024/1010-0652.22.34.261>
- Südkamp, A., & Praetorius, A.-K. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In Anna Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 13–18). Waxman.
- Tabachnick, B. G., & Fidell, L. S. (2006). *Using Multivariate Statistics* (5. Aufl.). Allyn & Bacon.
- Thoren, K., Wißmann, J., Harks, M., Wenger, M., Kinder, A., & Hannover, B. (2020). Förderung von Datennutzungskompetenzen in der Lehramtsausbildung: Konzeption und Evaluation dreier Seminare. In I. Gogolin, B. Hannover, & A. Scheunpflug, *Evidenzbasierung in der Lehrkräftebildung* (Bd. 4, S. 39–71). Springer. https://doi.org/10.1007/978-3-658-22460-8_3

- Tillema, H. H. (1995). Changing the professional knowledge and beliefs of teachers: A training study. *Learning and instruction*, 5(4), 291–318. [https://doi.org/10.1016/0959-4752\(95\)00020-8](https://doi.org/10.1016/0959-4752(95)00020-8)
- Verplanken, B., Hofstee, G., & Janssen, H. J. (1998). Accessibility of affective versus cognitive components of attitudes. *European journal of social psychology*, 28(1), 23–35. [https://doi.org/10.1002/\(SICI\)1099-0992\(199801/02\)28:1<23::AID-EJSP843>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1099-0992(199801/02)28:1<23::AID-EJSP843>3.0.CO;2-Z)
- Vohrmann, A. (2018). *Zeigt, was ihr könnt! Wirkung eines Motivations- und Selbststeuerungstrainings für besonders begabte Underachiever (MoSt) in Form eines Kleingruppentrainings im schulischen Kontext*. Waxmann Verlag.
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing Teacher Candidates' General Pedagogical/Psychological Knowledge: Test Construction and Validation. *Journal of Educational Psychology*, 103(4), 952–969. <https://doi.org/10.1037/a0025125>
- Webb, N. L. (2002, April). *Assessment Literacy in a Standards-Based Urban Education Setting*. Beitrag präsentiert auf Annual meeting of the American Educational Research Association, New Orleans, Vereinigte Staaten.
- Wedel, A., & Pfetsch, J. (2017, März). *Förderung pädagogischer Diagnostik von Lehramtsstudierenden mitfallbasiertem, problemorientierten Lernen*. Beitrag präsentiert auf der GEBF-Tagung „Durch Bildung gesellschaftliche Herausforderungen meistern“, Heidelberg, Deutschland.
- Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A. (2020). Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung—Effekte problemorientierten Lernens mit Textfällen. In *Evidenzbasierung in der Lehrkräftebildung* (pp. 95-121). Springer VS.
- Weidlich, J., & Spannagel, C. (2014). Die Vorbereitungsphase im Flipped Classroom. Vorlesungsvideos versus Aufgaben. In K. Rummler (Hrsg.), *Lernräume gestalten –Bildungskontexte vielfältig denken* (S. 237–249). Waxmann.
- Weinert, F. E. (2001). *Leistungsmessung in Schulen*. Beltz.
- Werning, R. (2017). Inklusive Beschulung—Eine Herausforderung. In L. Huck & A. Schulz (Hrsg.), *Lerntherapie und inklusive Schule* (S. 253–265). Duden-Verlag.
- Westphal, A., Zuber, J., & Vock, M. (2018). Welche Rolle spielen Selbstwirksamkeit, Motivation und Einstellungen zu Diagnostik für die Nutzung datenbasierter Rückmeldungen? *Zeitschrift für Bildungsforschung*, 8(3), 289–307. <https://doi.org/10.1007/s35834-018-0223-x>
- Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Hrsg.), *Development of achievement motivation* (S. 91–120). Elsevier.
- Wild, K.-P. (1991). *Identifikation hochbegabter Schüler: Lehrer und Schüler als Datenquellen*. Asanger.

- Wißmann, J., & Kinder, A. (2018). *Ein diagnostisches Fallinventar - Übungen zur Diagnostik von Lernbesonderheiten. Handreichung zur Lehr-Lerngelegenheit*. https://www.fu-berlin.de/sites/k2teach/news/Handreichung_Fallinventar_JW_2018-12-18a.pdf
- Wißmann, J., & Pöhler, M. (2020, März). *Datennutzungskompetenzen in der Lehramtsausbildung fördern: Der Erkenntniswert von testdiagnostischen Verfahren in der Schule*. Poster präsentiert auf der 49. Jahrestagung der dghd, Berlin, Deutschland. Abrufbar unter https://www.fu-berlin.de/sites/lehrqualifizierung/dghd2020/poster/transfermarkt/TM_202_WISSMANNPO_HLER_Karte-1_Vorbereitung-diagnostischer-Aufgaben.pdf
- Wurster, S., Richter, D., & Lenski, A. E. (2017). Datenbasierte Unterrichtsentwicklung und ihr Zusammenhang zur Schülerleistung. *Zeitschrift für Erziehungswissenschaft*, 20(4), 628–650. <https://doi.org/10.1007/s11618-017-0759-x>
- Ziegler. (2017). *Hochbegabung* (2. Aufl.). Ernst Reinhardt.

Anhänge

Anhang 2A. Zusammenstellung der Fallhefte für Studie 2

Tab. 2A Übersicht über Versionen von Fallheften für die Untersuchung des Fallinventars

