

Entwicklung diagnostischer Kompetenz angehender Lehrkräfte mit problemorientiertem Lernen und Textfällen

vorgelegt von
Dipl.-Päd.
Alexander Wedel
ORCID: 0000-0002-6909-0863

An der Fakultät I – Geistes- und Bildungswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Philosophie
– Dr. phil. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Johannes Meyser
Gutachterin: Prof.in Dr.in Angela Ittel
Gutachter: PD Dr. Jan Pfetsch
Gutachterin: Prof.in Dr.in Felicitas Thiel

Tag der wissenschaftlichen Aussprache: 15. Juni 2020

Berlin 2020

DANKSAGUNG

Eine Krux wissenschaftlichen Arbeitens ist die Zuschreibung von Leistung. Die Illusion von Autonomie und Kontrolle im eigenen Wirken ist im Wissenschaftssystem eng an die institutionalisierte Anerkennung einzelner Personen gebunden. In der Praxis ist wissenschaftliches Arbeiten aber ein Beziehungsgeschäft. Es geht nicht ohne kreativen Austausch, soziale Unterstützung und Zusammenarbeit. Wissenschaftliche Leistungen sind immer das Produkt mehrerer Personen, und seien diese historisch vorgängig. Das Produkt eines solchen gemeinsamen Arbeitsprozesses in dieser Dissertation zusammenzustellen, ist ein Privileg, dem ich allermindestens mit einer Danksagung begegnen kann.

Zuerst sei den Gutachtenden dieser Arbeit gedankt. Frau Prof.in Dr.in Angela Ittel, Herr Prof. Dr. Jan Pfetsch und Frau Prof.in Dr.in Felicitas Thiel waren stets zu einem interessierten Austausch bereit und haben dadurch eine gute Arbeitsatmosphäre hergestellt.

Bei Prof.in Dr.in Theresa Lempp möchte ich mich für die langjährige Wegbegleitung bedanken, die Offenheit, das Vertrauen und die gemeinsame Entwicklung von Gedanken. Es ist ein großes Glück, wenn Zusammenarbeit eine solche Form annehmen kann.

Dasselbe gilt für Dr.in Christin R. Müller, an deren Wissen und Können ich während der Dissertationszeit teilhaben durfte und die eine nicht ermüdende soziale Unterstützung leistete.

Ebenso möchte ich mich bei Prof. Dr. Lothar Böhnisch, Dr.in Simone Menz, Prof. Dr. Gerd Stecklina, Prof. Dr. Andreas Hanses und dem IRIS-Team für Ihre Rolle in meiner anfänglichen und weiteren wissenschaftlichen Sozialisation bedanken, was insbesondere eine tiefgehende theoretische Auseinandersetzung umfasst.

Weiterhin möchte ich mich bei Dominic Bonfert und Ellen Jande bedanken, die mich in der alltäglichen Projektarbeit unterstützt haben.

Außerdem bedanke ich mich bei Franziska Greiner für die schöne Zusammenarbeit während der Dissertationszeit und den Austausch über Erfolge, Misserfolge und andere Themen, die in jeder Dissertation auftauchen.

INHALT

Tabellen	iv
Abbildungen.....	v
Theoretischer Teil	6
1 Einleitung.....	6
2 Kompetenz als theoretisches Konstrukt	10
2.1 Generalistische, normative und funktionalistische Kompetenzbegriffe	11
2.2 Kognitive und psychometrische Modellierung von Kompetenz	18
2.3 Kritik am Kompetenzparadigma.....	23
2.4 Zusammenfassung des zweiten Kapitels	25
3 Diagnostische Kompetenz im professionellen Handeln von Lehrkräften	27
3.1 Definition und Modellierung diagnostischer Kompetenz	30
3.2 Erweiterung diagnostischer Kompetenz um kooperatives Handeln	50
3.3 Zusammenfassung und Definition diagnostischer Kompetenz	56
4 Entwicklung diagnostischer Kompetenz mit problemorientiertem Lernen	58
4.1 Diagnostische Fälle aus problemtheoretischer Perspektive	61
4.2 Problemorientiertes Lernen als Methode der Fallarbeit	64
4.3 Die Intervention POL-TF.....	65
Methodischer Teil	70
5 Methoden: Veränderungsmessung in kleinen Stichproben	70
5.1 Replikation und manifeste meta-analytische Regressionsmodelle.....	74
5.2 Manifeste Modellierung von Zusammenhangsveränderungen	80
6 Ziele, Fragestellungen und Hypothesen	84
Empirischer Teil	87
7 Studie 1: Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen	87
7.1 Einleitung und theoretischer Hintergrund.....	87
7.2 Fragestellung und Hypothesen	92
7.3 Methoden und Design	93
7.4 Ergebnisse	96
7.5 Diskussion	100

8	Study 2: Training Teachers' Diagnostic Competence with Problem-Based Learning: A Pilot and Replication Study	105
8.1	Introduction and Current State of Research	105
8.2	Research Questions and Hypotheses	111
8.3	Methods and Design	112
8.4	Results	116
8.5	Discussion	123
9	Studie 3: Evidenzbasierung und Zusammenhangsanalysen in der Lehrkräftebildung – Diagnosewissen, Selbstkonzept und Einstellungen zu Inklusion in manifesten Wachstumsmodellen	129
9.1	Einleitung und theoretischer Hintergrund	129
9.2	Fragestellung und Hypothesen	132
9.3	Methoden	133
9.4	Ergebnisse	137
9.5	Diskussion	141
10	Studie 4: Welchen Effekt haben Intelligenz, Empathie, Textverstehen und Textmerkmale auf das spontane diagnostische Urteil von Lehramtsstudierenden?	146
10.1	Einleitung und theoretischer Hintergrund	146
10.2	Fragestellung und Hypothesen	150
10.3	Methoden und Design	150
10.4	Ergebnisse	152
10.5	Diskussion	155
	Diskussionsteil	159
11	Gesamtdiskussion und Ausblick	159
11.1	Zusammenfassung zentraler Befunde	161
11.2	Einordnung der Ergebnisse in Interventionsforschung und Theorie zu diagnostischer Kompetenz in der Lehrkräftebildung	164
11.3	Grenzen und Stärken der Forschungsarbeit	172
11.4	Anschließende Fragestellungen und zukünftige Forschung	176
11.5	Implikationen für die Praxis der Lehrkräftebildung	177
	Literatur	179
	Anhang	222

TABELLEN

Tabelle 1	<i>Itemschwierigkeiten des PDW</i>	96
Tabelle 2	<i>Mittelwerte, Standardabweichungen und Mann-Whitney-U-Tests für Wissen, Motivation und Fähigkeitsselbsteinschätzungen in der Interventions- und Kontrollgruppe</i>	97
Tabelle 3	<i>ANOVA mit Messwiederholung für Wissen, Motivation und Selbstkonzept</i>	98
Table 4	<i>Design Elements for Constructing Text-Based Cases in the Domain of Teachers’ Diagnostics</i>	110
Table 5	<i>Means, Standard Deviations, and Mann-Whitney-U-Tests for Knowledge, Motivation, and Self-Concept</i>	116
Table 6	<i>Repeated Measures ANCOVAS for Knowledge, Motivation, and Self-Concept</i>	118
Table 7	<i>Multivariate meta-analytic regression models with comparative fit indices</i>	121
Tabelle 8	<i>Mittelwerte und Standardabweichungen (in Klammern) für alle Teilnehmenden sowie Interventions- und Kontrollgruppe</i>	137
Tabelle 9	<i>Korrelationen in Pearsons r zwischen Einstellungen, Wissen und Selbstkonzept und zugehörige p-Werte (in Klammern)</i>	138
Tabelle 10	<i>Logistische Regression für richtiges Urteil hoher und niedriger Spezifität</i>	153
Tabelle 11	<i>Logistische Regression für richtiges Urteil hoher und mindestens niedriger Spezifität</i> ...	230

ABBILDUNGEN

<i>Abbildung 1.</i>	Wissensbasiertes Modell fachdidaktischer Diagnose (übernommen von Ostermann, 2018).	36
<i>Abbildung 2.</i>	Modell wissensbasierter diagnostischer Informationsverarbeitung nach Monteiro und Norman (2013). Eigene Darstellung.	38
<i>Abbildung 3.</i>	Prozessmodell diagnostischen Handelns von Lehrkräften nach Klug (2011).	40
<i>Abbildung 4.</i>	Integriertes Prozess- und Strukturmodell der Diagnostik von Ingenkamp und Lissmann (2008).	47
<i>Abbildung 5.</i>	Prozessmodell diagnostischen Handelns auf individueller Ebene der Lehrkraft von Ingenkamp und Lissmann (2008).	48
<i>Abbildung 6.</i>	Diagnostische Anforderungssituationen für Lehrkräfte. Eigene Darstellung.	53
<i>Abbildung 7.</i>	Schritte der kooperativen Fallarbeit in POL-TF.	67
<i>Abbildung 8.</i>	POL-TF im 4C-ID Modell. Eigene Darstellung.	69
<i>Abbildung 9.</i>	Lineare Zusammenhangsveränderungen. Veränderung des Zusammenhangs r in den Gruppen k_1 und k_2 über die Zeit t . Eigene Darstellung.	83
<i>Abbildung 10.</i>	Effekt von Messzeitpunkt x Gruppe auf das Erkennen, Verstehen und Anwenden, Range 0–10.	99
<i>Abbildung 11.</i>	Effekt von Messzeitpunkt x Gruppe auf die Motivation.	100
<i>Abbildung 12.</i>	Effekt von Messzeitpunkt x Gruppe auf das Selbstkonzept.	100
<i>Figure 13.</i>	Effect of Measurement Point x Group on Recognizing (solid line), Understanding (dashed line), and Applying (dotted line).	118
<i>Figure 14.</i>	Effect of Measurement Point x Group on Motivation to engage in pedagogical diagnostics.	119
<i>Figure 15.</i>	Effect of Measurement Point x Group on diagnostic self-concept.	119
<i>Figure 16.</i>	Forest Plot for Standardized Mean Change Differences in Individual Studies and Estimated Multivariate Fixed-Effect and Random-Effects Models.	122
<i>Abbildung 17.</i>	Streudiagramme für Einstellung zu Inklusion, Diagnosewissen und diagnostisches Selbstkonzept; in den Streudiagrammen stellt jeder Punkt einen Messwert einer Person zu einem Messzeitpunkt dar.	140
<i>Abbildung 18.</i>	Mediation von kognitiver Trait-Empathie über affektive State-Empathie auf das richtige Urteil mit niedriger Spezifität.	155

THEORETISCHER TEIL

1 Einleitung

Diagnostisches Handeln und Entscheiden ist ein Kernbestandteil der pädagogischen Praxis von Lehrerinnen und Lehrern und damit auch ein zentraler Gegenstand der Lehrkräftebildung. Letztere steht vor der beständigen Herausforderung, evidenzbasierte Interventionen zur Vermittlung diagnostischer Kompetenz zu entwickeln (Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper, & Lautenbach, 2016). Evidenz bezieht sich dabei gleichermaßen auf empirische Erkenntnisse über diagnostische Prozesse wie auch über hochschuldidaktische Möglichkeiten der Kompetenzentwicklung. Daher lassen sich Forschungsbedarfe insbesondere aus bisherigen Erkenntnissen über die Integration dieser beiden Domänen ableiten.

Empirische Interventionsstudien zur Entwicklung diagnostischer Kompetenz in der ersten Phase des Lehramtsstudiums sehen nahezu durchgehend Übungstätigkeiten in diagnostischen Anforderungssituationen vor, die auf individuelles Handeln in der Interaktion von Lehrkräften mit Schülerinnen und Schülern zielen, wie etwa durch die videogestützte Reflexion der eigenen Wahrnehmung von Unterricht (Klug, 2011; Krammer, 2014; Santagata & Guarino, 2011; Seidel, Blomberg, & Stürmer, 2010; Trittel, Gerich, & Schmitz, 2014). Kooperative Anforderungssituationen im Bereich der Diagnostik, wie etwa die Bildung gemeinsamer Ziele und Handlungspläne, werden dagegen kaum aufgegriffen. Dieser Befund betrifft ebenfalls einen Großteil theoretischer Modelle diagnostischer Kompetenz von Lehrkräften (Herppich et al., 2018; Leuders, Dörfler, Leuders, & Philipp, 2018; Schrader, 2017).

Studien der empirischen Bildungsforschung zeigen, dass die Lehrkräftebildung damit nicht nur eine typische Anforderungssituation der schulischen Realität ausspart, sondern auch, dass in kooperativem diagnostischen Handeln ein bisher kaum ausgeschöpftes Potenzial zur systematischen Kompetenz- und Qualitätsentwicklung liegt (Jurkowski & Müller, 2018; OECD, 2018; Richter & Pant, 2016; Vangrieken, Meredith, Packer, & Kyndt, 2017). Dies wird auch im Blick auf andere Disziplinen deutlich, in denen Diagnostik zum professionellen Handeln gehört. So ist aus Meta-Reviews im Kontext medizinischer und psychiatrischer Diagnostik bekannt, dass Fallbesprechungen die Sensitivität in diagnostischen Prozessen steigern, d. h. tatsächlich vorhandene Störungsbilder werden durch kooperatives Handeln mit höherer Wahrscheinlichkeit auch als solche eingeordnet (Graber et al., 2012; McDonald et al., 2013). Andererseits weist die Forschung zur Lehrkräftebildung und -weiterbildung darauf hin, dass Lehramtsstudierende und Lehrkräfte ihrer Autonomie einen hohen Wert beimessen (Eder, Dämon, & Horl, 2011; Rothland, Biderbeck, Grabosch, & Heiligttag, 2018). Daher können einerseits Forschungsergebnisse aus anderen Disziplinen nicht ohne Weiteres auf die Domäne der Lehrkräfte übertragen werden und andererseits ist unklar, ob sich der Entwicklungsbedarf zu kooperativem Handeln auch in ein entsprechendes Lernverhalten umsetzt. Vor dem Hintergrund dieser Befunde ist es dringend notwendig, den kooperativen Aspekt von Diagnostik theoretisch herauszuarbeiten und hinsichtlich des Lernpotenzials für Lehramtsstudierende empirisch zu untersuchen.

In der Lehrkräftebildung wird in Interventionen zur Kompetenzentwicklung häufig fallbasiertes Lernen genutzt, da es die lernförderlichen Elemente der Realitätsnähe und der didaktischen Reduktion integriert (Bohl et al., 2016; Syring et al., 2015). Problemorientiertes Lernen ist eine Variante fallbasierten Lernens, dessen Effekt auf die Entwicklung diagnostischer Kompetenz in der Lehrkräftebildung bisher kaum untersucht wurde (Chernikova et al., 2019). Obwohl für problemorientiertes Lernen mit Textfällen im medizinischen Kontext insbesondere positive Effekte auf die Entwicklung prozeduralen Wissens und wissensbasierter Elaborations- und Anwendungsfähigkeiten bekannt sind (Strobel & van Barneveld, 2009; Bergstrom, Pugh, Phillips, & Machlev, 2016), wurden differenzielle Effekte dieser Lernmethode auf das Diagnosewissen von Lehramtsstudierenden noch nicht beforscht (Walker & Leary, 2009; Chernikova et al., 2019).

Im Anschluss an diese Befundlage liegt ein erster Schwerpunkt der vorliegenden Dissertation auf der Untersuchung von Effekten einer problemorientiert angelegten Intervention auf die Entwicklung diagnostischer Kompetenz von Lehramtsstudierenden. Im Bewusstsein um die sogenannte Replikationskrise in der Psychologie (Nosek et al., 2015) wird dieser Schwerpunkt ausgebaut, indem eine Replikationsstudie durchgeführt wird und geeignete statistische Modellierungen für die Schätzung von Effektstärken aus replizierten Studien diskutiert und angewendet werden.

Ein weiterer Schwerpunkt liegt in der vertieften Untersuchung der diagnostischen Urteilsbildung aus textförmigen und fallbezogenen Dokumentationen. In der Forschung zu diagnostischer Kompetenz bilden textbasierte Informationen eine bedeutsame Grundlage der Kompetenzmessung. Das betrifft sowohl Material als Einschätzungs- bzw. Urteilsgrundlage in der Expertiseforschung (Glock, Krolak-Schwerdt, Klapproth, & Böhmer, 2013) als auch vielfach angewendete Messinstrumente (Klug, 2011; Kunter et al., 2017). Dies ist auch kongruent zur diagnostischen Praxis von Lehrerinnen und Lehrern, in der verschiedene Arten textbasierter Dokumente eine Urteilsgrundlage bilden. Dazu gehören Texte von Schülerinnen und Schülern selbst, wie bspw. Klassenarbeiten oder Hausaufgaben, aber auch fallbezogene Dokumentationen von Personen aus der Schulpsychologie und -sozialarbeit sowie aus der Sonderpädagogik, der klinischen Psychologie und dem Kollegium. Eine systematische empirische Untersuchung der kognitiven Prozesse des Textverstehens bzw. der mentalen Modellierung von Textinformationen als Voraussetzung diagnostischer Urteilsbildung steht jedoch noch aus. Daher werden diese Zusammenhänge in einer Studie der vorliegenden Disseratation untersucht.

Zwar konzipieren neuere Modelle diagnostischer Kompetenz die Bildung mentaler Modelle über Schülerinnen und Schüler als Bestandteil des Diagnoseprozesses (Leuders, Dörfler, Leuders & Philipp, 2018; Ostermann, 2018). Eine empirische Überprüfung dieser theoretischen Annahme erfolgte bislang jedoch nicht. Die Expertiseforschung weist dahingehend auf Zusammenhänge zwischen einer effektiven und akkuraten Urteilsbildung und der Bildung komplexer und fall- bzw. personenbezogener mentaler Modelle hin (Bradley, Paul, & Seeman, 2006; Hirschfeld & Gelman, 1994). Dies stellt die potenzielle Bedeutung kognitiver und sozial-kognitiver Grundfähigkeiten für Prozesse der Urteilsbildung heraus. Damit ist für Kompetenzforschung aber ein kritischer Punkt erreicht, da diese von einer prinzipiellen Erlernbarkeit der kognitiven Dispositionen ausgeht, die Personen in domänenspezifischen Anforderungssituationen handlungsfähig machen (Klieme, Maag-Merki, & Hartig, 2007). Akkurate Urteile zu treffen, ist nicht nur die am umfangreichsten beforschte Anforderung an diagnostisches Handeln (Schrader, 2017; Südkamp, Kaiser, & Möller, 2012). Eine

Vielzahl an empirischen Studien nutzt die Urteilsakkuratheit von Lehrkräften oder Lehramtsstudierenden auch als Globalindikator diagnostischer Kompetenz (Karing, 2009; Kunter et al., 2013; Schrader, Helmke, Hosenfeld, Halt, & Hochweber, 2006). Nur wenige Studien untersuchten dabei, inwiefern auch kognitive und sozial-kognitive Grundfähigkeiten einen Effekt auf die diagnostische Urteilsakkuratheit haben (einzige Ausnahme: Kaiser, Helm, Retelsdorf, Südkamp, & Möller, 2012). Eine Untersuchung dieser Zusammenhänge erlaubt nicht nur einen tieferen Einblick in eine zentrale diagnostische Anforderung der im Rahmen dieser Dissertation entwickelten und untersuchten Intervention, sie ist auch darüber hinaus von grundlegender methodologischer Bedeutung für die Forschung zu diagnostischer Kompetenz.

2 Kompetenz als theoretisches Konstrukt

Kompetenz ist ein Konstrukt im Schnittpunkt von Wissenschaft und Gesellschaft. Während sich historisch sowohl eine Begriffsgeschichte in der Wissenschafts- als auch in der Alltagssprache nachzeichnen lässt, greifen die jeweiligen Diskurse seit jüngerer Zeit verstärkt ineinander (Winterton, 2009). Die mit dem Wandel zur Wissensgesellschaft einhergehende Auflösung starrer Bildungs- und Arbeitstrajekte stellt eine Beschränkung auf Lern- und Entwicklungsprozessen infrage, die ausschließlich auf Handlungsfähigkeit in festgelegten Situationen zielen (Böhnisch, Lenz & Schröer, 2009). Das Bildungssystem steht vielmehr vor der Herausforderung, Personen auf eine selbstgesteuerte Entwicklung ihrer eigenen Lernpotenziale vorzubereiten (Lempp, 2011). Für eine stärkere Bezugnahme formaler Bildungsprozessen auf die individuelle Entwicklungsdynamik bedarf es jedoch Konstrukte, die eben diesen Vorgang mess- und steuerbar machen. Der Kompetenzbegriff bietet eine solche erweiterte Perspektive auf die Voraussetzungen, Prozesse und Ergebnisse des formalen Lernens an.

Mit der Bindung des wissenschaftlichen Konstrukts der Kompetenz an das gesellschaftliche Bildungssystem ist theoriestrategisch die besondere Situation gegenseitiger Abhängigkeit beider Domänen gegeben. Theoriebildung und Forschung zu Kompetenzen ist daher eng an gesellschaftspolitische Rahmenbedingungen und Entscheidungsvorgänge geknüpft. Zu den Initiativen mit zentraler Bedeutung für die Entwicklung der Kompetenzorientierung, u. a. in der Lehrkräftebildung, gehören das Projekt Definition and Selection of Key Competencies (DeSeCo) der OECD, das Memorandum on Lifelong Learning der EU (Commission of the European Communities, 2000) sowie der 2008 anschließende Europäische Qualifikationsrahmen für lebenslanges Lernen (EQF LLL), die Gründung des Europäischen Zentrums für die Förderung der Berufsbildung (CeDeFOP) und die Verabschiedung der UN-Behindertenrechtskonvention. Ergebnisse dieser politischen Initiativen wurden und werden national von der Kultusministerkonferenz (KMK) und von der Hochschulrektorenkonferenz (HRK) in kompetenzorientierte Standards der Lehrkräftebildung übersetzt. An dieser Definitionsaufgabe hat sich auch das Wissenschaftssystem maßgeblich beteiligt, sodass richtungsweisende theoretische Auseinandersetzungen kurioserweise nicht im wissenschaftlichen Diskurs selbst aufgeworfen wurden, sondern im stärker auf gesellschaftliche Entwicklungen bezogenen Format der Forschungsexpertisen (Klieme, Maag-Merki, & Hartig, 2007; Nuissl, Schiersmann, & Siebert, 2002; Weinert, 1999). Zudem wurden internationale Vergleichsstudien mit Schülerinnen und Schülern (z. B. PISA, TIMSS) und Forschungsprogramme zur Professionalisierung der Lehrkräftebildung durchgeführt (z. B. KoKoHS, COACTIV, TEDS, QLB), die zur Etablierung von Bildungsstandards und zur Qualitätsentwicklung des Bildungssystems beigetragen haben (Blömeke & Kaiser, 2014; Blömeke, Suhl, & Kaiser, 2011; Prenzel & Gräsel, 2015; Kunter, Baumert, Blum, Klusmann, Krauss, & Neubrand, 2013; Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper, & Lautenbach, 2016).

Die gravierende wissenschaftliche und gesellschaftliche Bedeutung des Konstrukts der Kompetenz hat eine kontinuierliche theoretische Weiterentwicklung bewirkt, die aktuell in einen Begriffspluralismus gemündet ist (Koenen, Dochy, & Berghmans, 2015; Zlatkin-Troitschanskaia & Seidel, 2011).

Im kritischen Rationalismus haben Theorien die Funktion, ein Begriffssystem zu entwerfen, aus dem empirisch prüfbare Aussagen über einen Realitätsausschnitt gewonnen werden können (Popper, 1934). Zwar beinhalten alle theoretischen Definitionen von Kompetenz die Beschreibung veränderbarer Voraussetzungen zielgerichteter Handlungsfähigkeit, sie unterscheiden sich jedoch dahingehend, auf die Beschreibung welchen Realitätsausschnitts Aussagen zielen, die aus der Theorie abgeleitet werden können. Vorschläge zur Systematisierung vorhandener Definitionen von Kompetenz greifen an verschiedenen Aspekten des Begriffs an. Beispielsweise an der Kontextspezifität (Gonczi, 1994), der Konstruktstruktur (Weinert, 1999), der Situiertheit (Mulder, 2011), der Berufsbezogenheit (Winterton, 2005) und dem Handlungsbezug (Zlatkin-Troitschanskaia & Seidel, 2011). Diese Systematiken berücksichtigen jedoch nicht, dass Definitionen von Kompetenz eine Vorstellung des Verhältnisses von Wissenschaft und Gesellschaft zugrunde liegt. Ist Kompetenz ein gesellschaftsunabhängiges Konstrukt, so dass Wissenschaft lediglich eine möglichst genaue Beschreibung herstellen kann? Oder ist Kompetenz ein gesellschaftsabhängiges Konstrukt, so dass Wissenschaft entweder eine Steuerungsfunktion oder ebenfalls lediglich eine Beschreibungsfunktion hat? Zentrale Definitionen von Kompetenz unterscheiden sich in ihrer diesbezüglichen Positionierung, die für gesellschaftsbezogene Forschungsbereiche wie der Lehrkräftebildungsforschung von zentraler Bedeutung ist. Daher bildet die Grundlage der folgenden Diskussion zentraler Definitionen von Kompetenz eine Einteilung in generalistischen, normativen und funktionalistischen Ansätze, die auf dieser Positionierung aufbaut (Klieme, Hartig, & Rauch, 2008).

Aus dieser Kategorisierung von Kompetenzdefinitionen folgt ein enger Bezug zur disziplinären Herkunft der jeweiligen theoretischen Begriffe wie auch zu methodologischen Konsequenzen. Ein Großteil vorhandener Kompetenzmodelle greift auf Konstrukte wie Wissen, Fähigkeiten und Fertigkeiten zurück, die jedoch an anderer Stelle bereits hinreichend definiert und von Kompetenz abgegrenzt wurden (Hager & Gonczi, 1996; Winterton, Delamare-Le Deist, & Stringfellow, 2005). Im Folgenden werden grundlegende Dimensionen von Kompetenz aus zentralen Definitionen herausgearbeitet, um das Kompetenzverständnis der vorliegenden Forschungsarbeit und damit verbundene Annahmen darzulegen.