	Position der Fallvignette im Fallheft							
	1	2	3	4	5	6	7	8
Heft 1	Fall 2	Fall 5	Fall 3	Fall 7	Fall 8	Fall 6	Fall 1	Fall 4
1A	<i>Felix</i>	Julia	<i>Jan</i>	Sophie ¹	Hanna ¹	Lena ¹	<i>Luca</i> ¹	<i>Thomas</i> ¹
1B	Katrin	<i>Maximilian</i>	Leonie	<i>Lukas</i> ¹	<i>Jonas</i> ¹	<i>Christian</i> ¹	Jana ¹	Laura ¹
Heft 2	Fall 6	Fall 8	Fall 2	Fall 3	Fall 7	Fall 4	Fall 5	Fall 1
2A	Lena ¹	Hanna ¹	<i>Felix</i>	<i>Jan</i>	Sophie ¹	<i>Thomas</i> ¹	Julia	<i>Luca</i> ¹
2B	<i>Christian</i> ¹	<i>Jonas</i> ¹	Katrin	Leonie	<i>Lukas</i> ¹	Laura ¹	<i>Maximilian</i>	Jana ¹
Heft 3	Fall 3	Fall 1	Fall 5	Fall 2	Fall 4	Fall 8	Fall 6	Fall 7
3A	<i>Jan</i>	<i>Luca</i> ¹	Julia	<i>Felix</i>	<i>Thomas</i> ¹	Hanna ¹	Lena ¹	Sophie ¹
3B	Leonie	Jana ¹	<i>Maximilian</i>	Katrin	Laura ¹	<i>Jonas</i> ¹	<i>Christian</i> ¹	<i>Lukas</i> ¹
Heft 4	Fall 4	Fall 2	Fall 8	Fall 1	Fall 5	Fall 3	Fall 7	Fall 6
4A	<i>Thomas</i> ¹	<i>Felix</i>	Hanna ¹	<i>Luca</i> ¹	Julia	<i>Jan</i>	Sophie ¹	Lena ¹
4B	Laura ¹	Katrin	<i>Jonas</i> ¹	Jana ¹	<i>Maximilian</i>	Leonie	<i>Lukas</i> ¹	<i>Christian</i> ¹
Heft 5	Fall 1	Fall 6	Fall 7	Fall 4	Fall 3	Fall 2	Fall 8	Fall 5
5A	<i>Luca</i> ¹	Lena ¹	Sophie ¹	<i>Thomas</i> ¹	<i>Jan</i>	<i>Felix</i>	Hanna ¹	Julia
5B	Jana ¹	<i>Christian</i> ¹	<i>Lukas</i> ¹	Laura ¹	Leonie	Katrin	<i>Jonas</i> ¹	<i>Maximilian</i>
Heft 6	Fall 7	Fall 4	Fall 6	Fall 8	Fall 1	Fall 5	Fall 2	Fall 3
6A	Sophie ¹	<i>Thomas</i> ¹	Lena ¹	Hanna ¹	<i>Luca</i> ¹	Julia	<i>Felix</i>	<i>Jan</i>
6B	<i>Lukas</i> ¹	Laura ¹	<i>Christian</i> ¹	<i>Jonas</i> ¹	Jana ¹	<i>Maximilian</i>	Katrin	Leonie
Heft 7	Fall 8	Fall 3	Fall 1	Fall 5	Fall 6	Fall 7	Fall 4	Fall 2
7A	Hanna ¹	<i>Jan</i>	<i>Luca</i> ¹	Julia	Lena ¹	Sophie ¹	<i>Thomas</i> ¹	<i>Felix</i>
7B	<i>Jonas</i> ¹	Leonie	Jana ¹	<i>Maximilian</i>	<i>Christian</i> ¹	<i>Lukas</i> ¹	Laura ¹	Katrin
Heft 8	Fall 5	Fall 7	Fall 4	Fall 6	Fall 2	Fall 1	Fall 3	Fall 8
8A	Julia	Sophie ¹	<i>Thomas</i> ¹	Lena ¹	<i>Felix</i>	<i>Luca</i> ¹	<i>Jan</i>	Hanna ¹
8B	<i>Maximilian</i>	<i>Lukas</i> ¹	Laura ¹	<i>Christian</i> ¹	Katrin	Jana ¹	Leonie	<i>Jonas</i> ¹

Anmerkungen: In Heft 1 bis 8 wurden stets dieselben acht Fallvignetten vorgegeben, die je drei Schüler*innen ohne sowie fünf Schüler*innen mit einer Lernbesonderheit beschrieben (¹Schüler*in mit Lernbesonderheit). Die Fallvignetten wurden hierbei in variierender Reihenfolge vorgegeben, sodass über alle acht Fallhefte jeder Fall einmal an jeder Position vorkam (lateinisches Quadrat). Jedes Fallheft wurde außerdem in zwei parallelen Versionen präsentiert: Die jeweilige Version A wurde in der nicht-normativen Reihenfolgebedingung (Empfehlung vor Diagnose) und Version B in der normativen Bedingung (Empfehlung nach Diagnose) eingesetzt. Für die Versionen A und B wurden parallele weibliche und männliche Fallnamen verwendet, sodass in jedem Fallheft vier weibliche und vier männliche Fallvignetten enthalten waren (*Kursiv:* männliche Person).

Anhang 2B. Antworten für einzelne Fallvignetten in Studie 2

2B.1 Fall 1: Julia / Maximilian

Tab. 2B.1 Gewählte Antwortoptionen für Fallvignette 1

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	1	.8	ohne Angabe	2	1.6
normal begabte/r Schüler/in	36	28.1	sonderpädagogische Förderung	6	4.7
Lernstörung Lesen / Rechnen	26	20.3	sonderpädagogische Förderung, Förderung des Lesens	4	3.1
kombinierte Störung schulischer Fertigkeiten¹	40	31.3	sonderpädagogische Förderung, Förderung des Lesens, Förderung des Rechnens	1	.8
Schüler/in mit Lernbehinderung	2	1.6	sonderpädagogische Förderung, Förderung des Rechnens	1	.8
Hochbegabte/r Hochleister/in	0	.0	Förderung des Lesens	67	52.3
Hochbegabte/r Minderleister/in	16	12.5	Förderung des Lesens, Förderung des Rechnens²	33	25.8
mehrere Diagnosen ausgewählt	7	5.6	Förderung des Lesens, keine – nur Regelunterricht	1	.8
			Förderung des Rechnens	2	1.6
			Teilnahme AG „Club Einstein“	0	.0
			Teilnahme AG „Club Goethe“	0	.0
			keine – nur Regelunterricht	11	8.6
Gesamt	128	100.0	Gesamt	128	100.0

¹ korrekte Diagnose² korrekte Lösung für Empfehlungen

2B.2 Fall 2: Lena / Christian

Tab. 2B.2 Gewählte Antwortoptionen für Fallvignette 2

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	2	1.6	ohne Angabe	2	1.6
normal begabte/r Schüler/in¹	100	78.1	sonderpädagogische Förderung	8	6.3
Lernstörung Lesen / Rechnen	11	8.6	sonderpädagogische Förderung, Förderung des Lesens	6	4.7
kombinierte Störung schulischer Fertigkeiten	6	4.7	sonderpädagogische Förderung, Förderung des Lesens, Förderung des Rechnens	6	4.7
Schüler/in mit Lernbehinderung	3	2.3	sonderpädagogische Förderung, Förderung des Rechnens	1	.8
Hochbegabte/r Hochleister/in	0	.0	Förderung des Lesens	34	26.6
Hochbegabte/r Minderleister/in	0	.0	Förderung des Lesens, Förderung des Rechnens³	9	7.0
mehrere Diagnosen ausgewählt	6	4.7	Förderung des Lesens, keine – nur Regelunterricht	3	2.3
			Förderung des Rechnens	2	1.6
			Teilnahme AG „Club Einstein“	0	.0
			Teilnahme AG „Club Goethe“	0	.0
			keine – nur Regelunterricht²	57	44.5
Gesamt	128	100.0	Gesamt	128	100.0

¹korrekte Diagnose²korrekte Erstlösung für Empfehlungen³korrekte Alternativlösung für Empfehlungen