2.1 Generalistische, normative und funktionalistische Kompetenzbegriffe

Generalistische Ansätze zur Definition von Kompetenz zielen auf die Erklärung veränderbarer, problembezogener Handlungsfähigkeit über verschiedene Kontexte und Situationen hinweg. In der Psychologie wurde ein solcher Kompetenzbegriff erstmals maßgeblich von White (1959) eingeführt. Er nutzt das Konstrukt als theoretische Basis einer individualpsychologischen Motivationstheorie. In Anschluss an die strukturalistische Entwicklungspsychologie Piagets (1952) entwirft White (1959) Kompetenz als Motivationsziel in der Interaktion von Individuum und Umwelt und löst Motivation dadurch aus der damals prominenten Triebtheorie Freuds und der Annahme einer uni-direktionalen Beziehung zwischen endogenen innerpsychischen Grundbedürfnissen und ihrer Befriedigung durch die Auseinandersetzung mit der Umwelt. White (1959) formuliert dies so:

As used here, competence will refer to an organism's capacity to interact effectively with its environment. In organisms capable of but little learning, this capacity might be considered an innate attribute, but in the mammals and especially man, with their highly plastic nervous systems, fitness to interact with the environment is slowly attained through prolonged feats of learning. (S. 297)

Demzufolge beschreibt Kompetenz eine durch Lernprozesse erworbene Handlungsfähigkeit, die im Sinne eines generalisierten kognitiven Modul- und Regelsystems das Erreichen von Handlungszielen ermöglicht. Die Erfahrung einer effektiven Auseinandersetzung mit der Umwelt nimmt White (1959) als evolutionsbiologisch begründetes Grundbedürfnis an und benennt dieses als im Lebensverlauf ausdifferenzierbare „effectance motivation“ (S. 323). Dieser motivationstheoretischen Interpretation von Kompetenz liegt eine systematische theoretische Abgrenzung von Kompetenz und Motivation zugrunde (White, 1959, S. 305, 317f., 323). Kompetenz kommt die theoriestrategische Funktion zu, Motivation als dem psychischen Apparat zugehörig, motivationale Ursachen aber nicht rein innerpsychisch oder biologisch zu konzipieren. Da sich die Komplexität der Interaktion mit der Umwelt im Lebensverlauf ändert, schlussfolgert White, dass es unterschiedliche Kompetenzstufen geben müsse (1959, S. 323). Eine genaue inhaltliche Bestimmung von Kompetenz sieht er jedoch als empirische Frage, für die er verhaltensbasierte Kriteriums Stichproben vorsieht: „I now propose that we gather the various kinds of behavior just mentioned, all of which have to do with effective interaction with the environment, under the general heading of competence“ (White, 1959, S. 317).

Damit legt White (1959) grundlegende Dimensionen des Kompetenzbegriffs an. Dazu gehören die dialektische Beziehung von Individuum und Umwelt sowie von Kompetenz und Motivation, die Möglichkeit der Kompetenzentwicklung durch Lernprozesse und der Kompetenzmessung durch verhaltensbasierte Verfahren. Die Charakterisierung des damit eingeführten Kompetenzbegriffs durch Zlatkin-Troitschanskaia und Seidel (2011) im Sinne eines motivationalen Konstrukts, d. h. als „[...] ein Zielzustand (Handlungsfähigkeit), gleichzeitig aber auch „competence motivation““ (S. 222), ist dagegen stark verkürzt.

Eine neue Dimension des Kompetenzbegriffs wurde in der psycho-linguistischen Theorie von Chomsky (1965; 1981) eingeführt, der zur Begründung einer Universalgrammatik zwischen sprachbezogener generalisierter Kompetenz und Performanz unterscheidet. Chomsky geht davon aus, dass die im Sprechakt empirisch wahrnehmbare Grammatik eine durch Umwelteinflüsse verzerrte Form des zugrundeliegenden Regelsystems abbildet. Dieses grundlegende sprachliche Regelsystem sieht er als biologisch vererbte, neuropsychologische Konfiguration und verortet sie auf einer Ebene mit Voraussetzungen zur Embryonalentwicklung (Chomsky, 1981, S. 224, 231, 233).

We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations). Only under the idealization set forth in the preceding paragraph is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence. A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system he puts to use in actual performance. (Chomsky, 1965, S. 3-4)

Diese Definition von Kompetenz als latentem Konstrukt und empirisch beobachtbarem Verhalten als indirektem und mit Messfehlern behafteter Indikator hat nicht nur zur kognitiven Wende in der Psychologie beigetragen, sondern auch nachfolgende Kompetenzdefinitionen geprägt (Erpenbeck, 2002; Weinert, 1999). Ein zentraler Kritikpunkt an Chomskys Theorie bezieht sich auf die Annahme der kausalen Wirkrichtung von Kompetenz auf Performanz (Wood & Power, 1987). Wood und Power (1987) merken an, dass Performanz als analytischer Ausgangspunkt nur die Anteile an Kompetenz untersuchbar macht, die in Performanz münden und damit über Performanz operationalisierte Kompetenz allenfalls ein rein analytisches Konstrukt sein könne, jedoch nicht zwingend einen damit korrespondierenden physikalischen Sachverhalt. In Konsequenz dieser Kritik wäre eine theoretische Konzeption unterschiedlicher Kompetenzprofile oder latenter Kompetenzzustände notwendig, die Homogenität oder Varianz kognitiver Vorgänge über verschiedene Situationen hinweg messbar machen. Erst kürzlich konnte die Annahme Chomskys durch die Identifikation einer anatomischen neuronalen Struktur belegt werden (Fasciculus arcuatus; ein Axonbündel zwischen Wernicke-Areal und Broca-Zentrum), die zur Synthese und Verarbeitung komplexer grammatikalischer Strukturen notwendig ist (Friederici, Chomsky, Berwick, Moro, & Bolhuis, 2017).

Kompetenz ist bei Chomsky ein Konstrukt zur Beschreibung einer physikalischen Größe, die nur unter Zuhilfenahme von Indikatoren über Mess- und Instrumententheorien erfasst werden kann. Im Unterschied zu White (1959) ist Kompetenz nicht in der Interaktion von Individuum und Umwelt verortet. Vielmehr wird diese Interaktion im Sinne beobachtbaren Verhaltens als messfehlerbehaftete Performanz definiert. Eine Konsequenz dieser theoretischen Konzeption ist, dass zwischen Kompetenz und Performanz eine probabilistische Beziehung besteht. Das macht die Anwendung psychometrischer Modelle möglich.

Eine theoretische Lehrstelle in diesem Konzept betrifft die Übersetzung von Kompetenz in Performanz, etwa durch motivationale oder volitionale Prozesse und Merkmale. Diesem Kritikpunkt begegneten spätere entwicklungspsychologische Modelle mit der Einführung von Moderatoren und prozeduralen Kompetenzen (exemplarisch: Gelman & Greeno, 1989), ohne jedoch die angenommene einseitig gerichtete Wirkung von Kompetenz auf Performanz zu thematisieren (kritisch dazu: Ericsson, Krampe, & Tesch-Römer, 1993; Sophian, 1997). Eine wechselseitige Beziehung wäre aber zumindest dann theoretisch notwendig, wenn auch Kompetenzentwicklung möglich sein soll. Ein paralleles Vorhandensein beider Beziehungsarten könnte durch eine Unterscheidung zwischen strukturellen und funktionellen Voraussetzungen von Kompetenz beschrieben werden. Für die eher angewandte Forschung der Lehrkräftebildung sollten letztere besondere Relevanz haben.

Der von Chomsky geprägte Kompetenzbegriff wurde intensiv in der Erziehungswissenschaft rezipiert, etwa von Habermas (1971; 1984) und Roth (1971), die jeweils normative Ansätze einer Definition verfolgen. Habermas stellt in seiner Theorie kommunikativen Handelns die Konstrukte der kommunikativen Kompetenz und der Interaktionskompetenz in den Mittelpunkt der Ausbildung sozial anerkannter Identität und Handlungsfähigkeit. Bei Habermas ist das Individuum selbst an eigenen Entwicklungsprozessen beteiligt, da sich die Grundlagen für Kompetenzen „in einer zugleich konstruktiven und adaptiven Auseinandersetzung des Subjekts mit seiner Umwelt ausbilden“ (1984, S. 191). So wird durch die Aneignung von Lösungsstrategien in verschiedenen Situationen auch das Potenzial erweitert, diese Strategien konstruktiv weiterzuentwickeln. Kompetenzen unterliegen in

dieser Konzeption einem „Selbsterzeugungsprozess“ (1984, S. 192). Die interaktive Situation des Sprechakts, die bei Chomsky als messfehlerbehafteter Kompetenzausdruck konzipiert ist, ist für Habermas grundlegend an der Erzeugung der grammatikalischen Regeln beteiligt, die Kommunikation ermöglichen: „Ein Sprechakt erzeugt die Bedingungen dafür, dass ein Satz in einer Äußerung verwendet werden kann; aber gleichzeitig hat er selbst die Form eines Satzes“ (1971, S. 103). Durch die Berücksichtigung allgemeiner sozialer Strukturen kommunikativer Situationen überführt Habermas (1984) die Universalgrammatik in eine Universalpragmatik, in der sich Kompetenz nicht in einer universalen biologisch angelegten Entwicklungsstruktur erschöpft, sondern auch der Form sozialer Strukturen folgt.

Roth (1971) verwendet den Kompetenzbegriff zur Beschreibung von Mündigkeit, so dass ihm eher eine theoriestrategische Funktion zukommt. Mündigkeit ist bei Roth (1971) ein Ergebnis von Sozialisations- und Erziehungsprozessen, die zu intellektueller, sozialer und moralischer Entscheidungs- und Handlungsfähigkeit führen:

Mündigkeit, wie sie von uns verstanden wird, ist als Kompetenz zu interpretieren, und zwar in einem dreifachen Sinne: a) als Selbstkompetenz (self competence), d. h. als Fähigkeit, für sich selbst verantwortlich handeln zu können, b) als Sachkompetenz, d. h. als Fähigkeit, für Sachbereiche urteils- und handlungsfähig und damit zuständig sein zu können und c) als Sozialkompetenz, d. h. als Fähigkeit, für sozial, gesellschaftlich und politisch relevante Sach- oder Sozialbereiche urteils- und handlungsfähig und also ebenfalls zuständig sein zu können. (S. 180)

Der Kompetenzbegriff Roths (1971) setzt ähnlich wie bei Habermas (1984) und White (1959) an der Interaktion von Individuum und Umwelt an. Kompetenz ist hier jedoch nicht als generalisierte Voraussetzung für Handlungsfähigkeit konzipiert, sondern bekommt eine domänenspezifische Differenzierung (Sach-, Sozial-, Selbstkompetenz) und eine normativ-inhaltliche Ausrichtung. Über eine Persönlichkeitsintegration hinaus setzt Roth (1971, S. 512) selbst- und gesellschaftskritische Reflexivität als Entwicklungsziel. Diese Reflexivität entsteht durch eine kontinuierliche Aneignung des funktionellen Umgangs mit der dinglichen Umwelt (Sachkompetenz) sowie der Internalisierung und kritisch-konstruktiven Anwendung sozialer Normen (Sozialkompetenz). Beide Kompetenzen befähigen zum „Hinausgreifen über die gewussten, gekonnten und bewährten sozialen Rollen und Verhaltensmuster“ zur Lösung sozialer Probleme (Roth, 1971, S. 478). Selbstkompetenz entsteht erstmals durch die Gegenüberstellung von Familien- und Altersgruppennormen und ist deswegen an Reflexionsanstöße und Möglichkeiten zum Erwerb von Sozialkompetenz durch das Bildungssystem gebunden (Roth, 1971, S. 518). Daher ist Kompetenzentwicklung z. T. an das Lebensalter gebunden, insofern mit zunehmendem Alter eine immer größere Wahrscheinlichkeit gegeben ist, kritische Lern- und Entwicklungsanstöße zu erfahren und integriert zu haben (Roth, 1971, S. 500). Mit der Definition von Zuständigkeit als konstitutivem Bestandteil von Kompetenz wird ein konzeptuelles Element zum Transfer von Kompetenz zu Performanz eingeführt. Dieses besteht ähnlich wie bei Habermas (1984) in moralischen Prinzipien sowie in prosozialen und demokratischen Wertvorstellungen (Roth, 1971, S. 500).

Eine Integration der inhaltsbezogenen Differenzierung und der wertgebundenen Eigenschaften des Kompetenzkonstrukts der erziehungswissenschaftlichen Modelle sowie der performanzbezogenen Indikatoren der psychologischen Modelle stellt Erpenbeck (2002; 2014) gemeinsam mit Heyse (1997) auf.

Kompetenz wird von Erpenbeck und Heyse (1997) mit Hilfe des Dispositionsbegriffs in Abgrenzung vom Qualifikationsbegriff definiert und in die Differenz von Kompetenz und Performanz überführt. Der Qualifikationsbegriff beschreibt eine zweckgebundene, anforderungsorientierte und objektbezogene Beziehung zwischen Individuum und Umwelt, mit dem sich Voraussetzungen für ein Spektrum konkreter Handlungen zusammenfassen lassen (Erpenbeck, 2002, S. 5). Kompetenz beschreibt dagegen eine wertgebundene, selbstorganisierte und subjektbezogene Beziehung zur Umwelt, die wiederum Voraussetzung für eine selbstorganisierte Handlungsvielfalt auf Performanzebene ist. Die Selbstorganisationsdispositionen münden bei Erpenbeck (2004, S. 21) in ebenfalls weiter ausdifferenzierte personale Kompetenz, Umsetzungskompetenz, Fach- und Methodenkompetenz und sozial-kommunikative Kompetenz, wobei er dezidiert an den Europäischen Qualifikationsrahmen und die Forderung nach Schlüsselqualifikationen bzw. Key Competencies anschließt. Anders als in psychologischen Modellen besteht der theoretische Zweck des Kompetenzkonstrukts bei Erpenbeck (2002) nicht in der statistischen Vorhersage, sondern in der sinnverstehenden Rekonstruktion erfolgreicher Performanz „als kontingente, historisch je singuläre Form des Selbstorganisationsprozesses“ (Erpenbeck, 2002, S. 9). Erpenbeck und Heyse (1997) schreiben Kompetenz gegenüber soziodemografischen und situationalen Merkmalen ein größeres Potenzial für eine intentional gesteuerte Veränderung von Bildungsergebnissen zu und fordert daher vom Bildungssystem eine kompetenzorientierte Ausrichtung und Steuerung ein.

Zusammengefasst greifen die normativen Ansätze zur Definition von Kompetenz die Differenz von Kompetenz und Performanz auf, stellen jedoch daraus abgeleitete prognostische Aussagen infrage. Der Kompetenzbegriff dieser Ansätze operationalisiert Teilbereiche der Individual- und Persönlichkeitsentwicklung in Bezug auf ein konkretes Subjektverständnis. Mit der Konzeptualisierung ethisch-moralischer Werte als zentralem Bestandteil von Kompetenz ist diese grundlegend umwelt- und handlungsbezogen. Durch die stärkere Einbindung der Struktur der sozialen Umwelt in das Konstrukt selbst, kommt Bildungsinstitutionen nicht nur eine Vermittlungsfunktion, sondern auch eine konstitutive Funktion zu. In Konsequenz können Kompetenzen nicht ohne Modelle der Kompetenzentwicklung definiert werden, da die vermittelnden Lern- und Entwicklungsprozesse auch strukturgebend sind. Durch diesen theoretischen Kniff des konstitutiven Praxisbezugs wurde Erpenbecks (2004) Kompetenzmodell insbesondere in berufspädagogischen Konzepten rezipiert (Niedermaier, 2012).

Die in der Schul- und Lehrkräftebildung entwickelten bildungswissenschaftlichen Kompetenzmodelle beziehen sich wesentlich stärker auf funktionalistische Definitionsansätze. Diese zielen auf die Beschreibung von Voraussetzungen domänen- bzw. kontextspezifischer Handlungsfähigkeit.

Einen solchen Ansatz leitet Weinert (1999) aus der Gegenüberstellung verschiedener Definitionsvorschläge zur Auswahl und Beschreibung von Schlüsselkompetenzen im DeSeCo

Programm der OECD ab. Dazu formuliert Weinert zunächst zehn Probleme, die eine theoretische Kompetenzdefinition zufriedenstellend lösen sollte (1999, S. 15-20):

- (1) die Bestimmung des Abstraktionsniveaus des Konstrukts, das ‚unterhalb‘ generalisierter kognitiver Fähigkeiten aber ‚oberhalb‘ situations- und ausgabenspezifischer kognitiver Fähigkeiten liegt;
- (2) die Klärung der Unterscheidbarkeit und Verhältnissetzung von Kompetenz und Performanz, die inter-individuelle Unterschiede in intra-individuellen Veränderungen berücksichtigt;
- (3) die inhaltliche Bestimmung in Form erlernbarer kognitiver Fähigkeiten, im Kontrast zu entwicklungspsychologisch stabilen Konstrukten;
- (4) die inhaltliche Bestimmung durch den Kontext, in dem der Handlungsvollzug stattfindet;
- (5) die inhaltliche Bestimmung durch domänenspezifische Anforderungen (z. B. Schulfächer);
- (6) der Ausschluss kognitiver Basisfähigkeiten (z. B. Intelligenz, Arbeitsgedächtnis) und die Benennung domänenspezifischer kognitiver Fähigkeiten (z. B. für bestimmte Problemarten);
- (7) die Abgrenzung von Metakompetenzen (z. B. Fähigkeiten zur Adaptation vorhandener Kompetenzen an neue Anforderungen) und Schlüsselkompetenzen (z. B. transcurriculare Kompetenzen);
- (8) die Bestimmung von Möglichkeiten zur Kontrolle spezifischer motivationaler und volitionaler Konstrukte, die eine Kompetenzmessung potenziell konfundieren;
- (9) die Bestimmung des Kriteriums von Kompetenz bzw. erfolgreicher Performanz;
- (10) die Bestimmung instruktionaler oder erfahrungsgebundener Möglichkeiten der Kompetenzentwicklung zur Abgrenzung von entwicklungs- und persönlichkeitspsychologisch stabilen kognitiven Konstrukten.

Bisherige Definitionen ordnet Weinert (1999) einem konditionalen oder funktionalistischen Ansatz zu. Eine konditionale Definition würde das Kompetenzkonstrukt durch alle kognitiven Konstrukte beschreiben, die einen von generellen kognitiven Fähigkeiten unabhängigen Varianzanteil gezeigter Performanz beschreiben. Der funktionale Ansatz dagegen würde Kompetenz durch Beschreibung der spezifischen Konfiguration kognitiver Voraussetzungen definieren, die für die Bewältigung einer Klasse von Anforderungen, Situationen und Aufgaben notwendig sind.

Beide Definitionsansätze lassen sich in Bezug auf ihre methodologischen Folgen diskutieren. Der konditionale Ansatz öffnet den Kompetenzbegriff in die Breite, da ihm ein prinzipiell unabgeschlossener Raum potenziell performanzförderlicher Konstrukte zugrunde liegt. Die dadurch entstehenden inter-individuellen Unterschiede in kognitiven Strukturen unter demselben Begriff würden jedoch vergleichende Kompetenzmessungen und die Definition von Niveauunterschieden erschweren. Dagegen erfordert der funktionale Ansatz eine vorzuschaltende theoretische oder empirische Klassifikation von Anforderungen in einer Domäne, die Definition eines Kriteriums erfolgreicher Performanz und die anschließende empirische Erschließung inter-individuell vorhandener funktionaler kognitiver Strukturen. Diese Einengung des Kompetenzbegriffs erleichtert die Messung von Niveauunterschieden, eröffnet aber auch die Möglichkeit, dass eine einheitliche kognitive Struktur zur Bewältigung von Anforderungen in einem Kontext gar nicht vorhanden ist. Entsprechend ist jegliche Kompetenz immer zuerst ein analytisches Konstrukt, bevor sich eine funktionelle Struktur objektiv nachweisen lässt. Diese Entwicklung zeigt sich u. a. in Bezug auf

mathematische Kompetenz, da zunächst als latente Merkmale gemessene Teilkompetenzen nun auch manifest in neuronalen Netzwerken identifiziert wurden (Grabner et al., 2007; Price, Yeo, Wilkey, & Cutting, 2018; Skagerlund et al., 2018).

Weinert spricht sich weniger aus theoretischen als eher aus pragmatischen Gründen der Messbarkeit und der Anschlussfähigkeit an das Bildungssystem für eine funktionalistische Definition und die Beschränkung des Kompetenzkonstrukts auf „learned knowledge, skills and corresponding metaknowledge“ aus (1999, S. 28). In einer Argumentation für die Steuerung des Bildungssystems durch leistungsbezogene Vergleichsstudien entwirft Weinert (2001) eine konkrete Definition von Kompetenz:

Dabei versteht man unter Kompetenzen 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. (S. 27f)

Während diese Definition auch Konstrukte zur Übersetzung von Kompetenz zu Performanz in das Kompetenzkonstrukt integriert, trennt Weinert (1999) diese nachfolgend als Handlungskompetenzen von fachlichen Kompetenzen ab. Erstere ordnet er als Voraussetzung von Kompetenzentwicklung ein und knüpft damit an seine vorherige Argumentation zur Abgrenzung von Motivation und Kompetenz an (1999, S. 28; 2001, S. 28). In großen Teilen der Kompetenzforschung wurde dieser Punkt jedoch übersehen, so dass motivationale Konstrukte häufig als Teil von Kompetenz der Definition durch Weinert (1999) zugeschrieben und als Ziel und Evaluationskriterium für Interventionen zur Kompetenzentwicklung festgelegt werden (bspw. Klug et al., 2016; Lipowsky, 2006; Zlatkin-Troitschanskaia & Seidel, 2011; zu dieser Kritik siehe auch Hartig, 2006; Hartig & Klieme, 2006; Klieme, 2004).

Ähnlich wie Erpenbeck (2004) greift Weinert (1999) die normative Dimension der Verantwortungsübernahme auf, lagert diese jedoch in Abgrenzung zu kognitiven Fähigkeiten aus dem Kompetenzbegriff aus. Mit dem Fokus auf funktionelle kognitive Strukturen gelingt ein Zusammenschluss des biologischen Materialismus generalistischer Ansätze mit der umweltbezogenen Ausrichtung normativer Konzepte, ohne die inhaltliche Festlegung letzterer zu übernehmen.

Für eine Anwendung in der Steuerung des Bildungssystems wurde der funktionalistische Kompetenzbegriff weiter präzisiert (Koeppen, Hartig, Klieme, & Leutner, 2015). Klieme und Leutner (2006) definieren Kompetenz als „kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen“ (S. 879). Dabei beschreiben Domänen ausbildungs- und berufsbezogene Lernfelder und fachbezogene Leistungsbereiche, z. B. mathematisches Modellieren, kaufmännisches Entscheidungshandeln und informationstechnische Ablauforganisation. Klieme, Maag-Merki und Hartig (2007) weisen darauf hin, dass dieser Kompetenzbegriff nicht das gesamte Bildungs- und Entwicklungsgeschehen einer Person erfassen soll und dementsprechend institutionelle Steuerungsprozesse nicht ausschließlich auf Grundlage von Kompetenzmessungen erfolgen kann. Kompetenzmessungen sollen vielmehr zur Etablierung von Bildungsstandards im Sinne minimaler Ziele und Erwartungen an Lernende und

Bildungsinstitutionen beitragen. Damit wird die gesellschaftliche Bedeutung des Kompetenzkonstrukts auch deutlich von persönlichkeitsbezogenen Entwicklungsprozessen abgegrenzt. Für berufliche Kompetenzen von Lehrkräften fordern Klieme und Leutner (2006) einen Anschluss empirischer Untersuchungen an die Expertiseforschung sowie an theoretische Modelle professioneller Kompetenz von Lehrpersonen.

In der Lehrkräftebildungsforschung wurde diese Forderung aufgegriffen und zur Entwicklung einzelner Aspekte von Kompetenzmodellen genutzt, wie etwa Wahrnehmungs-, Interpretations- und Entscheidungsfähigkeiten als Mediatoren zwischen Kompetenz und Performanz (Blömeke & Zlatkin-Troitschanskaia, 2015; Blömeke, Gustaffson, & Shavelson, 2015; van Es & Sherin, 2008), differenzielle Effekte von institutionellem Kontext, Situation und Kompetenz auf Performanz (Santagata & Guarino, 2011; Santagata & Yeh, 2016; Scheerens & Blömeke, 2016), die kategoriale Unterscheidbarkeit von Kompetenz und Performanz (Blömeke et al., 2015; Blömeke & Kaiser, 2017), die Integration wertbezogener und motivationaler Konstrukte (Blömeke & Kaiser, 2017) sowie kollektiver Anteile (Amhag & Jakobsson, 2009).

2.2 Kognitive und psychometrische Modellierung von Kompetenz

Kompetenz kann durch eine Operationalisierung in psychometrischen Modellen empirisch gemessen werden. Psychometrische Modelle formulieren theoretische Annahmen über Beziehungen zwischen kognitiven Prozessen und Merkmalen mit statistischen Gleichungen aus (Eid, Gollwitzer, & Schmitt, 2011). Die empirische Kompetenzforschung leitet zentrale Annahmen über Voraussetzungen domänen- oder kontextspezifischen Performanz aus der Expertiseforschung ab (Klieme & Hartig, 2006).

Die Expertiseforschung fragt nach den Voraussetzungen überdurchschnittlich ausgeprägter domänenspezifischer Leistungen (Borko & Livingston, 1989; Sternberg & Horvath, 1995). Für Leistung existieren unterschiedliche Indikatoren, wie etwa die Reaktions- und Verarbeitungsgeschwindigkeit (Brochard, Dufour, & Despres, 2004), die abrufbare Informationsmenge und -akkuratheit nach Situationsanalysen (Chase & Simon, 1973), die visuelle Fixationshäufigkeit (van Gog, Paas, van Merriënboer, & Witte, 2005), die Fähigkeit zu domänenspezifischer Mustererkennung (Greenwood & King, 1995), die Komplexität von Wissens- und Problemrepräsentationen (Green & Gilhooly, 1992), die Problemlösefähigkeit (Ericsson, Delaney, Weaver, & Mahadevan, 2004) und die Adaptions- und Interpretationsfähigkeit (Ward, Gore, Hutton, Conway, & Hoffmann, 2017).

Für Lehrkräfte wurde Expertise insbesondere von Bromme (1997, 2001, 2008) und Berliner (1991, 1992, 2001) definiert. Bromme beschreibt Expertise im Anschluss an kognitionspsychologische Ansätze als „kohärentes Ensemble von Wissen und Können, das in einer Person vereinigt ist“ (1997, S. 187) und zu situational angemessener sowie flexibler Aufrechterhaltung multipler Zielorientierungen befähigt (Bromme, 2001). Konkret benennt Bromme (2008) automatisierte kognitive Prozesse, d. h. Routinen sowie ein breites und tiefes domänenspezifisches Wissen als Merkmale von Expertise. Berliner (2001) zieht vor dem Hintergrund empirischer Forschung zu inter-individuellen Unterschieden zwischen Lehrkräften multiple Definitionskriterien heran, ohne jedoch deren Zusammenhang zu thematisieren: (1) Automatisierung sich wiederholender Handlungsvorgänge, (2) Sensitivität gegenüber

Situationsmerkmalen, (3) Flexibilität im Unterrichten, (4) individualisierte kognitive Problemrepräsentationen, (4) schnelle und akkurate Mustererkennung, (5) Sinnverstehen in der Expertisedomäne und (6) Einbezug personenbezogener Informationen in Problemlösungen. Beide Autoren fokussieren auf ein hohes Niveau domänenspezifischen Wissens als zentrale Voraussetzung überdurchschnittlicher Leistung.