2B.3 Fall 3: Katrin / Felix

Tab. 2B.3 Gewählte Antwortoptionen für Fallvignette 3

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	3	2.3	ohne Angabe	3	2.3
normal begabte/r Schüler/in¹	81	63.3	sonderpädagogische Förderung	1	.8
Lernstörung Lesen / Rechnen	1	.8	sonderpädagogische Förderung, keine – nur Regelunterricht	1	.8
kombinierte Störung schulischer Fertigkeiten	0	.0	Förderung des Lesens	1	.8
Schüler/in mit Lernbehinderung	0	.0	Förderung des Lesens, Teilnahme AG „Club Einstein“	1	.8
Hochbegabte/r Hochleister/in	38	29.7	Förderung des Rechnens	0	.0
Hochbegabte/r Minderleister/in	0	.0	Teilnahme AG „Club Einstein“³	64	50.0
mehrere Diagnosen ausgewählt	5	3.9	Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“	19	14.8
			Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“, keine – nur Regelunterricht	3	2.3
			Teilnahme AG „Club Einstein“, keine – nur Regelunterricht	3	2.3
			Teilnahme AG „Club Goethe“	2	1.6
			Teilnahme AG „Club Goethe“, keine – nur Regelunterricht	1	.8
			keine – nur Regelunterricht²	29	22.7
Gesamt	128	100.0	Gesamt	128	100.0

¹korrekte Diagnose

²korrekte Erstlösung für Empfehlungen

³korrekte Alternativlösung für Empfehlungen

2B.4 Fall 4: Sophie / Lukas

Tab. 2B.4 Gewählte Antwortoptionen für Fallvignette 4

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	2	1.6	ohne Angabe	2	1.6
normal begabte/r Schüler/in	6	4.7	sonderpädagogische Förderung	1	.8
Lernstörung Lesen / Rechnen	0	.0	sonderpädagogische Förderung, Förderung des Rechnens, Teilnahme AG „Club Goethe“	1	.8
kombinierte Störung schulischer Fertigkeiten	0	.0	sonderpädagogische Förderung, Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“	3	2.3
Schüler/in mit Lernbehinderung	0	.0	Förderung des Lesens	0	.0
Hochbegabte/r Hochleister/in¹	111	86.7	Förderung des Rechnens	0	.0
Hochbegabte/r Minderleister/in	5	3.9	Teilnahme AG „Club Einstein“	5	3.9
mehrere Diagnosen ausgewählt	4	3.2	Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“²	57	44.5
			Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“, keine – nur Regelunterricht	3	2.3
			Teilnahme AG „Club Goethe“	52	40.6
			keine – nur Regelunterricht	4	3.1
Gesamt	128	100.0	Gesamt	128	100.0

¹korrekte Diagnose²korrekte Lösung für Empfehlungen

2B.5 Fall 5: Hanna / Jonas

Tab. 2B.5 Gewählte Antwortoptionen für Fallvignette 5

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	3	2.3	ohne Angabe	3	2.3
normal begabte/r Schüler/in¹	113	88.3	sonderpädagogische Förderung	1	.8
Lernstörung Lesen / Rechnen	2	1.6	sonderpädagogische Förderung, Teilnahme AG „Club Goethe“	2	1.6
kombinierte Störung schulischer Fertigkeiten	0	.0	Förderung des Lesens	0	.0
Schüler/in mit Lernbehinderung	0	.0	Förderung des Lesens, Förderung des Rechnens, keine – nur Regelunterricht	1	.8
Hochbegabte/r Hochleister/in	6	4.7	Förderung des Lesens, Teilnahme AG „Club Einstein“	1	.8
Hochbegabte/r Minderleister/in	2	1.6	Förderung des Rechnens	2	1.6
mehrere Diagnosen ausgewählt	2	1.6	Teilnahme AG „Club Einstein“	2	1.6
			Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“	8	6.3
			Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“, keine – nur Regelunterricht	1	.8
			Teilnahme AG „Club Goethe“	15	11.7
			Teilnahme AG „Club Goethe“, keine – nur Regelunterricht	5	3.9
			keine – nur Regelunterricht²	87	68.0
Gesamt	128	100.0	Gesamt	128	100.0

¹korrekte Diagnose

²korrekte Lösung für Empfehlungen

2B.6 Fall 6: Jana / Luca

Tab. 2B.6 Gewählte Antwortoptionen für Fallvignette 6

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	1	.8	ohne Angabe	3	2.3
normal begabte/r Schüler/in	1	.8	sonderpädagogische Förderung	5	3.9
Lernstörung Lesen / Rechnen	0	.0	sonderpädagogische Förderung, Förderung des Lesens	1	.8
kombinierte Störung schulischer Fertigkeiten	3	2.3	sonderpädagogische Förderung, Förderung des Lesens, Förderung des Rechnens	3	2.3
Schüler/in mit Lernbehinderung	0	.0	sonderpädagogische Förderung, Teilnahme AG „Club Einstein“	1	.8
Hochbegabte/r Hochleister/in	3	2.3	sonderpädagogische Förderung, Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“	1	.8
Hochbegabte/r Minderleister/in¹	116	90.6	Förderung des Lesens	3	2.3
mehrere Diagnosen ausgewählt	4	3.2	Förderung des Lesens, Förderung des Rechnens	2	1.6
			Förderung des Lesens, Teilnahme AG „Club Einstein“	1	.8
			Förderung des Rechnens	0	.0
			Teilnahme AG „Club Einstein“	36	28.1
			Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“²	58	45.3
			Teilnahme AG „Club Einstein“, Teilnahme AG „Club Goethe“, keine – nur Regelunterricht	1	.8
			Teilnahme AG „Club Goethe“	8	6.3
			keine – nur Regelunterricht	5	3.9
Gesamt	128	100.0	Gesamt	128	100.0

¹ korrekte Diagnose

² korrekte Lösung für Empfehlungen

2B.7 Fall 7: Leonie / Jan

Tab. 2B.7 Gewählte Antwortoptionen für Fallvignette 7

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	0	.0	ohne Angabe	1	.8
normal begabte/r Schüler/in	24	18.8	sonderpädagogische Förderung	0	.0
Lernstörung Lesen / Rechnen¹	90	70.3	sonderpädagogische Förderung, Förderung des Lesens	1	.8
kombinierte Störung schulischer Fertigkeiten	1	.8	Förderung des Lesens²	101	78.9
Schüler/in mit Lernbehinderung	0	.0	Förderung des Lesens, Teilnahme AG „Club Einstein“	17	13.3
Hochbegabte/r Hochleister/in	0	.0	Förderung des Lesens, keine – nur Regelunterricht	1	.8
Hochbegabte/r Minderleister/in	5	3.9	Förderung des Rechnens	2	1.6
mehrere Diagnosen ausgewählt	8	6.3	Teilnahme AG „Club Einstein“	0	.0
			Teilnahme AG „Club Goethe“	0	.0
			keine – nur Regelunterricht	5	3.9
Gesamt	128	100.0	Gesamt	128	100.0