Die Forschung zu Professionswissen als Kern von Expertise stellt die Struktur der Wissensorganisation als besonderes Merkmal heraus (Bromme, 2008). Während Novizinnen und Novizen über ein hierarchisch organisiertes und kompartimentiertes Wissen verfügen, zeichnen sich Expertinnen und Experten durch ein stärker vernetztes und integriertes Wissen aus, das sich um praktisch bedeutsame Anforderungen und Problemstellungen im Sinne einer fallbezogenen Struktur organisiert (Renkl, Mandel, & Gruber, 1996). Kognitionspsychologische Ansätze beschreiben diese Art der Wissensorganisation als Voraussetzung für den Vorgang der Spreading Activation (Anderson, 1983), d. h. ein externer Stimulus aktiviert nicht nur einzelne, damit assoziierte Wissensbausteine, sondern löst durch die wechselseitige Verknüpfung unterschiedlicher Wissensbausteine und -arten in Bezug auf den Stimulus eine weitreichende Aktivierungskette aus. Dieser kognitive Prozess ist eine Voraussetzung für die flexible Handlungsplanung, Mustererkennung und das schnelle Abrufen gespeicherter Denk- und Handlungsschemata (Raufaste, Eyrolle, & Mariné, 1998).

Dieses Modell der Wissensorganisation hat sich auch in der Forschung zu Expertise von Lehrkräften durchgesetzt (Bromme, 1997; Leuders, Nückles, Mikelskis-Seifert, & Phlipp, 2018; Seidel & Prenzel, 2006). Ausgehend von der Dimension der inhaltlichen Strukturierung des professionellen Wissens von Lehrkräften in pädagogisch-psychologisches Wissen, Fachwissen und fachdidaktisches Wissen (Baumert & Kunter, 2006; Borko & Putnam, 1996; König et al., 2018; Leuders et al., 2018; Shulman, 1986) wird im Hinblick auf die fallbezogene Integration zunehmend auch die Nähe des Wissens zur Schulpraxis als zweite Dimension der Wissensstruktur angenommen (Ball, Thames, & Phelps, 2008; Dreher, Lindmeier, Heinze, & Niemand, 2018). Diese zweite Dimension ist auch Ausgang von Modellen der Entwicklung von Expertise durch den Erwerb und die Organisation von Professionswissen.

Mit dem Ziel die instruktionale Einflussnahme auf die Entwicklung bzw. Veränderung von Expertise zu erklären, entwirft Berliner (2004) in Anschluss an Alexanders (2004) mit dem Modell des Domain Learning ein Stufenmodell für die professionelle Entwicklung von Lehrkräften. Die sechs Stufen von Noviziat bis zum Expertentum sind gekennzeichnet durch die zunehmende Kontextualisierung theoretischer Wissensbestände durch unterrichtspraktische Situationen und fallbezogenes Wissen, die Steigerung selektiver Aufmerksamkeit, die gesteigerte Fähigkeit zum abstrakten Vergleich von Unterrichtsstunden und die gesteigerte Fähigkeit zum flexiblen umschalten zwischen Unterrichtsskripts. Während Berliner (2004) in Anschluss an Ericsson et al. (2004) reflektierte und professionell begleitete Praxis als zentralen Lern- und Entwicklungsmodus benennt, kommen mehrere Reviews zu dem Schluss, dass sich diese These empirisch nicht stützen lässt (Baker & Young, 2014; Macnamara, Hambrick, & Oswald, 2014; Macnamara, Moreau, & Hambrick, 2016). Hambrick, Burgoyne, McNamara und Ullén (2018) schlagen statt dem unimodalen Entwicklungsmodell ein multifaktorielles Modell aus fähigkeitsbezogenen und umweltbezogenen Faktoren vor, die sich entwicklungsförderlich oder -hemmend auswirken können.

Insgesamt legen die Erkenntnisse der Expertiseforschung nahe, zur empirischen Untersuchung des Kompetenzkonstrukts drei zentrale Modellarten zu unterscheiden: (1) Strukturmodelle, (2) Niveaumodelle und (3) Veränderungsmodelle.

Strukturmodelle beschreiben die einzelnen kognitiven Module und Prozesse sowie deren Verbindungen untereinander, die zur Bewältigung aufgabenspezifischer Anforderungen notwendig sind. Die Modelle lassen sich als Verbindung von personenspezifischen und situationsspezifischen Leistungsfaktoren verstehen und benennen inter-individuelle und situationsübergreifende Merkmale, „zum Beispiel spezifische Fähigkeiten und Fertigkeiten oder bereichsspezifisches Wissen“ (Klieme, Maag-Merki, & Hartig, 2007, S. 13). Insofern diese Teilkomponenten und ihr funktionelles Zusammenspiel Kompetenz ausmacht, können diese Teilkomponenten als Dispositionen, d. h. sozial erwerbbar Anlagen für Kompetenz bezeichnet werden (Blömeke et al., 2015; Hartig & Klieme, 2006). Die differenzierte Abbildung leistungsförderlicher Teilkomponenten bildet eine entscheidende Grundlage der Identifikation geeigneter Interventionen zur Kompetenzförderung und für daran anschließende institutionelle Gestaltungsprozesse der Aus- und Weiterbildung von Lehrkräften (Curriculae, Lern- und Prüfungsformen etc.).

Niveaumodelle beschreiben die Ausprägung der Schwierigkeit situationsspezifischer Anforderungen, zu deren Bewältigung Kompetenz eine notwendige (aber keine hinreichende) Voraussetzung ist (Klieme, Hartig & Rauch, 2008). Zur Definition dieser Bewältigungsleistung können sowohl normative Ziele als auch empirische Leistungsmessungen aus der Expertiseforschung verwendet werden (Thilakaratne & Kvan, 2006). Eine solche inhaltliche Definition von erreichten Testergebnissen bzw. Skalenabschnitten erlaubt eine kriteriumsorientierte Beschreibung des Kontinuums von minimaler und maximaler Kompetenz, einen Vergleich zwischen empirisch vorliegendem und angestrebtem Lern- und Entwicklungsstand auf Individual- und Aggregatsebene sowie in Konsequenz die Formulierung von Bildungsstandards.

Veränderungsmodelle beschreiben Kompetenzentwicklung im Sinne des Anstiegs bzw. Abfalls eines erreichten Niveaus sowie hinsichtlich der Verlaufsformen. Zu den Verlaufsformen selbst existieren kaum konzeptuelle Ansätze (Klieme, Hartig, & Rauch, 2008). Häufig wird von einem kontinuierlichen und linearen Entwicklungsverlauf ausgegangen, der im Wesentlichen als Ergebnis der Akkumulation kognitiver Fähigkeiten gesehen wird (Prenzel et al., 2005; Klug et al., 2016; Blömeke et al., 2015). Seltener werden diskontinuierliche und nicht-lineare Entwicklungsprozesse angenommen, die durch qualitative Entwicklungsschwellen und -sprünge gekennzeichnet sind (von Aufschnaiter & Rogge, 2010; Rauner & Zeymer, 1991). Die Schwierigkeit der Modellierung von nicht-linearen Verlaufsmodellen besteht dabei vor allem in der dafür notwendigen hochfrequenten Erhebung longitudinaler Daten über eine sehr langen Zeit, da für die Entwicklung von Expertise ein Zeitraum von etwa 10 Jahren angenommen wird (Bromme, 2008). Aus der qualitativen Bildungsforschung sind eine Vielzahl solcher Modelle in Form sogenannter transformatorischer Bildungsprozesse vorhanden, die jedoch in der quantitativ orientierten Kompetenzforschung bislang kaum rezipiert werden (Kokemohr, 2007; Koller, 2018; Koller, Marotzki, & Sanders, 2015; Marotzki, 1990; Mezirow, 1991; Nohl, 2015; Peukert, 2000).

Ein generalisiertes Schwellen- und Stufenmodell für die Entwicklung von Expertise legen Dreyfus und Dreyfus (2004) vor. Sie gehen davon aus, dass zum Erwerb von Expertise spezifische

Lernsituationen durchlaufen und darin bestimmte Erfahrungen gesammelt werden müssen. Zum Wechsel von der Stufe Competence zur Stufe Proficiency bieten etwa Situationen eine Lern- und Entwicklungschance, in der bereits eigene Handlungspläne verfolgt werden, deren Erfolg aber noch nicht abgeschätzt werden kann. So stellen Dreyfus und Dreyfus (2004) fest:

Given this uncertainty, coping becomes frightening rather than merely exhausting. Prior to this stage, if the rules do not work, the performer, rather than feeling remorse for his or her mistakes, can rationalize that he or she had not been given adequate rules. But because at this stage the result depends on the learner's choice of perspective, the learner feels responsible for his or her choice. Often, the choice leads to confusion and failure. But sometimes, things work out well, and the competent student then experiences a kind of elation unknown to the beginner. (S. 178)

Ein ähnliches Modell schlagen Rauner und Zeymer (1991) vor. In ihrem Modell braucht es zum Wechsel von Kompetenten Akteuren zu Gewandten Professionellen Lernchancen in Form einer Konfrontation mit komplexen Handlungs- und Problemsituationen ohne vorgedachte Lösungen, die nur über erfahrungsgestützte Fähigkeiten lösbar sind.

Um diese Modelltypen mit empirischen Beobachtungen zu verbinden, lassen sich in psychometrischen Modellen die genauen Regeln spezifizieren, mit denen eine Zuweisung von Messwerten zu Modellgrößen erfolgt (Edelsbrunner & Dablander, 2019). Diese Instrumententheorien (Popper, 1934) zielen auf Aussagen darüber, wie zuverlässig ein Test ein Konstrukt messen kann, aber auch, welche individuellen Merkmalsausprägungen in einer Stichprobe vorliegen, wie sie verteilt sind und wie sie sich verändern.

Zur Erschließung von Kompetenzstrukturen und -niveaus eignen sich insbesondere variablenzentrierte Analysemodelle aus der klassischen Testtheorie (Strukturgleichungsmodellierung) und der probabilistischen Testtheorie (IRT Modellierung) (zum Überblick siehe Hartig & Frey, 2013). Der wohl bedeutendste Unterschied beider Messtheorien liegt in der Differenzierung von Personen- und Itemparametern (Bond & Fox, 2001). Personenparameter treffen Aussagen über die Personenstichprobe, bspw. gibt die Personenfähigkeit an, mit welcher Wahrscheinlichkeit eine Person die Fähigkeit zur Lösung einer Aufgabe hat. Itemparameter treffen Aussagen über die Itemstichprobe, bspw. gibt die Itemschwierigkeit an, mit welcher Wahrscheinlichkeit eine Aufgabe von Personen gelöst wird. In der klassischen Testtheorie wird die Itemschwierigkeit aus den Personenfähigkeiten der Stichprobe geschätzt und die Personenfähigkeit als Funktion der Itemschwierigkeiten (Geiser, 2011). Diese zirkuläre Annahme birgt einige Limitationen, welche die mathematisch komplexeren IRT Modelle durch eine Trennung von Iteminformationsfunktion (IIF) und Testinformationsfunktion (TIF) überwinden (Rost, 2004). Dazu wird die Itemschwierigkeit auf Basis der Anzahl von Personen mit korrekten Antworten geschätzt und die Personenfähigkeit auf Basis der Anzahl korrekt beantworteter Items im gesamten Test (Osteen, 2010). Diese Trennung ist besonders für Kompetenzmessungen interessant, da die vergleichende Analyse von Personen- und Itemparametern Aufschluss über kontextspezifische Anforderungen der Testsituation im Verhältnis zu kognitiven Leistungsdispositionen geben kann. Dementsprechend bestehen an Testinstrumente zur Kompetenzmessung neben den klassischen Testgütekriterien noch zusätzliche Anforderungen (Klieme,

Hartig, & Rauch, 2008). In Abgrenzung zur Messung situationsunabhängiger Konstrukte (z. B. Intelligenz) müssen Aufgaben zur Messung von Kompetenz sowohl die theoretisch angenommenen kognitiven Module und Prozesse abbilden, die zur Bewältigung kontextspezifischer Anforderungssituationen notwendig sind (Personenfaktor), als auch die Struktur dieser Anforderungen selbst (Situationsfaktor) (Hartig & Klieme, 2006). Dadurch steigt die Bedeutung von Instrumentenvalidierungen enorm an.

Die Analyse von Niveauveränderungen mit Längsschnittdaten ist im Rahmen beider Messtheorien mit variablenzentrierten Modellen nur dann ohne Umwege möglich, wenn ein zeitstabiles Strukturmodell vorliegt. Unter dieser Annahme können zur Schätzung von Veränderungen in Item- und Personenparametern sowohl bei unidimensionalen Konstrukten das Linear Logistische Testmodell aus der klassischen Testtheorie (LLTM; Fischer, 1983) als auch bei mehrdimensionalen Konstrukten das Multidimensional Rasch Model for Learning and Change (MRMLC; Embretson, 1991) aus der probabilistischen Testtheorie verwendet werden. Treten Niveau- und Strukturveränderungen jedoch gleichzeitig auf, etwa im Fall der Umorganisation einer kognitiven Struktur bei qualitativen Entwicklungssprüngen, könnte diese Veränderung durch die notwendige Annahme der Messinvarianz über die Messzeitpunkte nicht abgebildet werden. Demgegenüber ist in personenzentrierten Ansätzen der klassischen Testtheorie (z. B. latente Transitionsmodelle) zwar eine Veränderung von Kompetenzstrukturen als Veränderung einer latenten Klassenzugehörigkeit abbildbar, diese kann aber nicht gleichzeitig als Niveauveränderung interpretiert werden.

Einen vielversprechenden Ansatz bietet das hierarchische Mischverteilungs-Latent-State-Trait-Markov-Modell von Crayen, Eid, Lischetzke und Vermunt (2017), da das Modell u.a. Aussagen über die Wahrscheinlichkeit des Wechsels zwischen verschiedenen States von einem Messzeitpunkt zum anderen trifft. Angewendet auf Kompetenzveränderungen könnten darin Strukturveränderungen über die Trait-Komponente und Niveauveränderungen über die State-Komponente abgebildet werden. Ein bestechender Vorteil des Modells bezüglich der Modellierung qualitativer Strukturveränderungen besteht in der Markov-Eigenschaft, d. h. nur Effekte direkt aufeinanderfolgender Messzeitpunkte werden modelliert (z. B. t_2 auf t_3), für weiter entfernte Messzeitpunkte wird ein statistischer Gedächtnisverlust angenommen. Begrifflich arbeiten Markov-Modelle mit Zuständen und Übergangswahrscheinlichkeiten (Sonnenberg & Beck, 2013). Da mit Kompetenz das kognitive Potenzial zur Handlungsfähigkeit in kontextspezifischen Anforderungssituation gemessen werden soll, d. h. im messtheoretischen Sinn ein anforderungsbezogener Zustand, sollte ein Wechsel der Anforderungssituation auch zum Verlassen dieses Zustands führen. Umgekehrt lassen sich mit Markov-Modellen unter dieser Annahme auch die Grenzen von Anforderungssituationen identifizieren (z. B. in Experimentalstudien), was eine genauere Beschreibung einer spezifischen Kompetenz ermöglicht. In ähnlicher Form werden Markov-Modelle zur neuropsychologischen Modellierung von funktionellen kognitiven Zuständen sowie von Lern- und Gedächtnisprozessen genutzt (Niv, 2009; Ou et al., 2013; Rajapakse & Zhou, 2007; van Rossum & Shippi, 2013) aber auch zur Identifikation klinisch-psychologischer Störungsbilder (Martinez-Murcia, Górriz, & Ortiz, 2016) und für technologiegestütztes Lernverhalten (Jeong et al., 2008).

Zusammengefasst ermöglichen Struktur-, Niveau- und Veränderungsmodelle eine umfangreiche und tiefgehende Untersuchung von Kompetenz. Geeignete psychometrische Modelle sind für

unterschiedliche Analysezwecke vorhanden, setzen jedoch typischerweise größere Stichproben und aufwendige Erhebungsdesigns voraus, die in der Interventionsforschung (und darüber hinaus noch zugespitzt in der Lehrkräftebildung) jedoch nur selten gewonnen werden können (Bromme, Jäger, & Prenzel, 2014).

2.3 Kritik am Kompetenzparadigma

Am Kompetenzparadigma wird nicht zuletzt wegen der engen Verbindung zu bildungspolitischen Steuerungsprozessen aus verschiedenen wissenschaftlichen Perspektiven heraus Kritik geübt. Vor dem Hintergrund einer Ungleichheitstheoretischen Perspektive lässt sich etwa feststellen, dass in funktionalistischen Kompetenzbegriffen und daran anschließender Forschung gesellschaftliche Bewertungs- und Anerkennungsprozesse kaum berücksichtigt werden (Allmendinger, 2003). Während der Bildungsdiskurs in der Debatte zu informellem Lernen und lebensweltlich sinnhaftem Bewältigungshandeln auch die Voraussetzungen für Handlungsfähigkeit in Anforderungssituationen rekonstruiert, die außerhalb normalbiografischer Lebensverläufe liegen, setzt der Kompetenzdiskurs an interindividuell häufigsten und typischen Bewältigungsanforderungen an und riskiert dadurch die Ausgrenzung breiter Entwicklungsfelder und -dynamiken (Lempp, 2011). So ließe sich kritisch fragen, warum keine Kompetenzen im Bereich abweichenden Verhaltens definiert und untersucht werden, wenngleich deren Erwerb auch in formalen Bildungskontexten stattfindet (bspw. Erlernen der Praxis körperlicher Gewalt in Polizeischulen, Jacobsen, 2001; Erlernen peer-kulturell anerkannter Praxis des Betrügens in Schulleistungsmessung, Högberg, 2011). Zwar sind entsprechende Untersuchungen mit funktionalistischen Kompetenzbegriffen möglich, der wissenschaftliche Diskurs im Kompetenzparadigma reproduziert durch die Untersuchung ausschließlich gesellschaftskonformer Lern- und Entwicklungsfelder jedoch häufig unreflektiert eine normative bildungsbürgerliche Subjektkonzeption (Backes-Haase & Klinkisch, 2015). Mit den Konstrukten der Handlungsfähigkeit und Bewältigung, auf denen der Kompetenzbegriff basiert, ist jedoch eine prinzipielle Anschlussmöglichkeit an Sozialisationstheorien gegeben, die bereits zahlreiche Felder der Kompetenzentwicklung abseits normalbiografischer Bewältigungsanforderungen identifiziert haben.

Aus bildungstheoretischer Perspektive werden sowohl die Reichweite von Kompetenzmodellen, die Annahme der Steuerbarkeit von Lern- und Entwicklungsprozessen als auch der Leistungsbegriff kritisiert (Lempp, 2011). Bis auf die generalisierten Kompetenzbegriffe zielen die theoretischen Modelle auf eine Aussage über die situationale Handlungsfähigkeit. Auch das Konstrukt der Meta-Kompetenz (Weinert, 1999; 2001), befasst sich mit Strategien zur Organisation kognitiver Dispositionen in Anforderungssituationen. Das Kompetenzparadigma bezieht sich über den Situationsbezug stärker auf alltagszeitliche Konstrukte und stellt lebenszeitliche Konstrukte, wie die biografische Handlungsfähigkeit von Personen (Böhnisch et al., 2009), eher zurück. Da im lebensweltlichen Handeln von Personen jedoch auch situationale Entscheidungen vor dem Hintergrund und im Hinblick auf biografische Verläufe getroffen werden, ist die Reichweite des Kompetenzparadigmas kürzer, als eine langfristig ausgelegte Steuerung des Bildungssystems benötigt (Herzberg & Truschkat, 2009).

Der funktionalistische Kompetenzbegriff und die kognitive Modellierung zielen auf die Vorhersage von Handlungsfähigkeit und Leistung in Bezug auf definierte Anforderungen. Damit geraten jedoch nicht intendierte, aber subjektiv wertvolle und sinnhafte Bildungsprozesse und -ergebnisse aus dem Blickfeld. Zahlreiche Studien der Biografieforschung machen darauf aufmerksam, dass Lern- und Entwicklungsprozesse ein so konstruktives Geschehen sind, dass eine perspektivische Beschränkung auf normativ definierte Anforderungssituationen (und eine entsprechende Steuerung des Bildungssystems) die stärksten Bildungseffekte verpassen kann (zum Überblick siehe Lempp, 2011). Wenngleich Biografizität (Alheit, 1995) im Sinne biografischer Kompetenz (Schlüter, 2010) in der erziehungswissenschaftlichen Biografieforschung als Konstrukt angewendet wird, greift der quantitativ orientierte Kompetenzdiskurs diese Dimension bislang kaum auf.

Der Leistungsbegriff von funktionalistischen Kompetenzmodellen bezieht sich konzeptionell durchgängig auf erfolgreich bewältigte Situationen, d. h. auf ein Handlungsergebnis bzw. ein positives Outcome. Daraus leitet sich die Frage ab, wie intensives selbstgesteuertes, aber nicht erfolgreiches Lern- und Bewältigungsverhalten im Kontext von Leistung mit dem Kompetenzbegriff bewertet werden kann. Während in sozialisations- und biografiethoretisch angelegter Forschung die Produktivität des Scheiterns (Rödel, 2018; Villa, 2011) bereits umfangreich herausgearbeitet wurde, steht diese Diskussion für das Kompetenzparadigma noch aus. Auf diese Schwierigkeit kriterienorientierter Leistungsbewertung hat bereits Rheinberg (2001) in der Gegenüberstellung individueller, sozialer und kriterialer Vergleichsnormen hingewiesen. Der Diskurs um Chancengerechtigkeit zeigt jedoch, dass dieser Kritikpunkt weit in die gesellschaftliche Dimension hineinreicht (Weber, Danninger, & Feyerer, 2016). Die soziale Herkunft hat trotz entwicklungsförderlicher Lernumwelten von der Grundschule bis in akademische Berufe hinein einen starken Effekt auf Bildungserträge und -entscheidungen (Blossfeld, Blossfeld, & Blossfeld, 2015; Bos, Wendt, Köller, & Selter, 2012; Erikson, 2016; Reiss, Sälzer, Schiepe-Tiska, Klieme, & Köller, 2016).

Ein Großteil der Kritik bezieht sich auf ältere Kompetenzbegriffe, da diese Einschränkungen und Schwierigkeiten im Kompetenzparadigma bereits explizit reflektiert aber, noch nicht gelöst wurden (Klieme, Maag-Merki, & Hartig, 2007).

Aus methodologischer Perspektive ist das Verhältnis von Kompetenz und Performanz vielfach hinterfragt worden. Wood und Power (1987) wenden die Messfehlertheorie auf das Verhältnis an und stellen bei einem Indikatorverhältnis von Performanz und Kompetenz zwei Risiken heraus. Zum einen besteht bei ‚holistisch‘ angelegten Messverfahren ein erhöhtes Risiko falsch positiver Ergebnisse. Je komplexer die Anforderungssituation gestaltet ist, desto höher ist auch die Wahrscheinlichkeit, dass verschiedene Konfigurationen kognitiver Module und Prozesse zu einer mehr oder weniger erfolgreichen Leistung befähigen. In der Forschung zur Lehrkräftebildung zeigen etwa Santagata und Sandholtz (2018) in einer Studie mit Studierenden des Mathematiklehramts, dass zwei unterschiedliche, praxisnahe Testverfahren zwar jeweils eindeutige und starke Zusammenhänge zu kognitiven Dispositionen haben, jedoch zu sehr unterschiedlichen Ergebnissen hinsichtlich der Einschätzung von Kompetenz führen. Zum anderen besteht bei der Annahme eines Indikatorverhältnisses ein erhöhtes Risiko falsch negativer Ergebnisse, da unklar ist, wie viele Testsituationen nötig oder erlaubt sind, um Kompetenz abzubilden. Diese Kritik trifft jedoch nur generalisierte Kompetenzbegriffe, da funktionalistische Konzeptionen die Anforderungssituationen

explizit definieren. Ein weiterer Kritikpunkt von Wood und Power (1987) ist die Erschließung von Kompetenzstrukturen mit Hilfe von Personen, die nicht das Niveau von Expertise erreicht haben, da zu diesem Entwicklungspunkt noch kein konsistentes Testverhalten vorliegt. Dies ist jedoch eine Frage von Untersuchungsdesigns und lässt sich gegenüber der aktuellen Forschung nicht mehr in dem Maße kritisieren. Darüber hinaus entzerren neuere Definitionsansätze (s. o.) die Begriffe der Performanz und Kompetenz in Richtung unterschiedlicher Konstrukte, für die es jeweils geeignetere Indikatoren gibt, als das jeweils andere Konstrukt.

Ein weiterer Kritikpunkt in Bezug auf die Operationalisierung von Kompetenz ist die Begrenzung von Kompetenz auf individuelle Leistungsdispositionen. Sobald etwa eine Anforderungssituation auf den Aufbau sozialer Beziehungen zielt, genügt eine individuelle Disposition nicht, um handlungsfähig zu sein. Der Umgang mit Angewiesenheit ist zwar ein zentraler Bestandteil von entwicklungspsychologischen Theorien und Sozialisationstheorien, wird jedoch im Kompetenzparadigma größtenteils ausgeblendet. Eine Rezeptionsschwelle für die Integration dieser theoretischen Dimension wird in der statistischen Modellierung gesehen, in der soziale Gruppen als Additiv einzelner Personen betrachtet werden, ohne die besonderen Eigenschaften sozialer Gruppen einzubeziehen. Weinert (1999, S. 11) löst diese „individualized „competence barrier““ mit dem funktionalistischen Kompetenzbegriff bewusst nicht auf, sondern akzeptiert sie als Einschränkung des Konstrukts. In der Schulforschung und besonders im Bereich der Diagnostik von Lernvoraussetzungen wird eine kollektive Dimension von Kompetenz jedoch gegenüber der Verortung interaktiver Konstellationen als Kontextvariable weiterhin diskutiert (Santagatah & Yeh, 2016; Stahnke, Schueler, & Roesken-Winter, 2016).