¹korrekte Diagnose

²korrekte Lösung für Empfehlungen

2B.8 Fall 8: Laura / Thomas

Tab. 2B.8 Gewählte Antwortoptionen für Fallvignette 8

Diagnosen			Empfehlungen		
Antwortoption	Häufigkeit	Prozent	Antwortoption	Häufigkeit	Prozent
Ohne Angabe	2	1.6	ohne Angabe	2	1.6
normal begabte/r Schüler/in	0	.0	sonderpädagogische Förderung³	44	34.4
Lernstörung Lesen / Rechnen	9	7.0	sonderpädagogische Förderung, Förderung des Lesens	24	18.8
kombinierte Störung schulischer Fertigkeiten	36	28.1	sonderpädagogische Förderung, Förderung des Lesens, Förderung des Rechnens²	33	25.8
Schüler/in mit Lernbehinderung¹	70	54.7	sonderpädagogische Förderung, Förderung des Rechnens	1	.8
Hochbegabte/r Hochleister/in	0	.0	sonderpädagogische Förderung, keine – nur Regelunterricht	1	.8
Hochbegabte/r Minderleister/in	0	.0	Förderung des Lesens	10	7.8
mehrere Diagnosen ausgewählt	11	8.6	Förderung des Lesens, Förderung des Rechnens	13	10.2
			Förderung des Rechnens	0	.0
			Teilnahme AG „Club Einstein“	0	.0
			Teilnahme AG „Club Goethe“	0	.0
			keine – nur Regelunterricht	0	.0
Gesamt	128	100.0	Gesamt	128	100.0

¹korrekte Diagnose

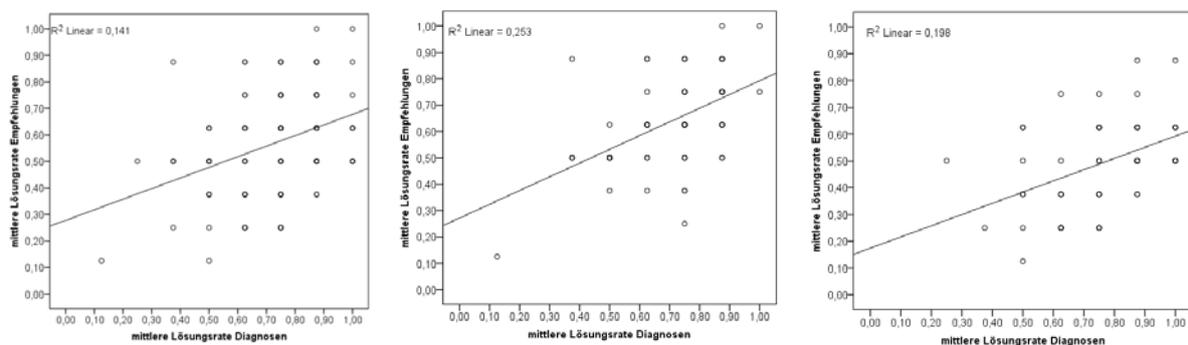
²korrekte Erstlösung für Empfehlungen

³korrekte Alternativlösung für Empfehlungen

Anhang 2C. Scatterplots zu Korrelationsanalysen in Studie 2

2C.1 Zusammenhang der Lösungsraten von Diagnosen und Empfehlungen

Die Scatterplots in Abb. 2C.1 zeigen, dass der Zusammenhang zwischen korrekten Diagnosen und Empfehlungen bedingungsübergreifend nur mittelgroß war. Bei der Zusammenhangsanalyse getrennt für beide Reihenfolgebedingungen ergaben sich etwas stärkere Zusammenhänge zwischen den Lösungsraten beider Bedingungen, die allerdings jeweils denselben Trend kennzeichneten wie die bedingungsübergreifende Zusammenhangsanalyse (vgl. Abb. 2C.1a sowie Abschnitt 7.3.2.2): In der normativen Bedingung (Empfehlung nach Diagnose) ergab sich ein mittelgroßer Zusammenhang zwischen den Lösungsraten, $r = .45$, $p = .001$, $N = 62$. In der nicht-normativen Bedingung (Empfehlung vor Diagnose) war dieser Zusammenhang minimal stärker: $r = .50$, $p < .001$, $N = 55$.



(a) gemittelt über beide Reihenfolgebedingungen

(b) nicht-normative Reihenfolgebedingung

(c) normative Reihenfolgebedingung

Abb. 2C.1 Zusammenhang der Lösungsraten von Diagnosen und Empfehlungen

2C.3 Zusammenhang des Nutzungsanteils formeller Diagnoseinformationen und Lösungsraten

Die nachfolgenden Scatterplots wurden in Ergänzung zur Korrelationsanalyse des Nutzungsanteils formeller Diagnoseinformationen und der Lösungsrate für Diagnosen und Empfehlungen unter Abschnitt 7.3.2.2 erstellt.

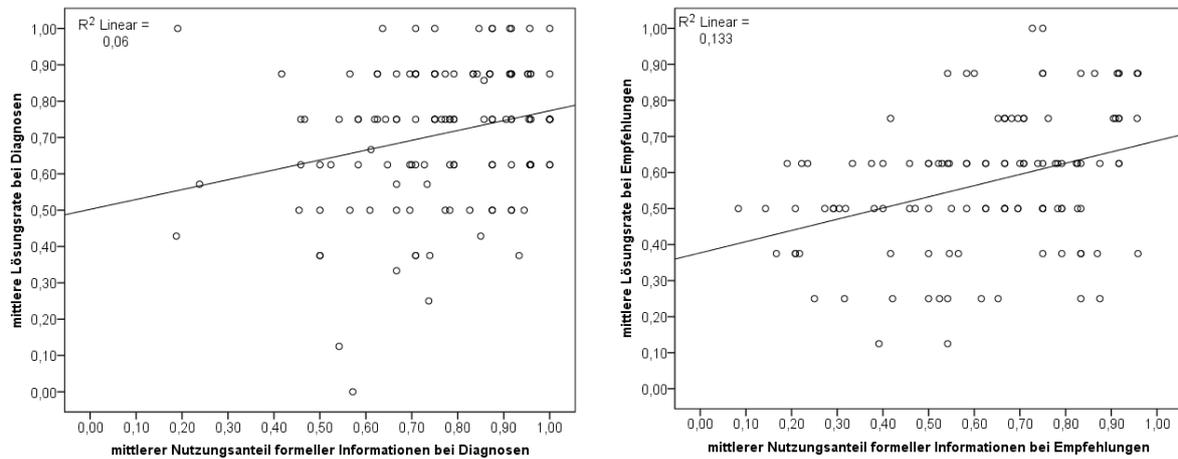


Abb. 2C.3 Zusammenhang vom Nutzungsanteil formeller Informationen und Lösungsrate für Diagnosen und Empfehlungen

Anhang 3.A Konzeptualisierung des Wissenstests in Studie 3

Tabelle 3.A Zweidimensionaler Wissenstest zur Erfassung von Grundlagen- und Anwendungswissen