2.4 Zusammenfassung des zweiten Kapitels

Kompetenz ist ein theoretisches Konstrukt, das Ergebnisse von Lern- und Entwicklungsprozessen auf struktureller Ebene verortet. Im Fokus stehen nicht die Aneignung konkreter inhaltlicher Wissensbestände und auf spezifische Aufgaben bezogene Fähigkeiten, sondern der Erwerb einer kognitiven Struktur, mit der Personen in unterschiedlichen Anforderungssituationen einer bestimmten Domäne handlungsfähig sind. Die Expertiseforschung zeigt, dass eine solche Kapazität weitestgehend unabhängig von einerseits stabilen Persönlichkeitsmerkmalen wie Intelligenz oder Verarbeitungsgeschwindigkeit und andererseits von hochgradig spezialisierten Fähigkeiten existiert und durch akkumulierte Lern- und Entwicklungsprozesse erwerbbar ist (Dreyfus & Dreyfus, 2004).

Die vorliegende Arbeit bezieht sich größtenteils auf den funktionalistischen Ansatz und bündelt die gewonnenen theoretischen Erkenntnisse in einer neuen Definition von Kompetenz im State-Trait Paradigma. Kompetenz bezeichnet die in formaler Bildung erlernbaren funktionellen Konfigurationen kognitiver Strukturen und Module zur Erbringung kognitiver und verhaltensbezogener Leistung in kontextspezifischen Anforderungssituationen. Dagegen bezeichnet Performanz die in Anforderungssituationen gezeigte und kriterial messbare Leistung. Angenommen wird eine probabilistische Beziehung von Kompetenz und Performanz. Dementsprechend kann Performanz ausschließlich unter strengen Experimentalbedingungen ein Indikator für Kompetenz sein. Kompetenz wird als State (Zustand) konzipiert, der gleichfalls in einer probabilistischen Beziehung zu den

kognitiven Strukturen und Modellen, d. h. Dispositionen steht. Letztere sind als Trait (Eigenschaft) konzipiert. Unterschieden wird je nach Anforderungssituation weiterhin zwischen Kompetenz als individuellem Merkmal und als Merkmal einer sozialen Gruppe. Letztere beschreibt ähnlich wie auf individueller Ebene nicht lediglich die Addition einzelner kompetenzbildender Bestandteile, sondern ihr funktionelles Zusammenwirken. Voraussetzungen zur Definition kontextspezifischer Kompetenz bestehen in der Beschreibung von Anforderungssituationen und den konkreten kognitiven und sozialen Strukturen und Prozessen, die zu Leistung in diesen Situationen befähigen. Inwiefern es sich bei den kognitiven Strukturen und Prozessen bspw. um wissensbezogene Gedächtnisinhalte, reizspezifische Verarbeitungsmodi oder motorische Routinen wie Blickbewegungen handelt und bei den sozialen Strukturen und Prozessen um wertbezogene Einstellungskonstellationen oder problemspezifische Kommunikations- und Entscheidungsabläufe, muss für jeden Kontext differenziert beschrieben werden. Motivationale Prozesse können zur Aktivierung der funktionellen kognitiven Struktur beitragen und sollten daher zur exakten Bestimmung eines Kompetenzniveaus kontrolliert werden.

Der Bezug auf *funktionelle* Konfigurationen fordert von Beschreibungen einer konkreten Kompetenz ein, sowohl die an anforderungsbezogener Leistung beteiligten kognitiven Strukturen und Module zu benennen als auch die Mechanismen ihres Zusammenwirkens. Damit besteht ein Anknüpfungspunkt zu kognitionspsychologischen Theorien der Informationsverarbeitung, wie etwa zu Modellen des Arbeitsgedächtnisses, zu Aufmerksamkeitsprozessen und zu Zwei-Prozess-Modellen (Chaiken & Trope, 1999). Dieses Definitionsmerkmal erleichtert auch den von Klieme und Leutner (2006) geforderten Anschluss an die Expertiseforschung, die bspw. die zielbezogene Anwendung systematischer und heuristischer Informationsverarbeitung neben erfolgreicher Aufgabenbewältigung als Kriterium zur Leistungsmessung verwendet (Swanson, O'Connor, & Cooney, 1990). Diese Definitionserweiterung hat zur Folge, dass Kompetenz nicht mehr als ein Konstrukt verstanden werden kann, dessen Ausprägung *vollständig* unabhängig von kognitiven Basisfunktionen ist. Zugespitzt kann sich zur Veranschaulichung die klinisch-psychologische Frage gestellt werden, in welchem Ausmaß Personen mit gravierenden Funktionsstörungen allgemeiner kognitiver Strukturen (z. B. semantische Gedächtnisstörungen wie demenzielle Erkrankungen) alltäglichen beruflichen Handlungsanforderungen vollumfänglich ausgesetzt sein sollten und in diesen kompetent handeln können (Henry, Merten, & Wallasch, 2008). Es ist plausibler anzunehmen, dass sich Kompetenz nicht in kontextspezifischen kognitiven Modulen und Strukturen erschöpft, sondern darüber hinaus ein Aktivierungsmusters dieser *im Rahmen allgemeiner kognitiver Strukturen und Module* darstellt. Dementsprechend wird in der vorliegenden Arbeit sowohl Forschung zu ausgewählten kognitiven Strukturen als auch zu funktionellen Prozessen diskutiert, die zu Leistung in diagnostischen Anforderungssituationen befähigen.

3 Diagnostische Kompetenz im professionellen Handeln von Lehrkräften

Diagnostizieren ist ein Bestandteil verschiedener Professionen, wie etwa der Medizin und Psychologie, in der bereits eine lange gemeinsame Forschungstradition zum Diagnostizieren als Voraussetzung klinischer Praxis besteht (im Überblick: Elstein, 2016; Shulman & Elstein, 1975). Eine Erklärung des besonderen Stellenwerts von Diagnostik für professionelles Handeln auf gesellschaftlicher Ebene leistet Oevermann (1996; 2008) in seinem strukturtheoretischen Ansatz der Professionalisierung.

Oevermann (1996) geht davon aus, die Herstellung von Autonomie im Lebenszusammenhang von Personen ein grundlegender Wert moderner Gesellschaften ist. Die selbstständige Gestaltung des Lebens gerät jedoch immer wieder an gesellschaftliche oder biologische Grenzen. Die Erfahrung einer solchen Begrenzung ist soweit unproblematisch, bis sie ein solches Maß an negativem Stress erzeugt, dass das Stresserleben eine autonome Lebensgestaltung im Rahmen der gesellschaftlichen Möglichkeiten einschränkt. Oevermann (1996) bezeichnet diesen Zustand als Krise, d. h. als Situation, in der individuell erworbene Handlungsroutinen zur Wiederherstellung einer autonomen Lebenspraxis nicht greifen oder nicht zur Verfügung stehen. Diese Diskrepanz von eigener Lebenspraxis auf der einen Seite und gesellschaftlichen Anforderungen sowie zugehörigen kulturell sedimentierten Bewältigungsmöglichkeiten auf der anderen Seite ist nach Oevermann (1996) zwar eine Voraussetzung für das Erleben von Autonomie. Diese kann jedoch nur dann als solche erlebt und realisiert werden, wenn sich tatsächlich neue Bewältigungsmöglichkeiten auftun. In Fällen, in denen dieser Zustand nicht erreicht wird, greift professionelle Praxis. Diese versteht Oevermann (1996) als stellvertretende Deutung bzw. stellvertretende Krisenbewältigung mit dem Ziel der Wiederherstellung und Stärkung des persönlichen Autonomiepotenzials. Entscheidungen für eine andere Person zu treffen geht jedoch mit einer Explikations- und Begründungsverpflichtung einher, die professionell Handelnde einerseits durch generalisiertes wissenschaftliches Wissen und andererseits durch fallbezogene Rekonstruktionen von Deutungs- und Lösungsangeboten einlösen. Darüber hinaus ist die Wiedererlangung von Autonomie nur dann möglich, wenn professionell Handelnde die geplanten Lösungsschritte nicht selbst durchführen, sondern zu deren Umsetzung innerhalb eines Arbeitsbündnisses befähigen (Oevermann, 2008).

Um Diagnostik als Bestandteil professionellen Handelns von Lehrkräften theoretisch zu beschreiben, ist also eine Perspektive gefragt, die einen spezifischen Ausschnitt der stellvertretenden Deutung und Krisenbewältigung von Lehrkräften für Schülerinnen und Schüler umfasst. Diagnostizieren im Sinne professionellen Handelns stellt die strukturelle Anforderung an Lehrkräfte, ihre pädagogischen und didaktischen Entscheidungen einerseits von generalisiertem Regelwissen und fallbezogenem Wissen über Schülerinnen und Schüler abzuleiten und andererseits die darauf basierenden Handlungspläne in einer gemeinsamen bzw. geteilten Handlungspraxis umzusetzen.

Aus dem gemeinsamen Forschungsfeld psychologischer und medizinischer Diagnostik heraus wurden bereits allgemeine Theorien zu Diagnostik entwickelt, die zunächst von hypothetisch-deduktiven Regelmodellen ausgingen (Matarazzo, 1983; Newell & Simon, 1972; Reiter, 1987), anschließend zu kontextspezifischen Modellen heuristischer und systematischer Informationsverarbeitung spezifiziert wurden (Brush, Sherbino, & Norman, 2017; Croskerry, 2017;

Norman et al., 2017; Monteiro & Norman, 2013; Nystrom, Williams, Paull, Graber, & Hemphill, 2016), um den sozialen Kontext des Diagnostizierens, anschließende Interventionspraxis und Kompetenzentwicklung erweitert wurden (Cook, Sherbino, & Durning, 2018; Ilgen et al., 2012; Lambe, O'Reilly, Kelly, & Curristan, 2016) und schließlich an neuropsychologische Forschung angeschlossen wurden (Scholz, Krems, & Jahn, 2017; Rotgans et al., 2019). Ein Transfer dieser Forschungstradition in die Unterrichts- und Expertiseforschung zu Lehrkräften sowie in die Lehrkräftebildungsforschung findet überraschender Weise kaum statt (z. B. Herppich et al., 2018; Klug, Bruder, & Schmitz, 2016; Schrader, 2013; Ausnahmen: Leuders, Dörfler, Leuders, & Philpp, 2018; Shulman & Elstein, 1975). Shulman (1981) beobachtete diese fehlende Bezugnahme bereits in den 80er Jahren und stellte fest, „Educational researchers are typically eager to distinguish their work from other forms of discourse which, for them, cannot lay claim to being research“ (S. 5). Dennoch lässt sich eine stetig zunehmende Bedeutung von Diagnostik in der jüngeren Geschichte der Ausbildung von Lehrerinnen und Lehrern nachzeichnen.

Im nationalen Kontext gewann die systematische institutionelle Schulung von Lehrkräften im Bereich Diagnostik verstärkt mit der geforderten Professionalisierung der Lehrenden- und Lehrerbildung und der Neuausrichtung und -verortung des sonderpädagogischen Bereichs in den siebziger Jahren an programmatischer Bedeutung. Der 1970 vom Deutschen Bildungsrat erstellte Strukturplan für das Bildungswesen sah für Lehrkräfte die Ausbildungsbereiche Lehren, Erziehen, Beurteilen, Beraten und Innovieren vor. Zwar fand das gesamte Sonderschulwesen in diesem Plan keine Berücksichtigung, es schloss sich jedoch die Gründung des Fachausschusses Sonderpädagogik durch den Bildungsrat an, der sowohl in die folgende Empfehlung der Kultusministerkonferenz (KMK) „Zur Ordnung des Schulwesens“ von 1972, in die Empfehlung des Bildungsrates „Zur pädagogischen Förderung behinderter und von Behinderung bedrohter Kinder und Jugendlicher“ von 1973 sowie in den Bildungsgesamtplan der Bund-Länder-Kommission von 1973 den Begriff der „Pädagogischen Diagnostik“ einbrachte und das Ziel eines inklusiven Unterrichts ohne Trennung von Schultypen verfolgte:

Zahlreiche Wissenschaften wie beispielsweise Soziologie, Pädagogik, Psychologie, Nationalökonomie müssen in der allgemeinen Bildungsforschung interdisziplinär zusammenwirken. Dadurch werden die Grundlagen geschaffen, auf denen einzelne Bereiche der Bildungsforschung (Curriculumforschung, Pädagogische Diagnostik, Bildungstechnologie, Bildungsökonomie) Beiträge für die Entwicklung neuer Strukturen und Inhalte des Bildungswesens leisten können. (S. 42)

Pädagogische Diagnostik wurde im Bildungsgesamtplan von 1973 wie folgt definiert:

Unter pädagogischer Diagnostik werden alle Maßnahmen zur Aufhellung von Problemen und Prozessen sowie zur Messung des Lehr- und Lernerfolges und der Bildungsmöglichkeiten des einzelnen im pädagogischen Bereich verstanden, insbesondere solche, die der individuellen Entscheidung über die Wahl der anzustrebenden Qualifikationen der Schullaufbahn, des Ausbildungsganges im tertiären Bereich und der Berufsausbildung sowie der Weiterbildung dienen. (S. 45)

Neben dem Fokus auf Übergangsempfehlungen bestanden weitere Ziele der Kommission in der Entwicklung von Messinstrumenten zur Erfassung von „Lerndispositionen“ mit dem Zweck der Bildungsberatung, sowie der objektiven Leistungsmessung mit dem Zweck innerer und äußerer Differenzierung. Der Bildungsgesamtplan sah zur Umsetzung dieser Ziele in der Lehrkräftebildung die Einrichtung einer „Abteilung pädagogische Diagnostik“ in der ebenfalls einzurichtenden „Zentralstelle zur Förderung und Dokumentation der Curriculumforschung und -entwicklung“ vor. Deren Finanzierung konnte in der Auseinandersetzung von Regierungskoalition und Opposition jedoch nicht beschlossen werden (zur soziologischen Ursachenforschung siehe Naumann, 1974). Stattdessen wurde eine Kommission „Perspektiven für die Lehrerbildung in Deutschland“ von der KMK eingesetzt und mit der Entwicklung von Standards für die Lehrkräftebildung beauftragt (KMK, 2004). In der letztendlichen Ausformulierung der Standards schließt sich die KMK einem Expertengutachten von Terhart (2002) an. Die Dokumentation der jeweiligen institutionellen Kompetenzdefinition für die Lehrkräftebildung erfolgt letztlich in den Universitäten und richtet sich nach Empfehlungen der KMK. Neben diesen institutionalisierten Professionalisierungsschritten stieg auch die Anzahl verfügbarer Testinstrumente zur Feststellung der Lern- und Leistungsmerkmale von Schülerinnen und Schülern wie auch von Studierenden (Ingenkamp & Lissmann, 2008).