Subskala (Anzahl Items)	Kurzbeschreibung der Inhalte (Anzahl Items)
Diagnostisches Grundlagenwissen (29)	<p>Themenbereich: Kognitive Lernvoraussetzungen / Intelligenz¹ (14)</p> <p>Interpretation von IQ-Werten in Bezug auf die Standardnormalverteilung und Eigenschaften der IQ-Skala</p> <p>Befundmuster zu intelligenzbezogenen Lernbesonderheiten zuordnen²</p> <p>Diagnosekriterien und notwendige diagnostische Informationen zur Feststellung von Hochbegabung, Underachievement und allgemeiner Lernschwäche („Lernbehinderung“)</p> <p>Prinzip symptomorientierter Förderung</p> <p>Themenbereich: Lernstörungen (8)</p> <p>Befundmuster zu Lernstörungen zuordnen²</p> <p>Diagnosekriterien und notwendige diagnostische Informationen zur Feststellung von Lese-Rechtschreibstörung, kombinierter Störung schulischer Fertigkeiten und allgemeiner Lernschwäche</p> <p>Prinzip symptomorientierter Förderung</p> <p>Themenbereich: Standardisierte Tests (7)</p> <p>Interpretation von T-Werten und z-Werten in Bezug auf die Standardnormalverteilung</p> <p>Umrechnung eines z-Wertes in einen T-Wert</p> <p>Interpretation von Prozenträngen</p>
Diagnostisches Anwendungswissen – Szenariotest ³ (12)	<p>Vergabe von Diagnosevermutungen für Fallbeschreibungen (6)</p> <p><i>Hochbegabter Minderleister</i> mit ausgeprägten Interessen für sozialpolitische Themen, aber schlechten Noten in fast allen Fächern</p> <p>Schüler mit <i>Lese-Rechtschreibschwäche</i>, der umschriebene Schwierigkeiten im Bereich der Lesefähigkeiten und der Rechtschreibung zeigt</p> <p>Schülerin mit <i>Rechenschwäche</i>, die in allen Fächern gute bis sehr gute Leistungen zeigt, außer in Mathematik</p> <p>Schüler mit <i>allgemeiner Lernschwäche</i>, der in fast allen Fächern deutliche Entwicklungsverzögerungen zeigt und sich Inhalte schwer einprägen kann</p> <p>Schülerin mit <i>Hochbegabung</i> und entsprechend sehr guten Schulleistungen</p> <p>Schülerin mit einer <i>kombinierten Störung schulischer Fertigkeiten</i>, die im Lesen, Rechtschreiben und Rechnen deutliche Defizite zeigt, deren kognitive Fähigkeiten aber im durchschnittlichen Bereich liegen</p> <p>Abgabe von Förderempfehlungen für Fallbeschreibungen (6)</p> <p>(gleiche Fälle wie oben)</p>

Anmerkungen: ¹Die ersten vier Items zu diesem Themenbereich könnten alternativ auch dem Inhalt „standardisierte Tests“ zugeordnet. Damit würde sich folgende Item-Anzahl pro Inhaltsbereich ergeben: Kognitive Lernvoraussetzungen / Intelligenz: 10 Items, Lernstörungen: 8 Items, standardisierte Tests: 11 Items. ²Die Items hierzu wurden aufgrund geringer Trennschärfen aus dem Test ausgeschlossen. ³Die ursprüngliche Konzeptualisierung des Wissenstests sah vor, dass die Diagnose- und Empfehlungsentscheidungen im Szenariotest einer gemeinsamen Skala zugeordnet werden, die prozedurales Wissen erfasst.

Anhang 3.B Itembeschreibungen und –statistiken für Studie 3

3.B.1 Evaluationsebene 1: Rezeption

Einführungstext: Wie haben Sie die Übung in Hinsicht auf die untenstehenden Kriterien empfunden? Es gibt hier keine richtigen oder falschen Antworten. Es geht um Ihre persönliche Meinung oder Einschätzung. Bitte geben Sie den Grad Ihrer Zustimmung zu den einzelnen Aussagen an (1 = ich stimme gar nicht zu, 6 = ich stimme voll zu).

Tab. 3.B.1 Kognitive und affektive Bewertung der Übungen

Variable	Itemkürzel	Item	Lerngelegenheit	
			Aufgaben	Fallinventar
Kognitive Bewertung für Klausur	Rel_K	Die Inhalte der Übung sind relevant für die Klausur.	5.41 (.83)	5.34 (.90)
	Nutz_K	Die Übung ist nützlich zur Klausurvorbereitung.	5.39 (.85)	5.24 (.91)
Kognitive Bewertung für Praxis	Rel_P	Die Inhalte der Übung sind relevant für die spätere Praxis.	4.78 (1.15)	5.24 (1.04)
	Nutz_P	Die Übung ist nützlich zur Vorbereitung auf die spätere Praxis.	4.78 (1.12)	5.03 (1.09)
Affektive Bewertung	Int	Die Inhalte der Übung finde ich persönlich interessant.	4.19 (1.05)	4.35 (.99)
	Freu	Die Bearbeitung der Übung hat mir persönlich Freude bereitet.	4.94 (1.17)	4.85 (1.10)

Anmerkung: Angegeben sind die Mittelwerte (Standardabweichungen) der Items.

3.B.2 Evaluationsebene 2: Lernen

Indikatoren für Lernzuwachs

Einführungstext: Wie haben Sie die Übung in Hinsicht auf die untenstehenden Kriterien empfunden? Es gibt hier keine richtigen oder falschen Antworten. Es geht um Ihre persönliche Meinung oder Einschätzung. Bitte geben Sie den Grad Ihrer Zustimmung zu den einzelnen Aussagen an (1 = ich stimme gar nicht zu, 6 = ich stimme voll zu).

Tab. 3.B.2.1 Indikatoren für Lernzuwachs

Variable	Itemkürzel	Item	Lerngelegenheit	
			Aufgaben	Fallinventar
Subjektiver Lernzuwachs	Verb_1	Ich habe durch die Übung mein Wissen vertiefen können.	4.46 (1.22)	4.96 (1.09)
	Verb_2	Ich habe durch die Übung mein Wissen erweitern können	4.07 (1.26)	4.56 (1.22)
	Verb_3	Ich habe durch die Übung meine Beurteilungskompetenz verbessert.	4.14 (.98)	4.43 (.94)
Kompetenzerleben aktuell	Kom_a	Die Bearbeitung der Übung ist mir leicht gefallen.	4.88 (1.19)	4.77 (1.18)
Kompetenzerleben prognostisch.	Kom_b	Die Inhalte der Übung werde ich in der Zukunft gut umsetzen können.	4.43 (1.19)	4.46 (1.07)

Anmerkung: Angegeben sind die Mittelwerte (Standardabweichungen) der Items.

Diagnostikbezogene Selbsteinschätzungen

Einführungstext: Wie haben Sie die Übung in Hinsicht auf die untenstehenden Kriterien empfunden? Es gibt hier keine richtigen oder falschen Antworten. Es geht um Ihre persönliche Meinung oder Einschätzung. Bitte geben Sie den Grad Ihrer Zustimmung zu den einzelnen Aussagen an (1 = ich stimme gar nicht zu, 6 = ich stimme voll zu).

Tab. 3.B.2.2 Diagnostische Kompetenz (Skala)

Item-kürzel	Item	M	SD	Trennschärfe r_{it}	Alpha bei Elision
DK_1	Verschiedene Schüler*innen einer Klasse in Bezug auf Ihre Gesamtleistung in eine Rangreihe bringen	4.14	.96	.49	.82
DK_2	Den Leistungsfortschritt von Schüler*innen mit unterschiedlichen Methoden feststellen.	3.89	1.01	.59	.80
DK_3	Das durchschnittliche Leistungsniveau einer Klasse im Vergleich zu anderen Klassen korrekt einschätzen	4.02	.94	.58	.81
DK_4	Die mündlichen Leistungen von Schüler*innen gerecht beurteilen	3.83	1.08	.40	.83
DK_5	Die Leistungsvariation in einer Klasse im Vergleich zu anderen Klassen korrekt einschätzen	3.74	.91	.56	.81
DK_6	Den Unterricht an das individuelle Leistungsniveau der Schüler*innen anpassen	3.94	.96	.53	.81
DK_7	Den Schwierigkeitsgrad einzelner Aufgaben abschätzen	4.02	.97	.44	.82
DK_8	Das Leistungsniveau einzelner Schüler*innen im Vergleich zur Altersgruppe einschätzen ¹	3.74	.99	.61	.80
DK_9	Einschätzen, ob die Schüler*innen bestimmte Lernziele erreicht haben ¹	4.38	.90	.62	.80

Anmerkungen: Diese Skala ist eine Adaption der Skala „Diagnostische Kompetenz“ aus dem BilWiss-Projekt (Kunter et al., 2015). ¹ Diese Items wurden der ursprünglichen Skala hinzugefügt.

Tab. 3.B.2.3 Diagnostische Expertise (Skala)

Item- kürzel	Item	M	SD	Trenn- schärfe r_{it}	Alpha bei Elision
DE_5	Ein geeignetes Diagnoseverfahren für den gegebenen Anlass auswählen	3.61	1.05	.54	.75
DE_6	Im Unterricht einen Test entwickeln, ihn einsetzen, auswerten und die Ergebnisse zurückmelden	4.20	1.12	.50	.76
DE_7	Einschätzen, ob Testleistungen zuverlässigere Messungen ergeben als subjektive Eindrücke der Lehrkraft	4.06	1.01	.57	.75
DE_8	Diagnostische Informationen gezielt zur individuellen Förderung von Schüler*innen und zur Planung des eigenen Unterrichts einsetzen	4.03	2.55	.28	.86
DE_9	Diagnostische Informationen zur Unterrichts- und Schulentwicklung gezielt einsetzen	3.66	.96	.66	.74
DE_10	Unterschiedliche diagnostische Informationen zu einer zusammenfassenden Beurteilung entwickeln	3.70	1.07	.66	.74
DE_11	Aus einer Diagnose Einschätzungen über die zukünftige Entwicklung (Prognosen) abgeben	3.38	1.05	.58	.75
DE_12	Diagnoseergebnisse verständlich und konstruktiv an die Schüler*innen oder Eltern vermitteln	4.08	1.07	.48	.76
DE_13	Entsprechend der Diagnose geeignete Fördermaßnahmen erarbeiten	4.06	1.03	.60	.75

Anmerkungen: Diese Skala ist eine Eigenentwicklung nach Helmke (2009). Das Item „DE_8“ wurde trotz relativ ungünstiger Itemcharakteristika bei der Analyse der Interventionseffekte beibehalten, da es inhaltlich relevant ist.

Diagnostikbezogene Einstellungen

Einführungstext: Wie haben Sie die Übung in Hinsicht auf die untenstehenden Kriterien empfunden? Es gibt hier keine richtigen oder falschen Antworten. Es geht um Ihre persönliche Meinung oder Einschätzung. Bitte geben Sie den Grad Ihrer Zustimmung zu den einzelnen Aussagen an (1 = ich stimme gar nicht zu, 6 = ich stimme voll zu).

Tab. 3.B.2.4 Einstellung zu Diagnostik (Skala)

Itemkürzel	Item	M	SD	Trennschärfe r_{it}	Alpha bei Elision
ED_allg_1	Diagnostik ist einer der wichtigsten Bestandteile der täglichen Arbeit einer Lehrkraft.	4.18	1.18	.53	.69
ED_allg_3r	Für die Unterrichtsgestaltung kann man meist auf Diagnostik verzichten.	4.69	1.12	.51	.69
ED_allg_4	Ich interessiere mich für die Methoden der pädagogischen Diagnostik.	4.31	1.20	.50	.69
ED_allg_5r	Diagnostik zu betreiben bereitet mir wenig Freude.	4.07	1.28	.46	.70
ED_allg_6	Lehrkräfte sollten gezielt Diagnostik betreiben, um ihre eigenen Eindrücke abzusichern.	4.79	1.97	.23	.78
ED_allg_7	Diagnostik zu betreiben ist unverzichtbar, um Fördermöglichkeiten für Schülerinnen und Schüler zu erarbeiten.	5.04	1.04	.58	.68
ED_allg_8r	Diagnostische Informationen können nur wenige Fragestellungen beantworten.	4.33	1.10	.42	.71
ED_allg_9r	Ich kann mich auf meine subjektiven Urteile gut verlassen und brauche daher keine spezielle Diagnostik.	4.91	1.03	.43	.71

Anmerkungen: Items, die mit „r“ enden, waren negativ gepolte Items und wurden vor den Berechnungen rekodiert.

Tab. 3.B.2.5 Einstellung zu standardisierten Tests (Skala)

Item	Ausformuliertes Item	M	SD	Trennschärfe r_{it}	Alpha bei Elision
ED_ST_1	Standardisierte Verfahren sollte man in der Schule unbedingt einsetzen, da sie sehr objektive Informationen liefern.	4.09	1.12	.67	.78
ED_ST_3r	Über die Kompetenzen von Schülerinnen und Schülern kann man durch standardisierte Verfahren nicht viel erfahren.	4.21	1.26	.56	.81
ED_ST_4	Standardisierte Verfahren sind in der Schulpraxis unverzichtbar, da sie sehr aussagekräftige Informationen bereitstellen.	3.88	1.12	.67	.78
ED_ST_5r	Die Verwendung standardisierter Tests ist im Schulalltag unnötig.	4.73	1.22	.59	.80
ED_ST_6	Alle Lehrkräfte sollten, wenn immer möglich, auf standardisierte Verfahren zurückgreifen.	3.33	1.27	.58	.80
ED_ST_7r	Standardisierte Verfahren können den Leistungsstand einzelner Schüler*innen nur ungenügend abbilden.	3.62	1.30	.51	.82

Anmerkungen: Items, die mit „r“ enden, waren negativ gepolte Items und wurden vor den Berechnungen rekodiert.

Diagnostisches Wissen

Tab. 3.B.2.6 Deklaratives Wissen

Itemkürzel	Itemkurzbeschreibung	Antwortoptionen (Format)	M	SD	Trennschärfe r_{it}	Alpha bei Elision
<i>Themenbereich: Kognitive Lernvoraussetzungen / Intelligenz</i>						
Int_3	Interpretation IQ-Wert (IQ = 83)	5 (MC)	.71	.45	.45	.81
Int_5	Mittelwert IQ-Skala	5 (MC)	.89	.32	.37	.81
Int_6	Standardabweichung IQ-Skala	5 (MC)	.77	.42	.46	.81
Int_18	Notwendige Tests zur Diagnostik einer Hochbegabung	9 (MA)	.36	.48	.22	.82
Int_7_1	IQ-Kriterium Hochbegabung	Keine (HO)	.71	.45	.38	.81
Int_7_2	Diskrepanzkriterium hochbegabte/r Minderleister/in	Keine (HO)	.64	.48	.39	.81
Int_10_Z	Diagnosekriterien Lernbehinderung	6 (MA)	.44	.50	.28	.82
Int_15_Z	Indikation zur Förderung kognitiver Fähigkeiten	2 (WF)	.56	.50	.15	.82
Int_4_Z	Interpretation IQ-Wert (IQ = 111)	5 (MC)	.83	.38	.40	.81
Int_12_Z	Notwendige Tests zur Diagnostik einer Lernbehinderung	5 (MA)	.57	.50	.36	.81
<i>Themenbereich: Lernstörungen</i>						
Lern_6	Prinzip der symptomorientierten Förderung bei Lernstörungen	2 (WF)	.74	.44	.21	.82
Lern_3	Notwendige Tests zur Diagnostik einer Lese-Rechtschreibschwäche	6 (MA)	.53	.50	.35	.81
Lern_11	Notwendige Tests zur Diagnostik einer kombinierten Störung schulischer Fertigkeiten	5 (MA)	.75	.44	.40	.81
Lern_10	Notwendige Tests zum Ausschluss einer allgemeinen Lernschwäche / Lernbehinderung	5 (MA)	.64	.48	.43	.81
<i>Themenbereich: Standardisierte Tests</i>						
Test_3	Interpretation Standardnormwert (T = 32)	5 (MC)	.71	.46	.35	.81
Test_5	Interpretation z-Wert (z = -0.1)	5 (MC)	.62	.49	.53	.80
Test_7	Interpretation Standardnormwert (T = 78)	5 (MC)	.61	.50	.42	.81
Test_9	Umrechnung z-Wert in T-Wert	2 (WF)	.57	.50	.49	.81
Test_4	Interpretation Prozentrang (PR = 64)	5 (MC)	.71	.45	.30	.8
Test_2_Z	Interpretation z-Wert (z = 1.3)	5 (MC)	.74	.44	.30	.81
Test_11_Z	Interpretation z-Wert (z = 0)	2 (WF)	.64	.48	.44	.81
<i>Diagnosevermutungen im Szenariotest</i>						
Fall1_D	Hochbegabter Minderleister	5 (MC)	.93	.25	.26	.82
Fall3_D	Schüler mit Lese-Rechtschreibschwäche	5 (MC)	.78	.42	.32	.81

Tabelle 3.B.2.6 Fortsetzung

Fall4_D	Schülerin mit Rechenschwäche	5 (MC)	.40	.48	.32	.81
Fall5_D	Schüler mit allgemeiner Lernschwäche / Lernbehinderung	5 (MC)	.84	.36	.24	.82
Fall6_D_Z	Schülerin mit Hochbegabung ohne Minderleistung	Keine (O)	.34	.48	.27	.82
Fall2_D_Z	Schülerin mit kombinierter Störung schulischer Fertigkeiten	5 (MC)	.59	.49	.23	.82

Anmerkungen: Die Aufgaben entstammen überwiegend dem Grundlagenteil des Wissenstests. Die Diagnosevermutungen des Szenariotests wurden entsprechend der Validitätsuntersuchungen ebenfalls dieser Subskala deklarativen Wissens zugeordnet. Abkürzungen: MC = Multiple Choice, MA = Multiple Answer, HO = Halboffenes Antwortformat, WF = Wahr-Falsch.

Tab. 3.B.2.7 Prozedurales Wissen

Item	Itemkurzbeschreibung	Antwortoptionen (Format)	M	SD	Trennschärfe r_{it}	Alpha bei Elision
Fall1_E	Hochbegabter Minderleister	Keine (O)	.46	.25	.36	.68
Fall3_E	Schüler mit Lese-Rechtschreibschwäche	Keine (O)	.40	.23	.38	.67
Fall4_E	Schülerin mit Rechenschwäche	Keine (O)	.47	.28	.55	.61
Fall5_E	Schüler mit allgemeiner Lernschwäche (Lernbehinderung)	Keine (O)	.42	.25	.51	.63
Fall6_E	Schülerin mit Hochbegabung ohne Minderleistung	Keine (O)	.38	.31	.32	.70
Fall2_E	Schülerin mit kombinierter Störung schulischer Fertigkeiten	Keine (O)	.53	.24	.47	.64

Anmerkungen: Die Aufgaben dieser Skala enthalten die Empfehlungen aus dem Szenariotest. Abkürzungen: O = offenes Antwortformat

Anhang 3.C Voranalysen zu potentiellen Störvariablen für Studie 3

Tabelle 3.C Randsummen und Prüfstatistiken zu potentiellen Störvariablen

Variablen Stufen	Interventionsgruppe			N gesamt	Prüfstatistik		
	Aufgaben	Fallinventar	kombiniert		$\chi^2_{Pearson}$	df	p
Geschlecht							
Gesamt	103	74	79	256	5.86	4	.21
Weiblich	73	46	57	176			
Männlich	29	23	20	72			
Ohne Zuordnung	1	5	2	8			
Alter							
Gesamt	103	74	79	256	4.08	4	.40
20 - 25 Jahre	62	39	44	145			
26 - 30 Jahre	16	14	20	50			
> 30 Jahre	25	21	15	61			
Studienphase							
Gesamt	105	75	79	256	2.58	2	.28
Bachelorstudium	8	11	7	26			
Masterstudium	97	64	72	233			
Studiengang							
Gesamt	106	76	78	260	2.53	2	.28
Grundschulpäd.	46	25	27	98			
ISS Gymnasium	60	51	51	162			
Erfahrungen in Diagnostik							
Gesamt	76	105	79	260	2.06	2	.36
Mit Erfahrung	4	11	9	24			
Ohne Erfahrung	72	94	70	236			
Erfahrungen in der Praxis							
Gesamt	76	105	79	260	.98	2	.61
Mit Erfahrung	44	53	42	139			
Ohne Erfahrung	32	52	37	121			
Seminarleiterin							
Gesamt	76	106	79	261	7.47	6	.28
Seminarleiterin 1	16	24	19	59			
Seminarleiterin 2	40	44	26	110			
Seminarleiterin 3	11	21	16	48			
Seminarleiterin 4	9	17	18	44			

Anmerkungen. Zur Bestimmung signifikanter Ungleichverteilungen hinsichtlich der Prüfvariablen wurde die Prüfgröße χ^2 nach Pearson berechnet. Die Berechnung der p-Werte erfolgte asymptotisch.

Anhang 3.D Faktorenanalyse des zweidimensionalen Wissenstests in Studie 3

Tabelle 3.D Faktorladungen der Items aus dem Wissenstest

Items	Unrotierte Lösung ¹		Rotierte Lösung ²	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2
Int_3	.52	-.30	.60	.03
Int_5	.44	-.28	.52	.00
Int_6	.57	-.19	.58	.15
Int_18	.26	.037	.20	.17
Int_7_1	.46	-.16	.47	.11
Int_7_2	.45	-.20	.49	.08
Int_10	.33	.15	.20	.36
Int_15	.18	.04	.13	.13
Int_4	.45	-.20	.49	.07
Int_12	.45	-.04	.40	.21
Lern_6	.22	-.17	.28	-.03
Lern_3	.43	-.07	.40	.17
Lern_11	.43	-.07	.40	.18
Lern_10	.53	-.05	.47	.24
Test_3	.42	-.27	.50	-.00
Test_5	.61	-.17	.60	.19
Test_9	.57	-.17	.57	.17
Test_7	.47	-.17	.49	.11
Test_4	.33	.02	.26	.20
Test_11	.36	.09	.26	.27
Fall1_D	.36	.06	.27	.24
Fall1_E	.27	.46	-.02	.53
Fall3_D	.40	-.04	.36	.19
Fall3_E	.41	.52	.07	.66
Fall4_D	.37	-.19	.41	.04
Fall4_E	.39	.35	.14	.51
Fall5_D	.31	.14	.19	.28
Fall5_E	.41	.62	.01	.74
Fall6_D	.30	.00	.25	.17
Fall6_E	.28	.41	.01	.50
Fall2_D	.28	.14	.16	.27
Fall2_E	.39	.54	.04	.67

Anmerkungen: Extraktionsmethode Hauptkomponentenanalyse; fettgedruckte Werte zeigen die Zuordnung zu den Faktoren an. ¹ Die nichtrotierte Lösung ergab eindeutiger Item-Faktoren-Zuordnungen als die rotierte Faktorenlösung und wurde daher bevorzugt sowie zur Faktorenbenennung herangezogen. ² Rotationsmethode: Varimax mit Kaiser-Normalisierung, Rotation ist in 3 Iterationen konvergiert. Abkürzungen: MC = Multiple Choice, MA = Multiple Answer, HO = Halboffenes Antwortformat, WF = Wahr-Falsch. Zur Bedeutung der Itemkürzel vgl. Tab. 3.B.2.6 und 3.B.2.7.

Anhang 3.E Einschätzungen der Lerngelegenheiten durch die Dozierenden

Tabelle 3.E Bewertung der Übungen durch die Seminarleiter*innen

Indikator	Items	SL 1		SL 2		SL 3		SL 4		Alle SL	
		AÜ	FI	AÜ	FI	AÜ	FI	AÜ	FI	AÜ	FI
Bewertung	Relevanz	6	6	6	5	6	5	6	5	6	5
kognitiv –	Nutzen	6	6	6	5	6	5	6	3	6	5
Klausur	<i>Mittelwert</i>	6	6	6	5	6	5	6	4	6	5
Bewertung	Relevanz	4	5	5	5	5	6	5	6	5	6
kognitiv –	Nutzen	4	5	4	5	5	6	6	6	5	6
Praxis	<i>Mittelwert</i>	4	5	5	5	5	6	6	6	5	6
Bewertung	Freude	4	6	3	4	3	5	5	6	4	5
affektiv	Nutzen	2	5	4	5	4	6	6	5	4	5
	<i>Mittelwert</i>	3	6	4	5	4	6	6	6	4	5
Empfundener	Wissen vertiefen	1	4	5	4	3	4	2	6	3	5
Lernzuwachs	Wissen erweitern	1	5	5	4	3	4	4	6	3	4
	Beurteilungskompetenz	2	2	2	3	3	4	2	3	2	3
	<i>Mittelwert</i>	1	4	4	4	3	4	3	5	3	4
Kompetenz-	aktuell	3	5	5	5	5	5	5	3	5	5
erleben bei	Prognostisch	4	5	6	6	5	6	3	4	5	5
Übung	<i>Mittelwert</i>	4	5	6	6	5	6	4	4	5	5

Anmerkungen: SL = Seminarleitung, AÜ = aufgabenbasierte Übung, FÜ = fallbasierte Übung;

für alle Maße stand eine sechsstufige Likert-Skala) zur Verfügung, mittels derer die Dozierenden ihre Zustimmung zu den einzelnen Items angeben konnten (1 = stimme gar nicht zu, 6 = stimme voll zu). Die Mittelwerte pro Seminarleiter*in wurden .5 aufgerundet. Die Spalte „alle SL“ gibt den Mittelwert über alle Seminarleiter*innen an.

Anhang 3.F Deskriptive Ergebnisse der Studierenden ohne Übung

Tabelle 3.F Deskriptiver Vergleich der Studierenden mit und ohne Intervention

	Gruppe „Aufgaben“ (N = 75)	Gruppe „Fallinventar“ (N = 105)	Kombinierte Gruppe (N = 79)	Ohne Übung (N = 11)
Ebene 2 (Lernen): Kognitive Voraussetzungen				
Wissenstest – deklarativ	.69 (.19)	.58 (.21)	.73 (.17)	.48 (.26)
Wissenstest - Anwendung	.40 (.14)	.47 (.16)	.49 (.13)	.31 (.27)
Wissenstest – Diagnosen	.66 (.22)	.60 (.19)	.69 (.20)	.53 (.20)
Ebene 2 (Lernen): Affektiv-motivationale Voraussetzungen				
Diagnostische Kompetenz	3.89 (.59)	3.92 (.59)	4.12 (.74)	3.94 (.41)
Diagnostische Expertise	3.80 (.86)	3.81 (.79)	4.05 (.73)	4.01 (.58)
Einstellung Diagnostik	4.42 (.85)	4.65 (.68)	4.59 (.70)	4.54 (.58)
Einstellung Tests	3.92 (.85)	4.06 (.98)	4.00 (.84)	4.03 (.66)

Anmerkungen: Bei den kognitiven Voraussetzungen sind mittlere Lösungsraten (SD) angegeben (Wertebereich 0 - 1). Für die affektiv-motivationalen Variablen sind die Skalenmittelwerte (SD) angegeben (Wertebereich 1 - 6).

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die Dissertation mit dem Titel

„Ein diagnostisches Fallinventar: Entwicklung, Erprobung und Evaluation einer fallbasierten Lerngelegenheit zur Förderung von diagnostischen Kompetenzen im Lehramtsstudium“

selbstständig verfasst habe. Sämtliche Hilfsmittel, die ich verwendet habe, sind angegeben.

Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Berlin, 30.06.2020