Einen grundlegenden wissenschaftlicher Impuls zur Aufnahme diagnostischen Handelns in professionelle Standards gab Shulman (1987) mit einer konzeptuellen Strukturierung des Professionswissens von Lehrkräften. Während die Kategorien pädagogisches Wissen, Fachwissen und fachdidaktisches Wissen auch weiterhin in empirisch gestützten theoretischen Modellen Verwendung finden (Baumert & Kunter, 2006; Kunter, Baumert, & Blum, 2011), sah Shulman (1987, S. 8) vier weitere Dimensionen des Professionswissens vor: „curricular knowledge [...], knowledge of learners and their characteristics [...], knowledge of educational contexts [...], knowledge of educational ends, purposes, and values, and their philosophical and historical grounds“, die mittlerweile unter die anderen drei Facetten subsumiert wurden. Auf Shulman (1987) und eigenen empirischen Untersuchungen aufbauend, schlugen Weinert, Schrader und Helmke (1990, S. 172) eine Unterteilung des Professionswissens in „knowledge of classroom management [...], knowledge of instructional techniques [...], subject-matter knowledge and diagnostic knowledge“ vor. Dieser Vorschlag wurde von Oser (1997) aus einer praxisorientierten Perspektive heraus ausdifferenziert, wobei er eine Gruppe von professionellen Standards mit dem Inhalt schülerunterstützendes Handeln und Diagnose beibehielt. Terhart (2002) erstellte darauf bezogen ein Gutachten, das einen stärkeren Wissens- und Kompetenzbezug einforderte, aber weiterhin auch Standards zu Lerndiagnostik sowie Beurteilung, Diagnose und Förderung enthielt. An diesem orientierte sich letztlich auch die KMK (2004) in der Veröffentlichung der aktuell gültigen Bildungsstandards, in der diagnostisches Handeln weiterhin enthalten ist, bspw. „Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern“ (S. 8).

Dieses Kapitel diskutiert theoretische Definitionen diagnostische Kompetenz, Modelle diagnostischen Handelns sowie empirischer Befunde zu diagnostischer Kompetenz von Lehrkräften. Dabei wird auch Forschung zu Diagnostik in anderen Professionen einbezogen. Das Kapitel mündet in einer Definition diagnostischer Kompetenz, auf der die Forschung der vorliegenden Arbeit basiert.

3.1 Definition und Modellierung diagnostischer Kompetenz

Zur Definition diagnostischer Kompetenz von Lehrkräften existieren verschiedene Ansätze, wie etwa die disziplinspezifische Einteilung in fachdidaktische, pädagogische und sonderpädagogische Diagnostik (Füchter, 2011), die strukturelle Einteilung in produkt- und prozessbezogene Diagnosemodelle (Südkamp & Praetorius, 2017), die methodologische Einteilung in formelle und informelle Diagnostik (Schrader, 2013), die zielbezogene Einteilung in pädagogische Selektions- und Modifikationsdiagnostik (Ingenkamp & Lissmann, 2008) sowie die zeitbezogene Einteilung in Status-, Prozess-, Verlaufs- und Veränderungsdiagnostik (von Aufschnaiter et al., 2015). Die Pluralität der Definitionsansätze geht mit einer entsprechenden Vielfalt an Operationalisierungen in Modellen und Konstrukten einher (von Aufschnaiter et al., 2015), so dass eine gemeinsame Darstellung von Definitionen, zugehörigen Modellen und zugehörigen empirischen Befunden erfolgt. Die o. g. Systematiken unterteilen Modelle diagnostischer Kompetenz entlang eines ausgewählten Ausschnitts der diagnostischen Praxis von Lehrkräften. Diese Art *beschreibender* Systematisierungen bietet jedoch keine *Erklärung* für die jeweils unterschiedlichen Perspektiven auf diesen Ausschnitt an. Dieser Umstand erschwert die Reflexion von Vor- und Nachteilen der jeweiligen Modelle, die auf grundlegende theoretische Positionen und Perspektiven zurückzuführen sind. Daher erfolgt in diesem Kapitel eine Einteilung der Definitionen und Modelle diagnostischer Kompetenz vor dem Hintergrund ihrer meta-theoretischen Perspektive. Die bestehende Vielfalt an Definitionen wird dabei weniger als zu lösendes Problem und mehr als Bereicherung für den Bereich der Diagnostik im schulischen Kontext gesehen.

Kognitive Ansätze

Der Gegenstand kognitiver Theorien sind individuelle Denk- und Verarbeitungsprozesse sowie deren Entwicklung, wie etwa in der Gedächtnistheorie von Baddeley (1992). Da kognitive Theorien auf die Erklärung intra-individueller Strukturen, Prozesse und Ergebnisse der Informationsverarbeitung zielen, spielt die Modellierung vor- und nachgängiger Strukturen sowie Veränderungen in der Umwelt nur eine geringfügige Rolle. So haben etwa große Teile der Expertiseforschung die Entwicklung kognitiver Modelle zum Gegenstand, die zur Identifikation möglichst effizienter Strategien der Informationsverarbeitung genutzt werden. Für die Diagnostik besteht dementsprechend Erkenntnispotenzial in der Identifikation kognitiver Module und Strukturen, die inter-individuell konsistent im professionellen Handeln genutzt werden. Umgesetzt in psychometrische Modelle können so auch Kriterien für Kompetenzniveaus bestimmt werden. Eine zentrale Rolle spielt daher die Modellierung des Produkts und der Prozesse in der diagnostischen Informationsverarbeitung.

Urteilsgenauigkeit als Produkt individueller Informationsverarbeitung. Coladarci (1986) ging in Anschluss an Hoge (1983) in einer der ersten Untersuchungen zu Urteilsgenauigkeit von Lehrkräften davon aus, dass Entscheidungen in der Unterrichtsplanung zu einem großen Teil auf Einschätzungen der kognitiven Fähigkeiten von Schülerinnen und Schülern basieren. Wenngleich er explizit anmerkt,

dass allein aus den Ergebnissen zur Urteilsgenauigkeit keine direkten Implikationen für Modelle der Entscheidungsbildung und Kognition von Lehrkräften gezogen werden sollten (Coladarci, 1986: 145), setzen Helmke und Schrader (1987) die Urteilsgenauigkeit begrifflich (jedoch nicht methodisch) mit der Sensitivität einer Diagnose und mit diagnostischer Kompetenz gleich (Helmke & Schrader, 1987, Abb. 3). Schrader (1989) unterscheidet drei Dimensionen diagnostischer Urteile: personenbezogene, aufgabenbezogene und aufgabenspezifische Urteile. Personenbezogene Urteile betreffen die Einschätzung einer Schülerin oder eines Schülers hinsichtlich eines Merkmals (z. B. Mathematikfähigkeit). Das aufgabenbezogene Urteil bezieht sich auf die Einschätzung der Lösungshäufigkeit einer Aufgabe in einer Schulklasse, d. h. die gruppenbezogene Itemschwierigkeit. Die aufgabenspezifische Beurteilung fasst diese beiden Dimensionen zusammen, gefragt ist die Einschätzung zur Lösung einer bestimmten Aufgabe durch eine bestimmte Schülerin oder einen bestimmten Schüler.

Innerhalb der ersten beiden Dimensionen differenziert Schrader (1989) drei Komponenten eines genauen Urteils aus: die Rangordnungs-, Niveau-, und Differenzierungskomponente. Die Rangordnungskomponente gibt an, wie genau eine Lehrkraft die Rangreihenfolge eines einzuschätzenden Merkmals in einer Schulklasse vorhersagt. Die Niveauelemente beschreibt die korrekte Einschätzung des mittleren Leistungsniveaus der Klasse, gemessen in der Differenz aus Merkmalausprägungen und Lehrkräfteurteilen. Sie drückt daher auch Tendenzen zur Über- oder Unterschätzung aus. Die Differenzierungskomponente bezieht sich auf die Einschätzung der Streuung eines Merkmals in einer Schulklasse, gemessen im Quotienten aus der Streuung der Lehrkräfteurteile und der Streuung der Merkmalsausprägungen. Untereinander korrelierten die Genauigkeitskomponenten in der Studie von Schrader (1989) und folgenden Untersuchungen nur sehr gering (Schrader, Helmke, Hosenfeld, Halt, & Hochweber, 2006; Karst 2012). In Folge dessen wird häufig lediglich die Rangordnungskomponente als Indikator für die Urteilsgenauigkeit verwendet, jedoch alle drei Komponenten zu Vergleichszwecken berichtet (Helmke, Hosenfeld, & Schrader, 2004; Karst, 2012; Schrader, 1989; Praetorius, Lipowsky, & Karst, 2012).

Schrader (2009, S. 237) definiert vor diesem Hintergrund diagnostische Kompetenz als „Fähigkeit von Lehrkräften, Schülerinnen und Schüler sowie lern- und leistungsrelevante Sachverhalte zutreffend zu beurteilen“. Zwar benennt Schrader (2008, S. 168) auch Voraussetzungen genauer Urteile als „das damit zusammenhängende Wissen sowie die Kenntnis und Beherrschung dazu geeigneter diagnostischer Methoden“, differenziert diese jedoch nicht weiter aus. Es bleibt daher unklar, ob mit Kompetenz ein inter-individuell stabiles Merkmal gemeint ist oder ob von einem situativ-holistischen Kompetenzverständnis ausgegangen wird, in dem sich die konkreten Voraussetzungen für Handlungsfähigkeit je nach Anforderungssituation inter-individuell oder auch intra-individuell unterscheiden können. In dieser Form ist diagnostische Kompetenz auch im COACTIV-Modell definiert, auf dem zahlreiche Studien in der Lehrkräftebildungsforschung basieren (Kunter, Baumert, Blum, Klusmann, Krauss, & Neubrand, 2013).

Die Kritik an der Definition und Modellierung diagnostischer Kompetenz über Urteilsgenauigkeit ist bereits sehr umfangreich und umfasst theoretische, methodologische sowie empirische Argumente. Auf theoretischer Ebene merkt Abs (2007) an, dass eine genaue Beurteilung allein keine Ableitung pädagogischer oder didaktischer Anschlusshandlungen zulässt und daher die Relevanz von

Urteilsgenauigkeit für das Konstrukt der diagnostischen Kompetenz unklar ist. van Ophuysen (2010) wirft daran anschließend auf, dass diagnostisches Handeln auch dem Urteil vor- und nachgelagerte Aspekte umfasst. Dieses Argument lässt sich dahingehend weiterführen, dass neben einer mehr oder weniger finalen diagnostischen Einschätzung andere Faktoren existieren, die sowohl den diagnostischen Prozess als auch eventuelle Anschlusshandlungen anleiten (z. B. Wertvorstellungen oder verfügbare Ressourcen). Weiterhin müssen Informationen über lern- und leistungsrelevante Merkmale von Schülerinnen und Schülern nicht zwangsläufig für ein personenbezogenes Urteil genutzt werden, sondern können bspw. auch die Entwicklung von Unterrichtsaufgaben informieren. Klug, Bruder und Schmitz (2016) schätzen die Modellierung durch Urteilsgenauigkeit in zugespitzter Form als methodologisches Artefakt ein, dessen Relevanz eher im Wissenschafts- als im Bildungssystem bestehe.

Karing (2009) sowie Südkamp und Möller (2009) führen weiterhin an, dass Unklarheit über Standards und gerechtfertigte Ansprüche besteht, da sich Beurteilungssituationen in der Berufspraxis nicht ausreichend steuern lassen. Eine große Heterogenität in der Schulklasse vereinfacht bspw. die Beurteilung durch Lehrkräfte (Paleczek, Seifert, & Gasteiger-Klicpera, 2017). Auch Artelt (2016) weist darauf hin, dass genaue Urteile nicht zwangsläufig erwartet werden können: „Teachers should not be expected to deliver high quality and accurate judgments independent of the goals, demands, and standards placed on their judgements“ (S. 4). Im Kern zielt die theoretische Kritik darauf, dass die Genauigkeit eines diagnostischen Urteils eine mögliche Outcome-Variable einer ausgewählten Performanzsituation darstellt, ihr jedoch kein theoretisches Modell diagnostischer Kompetenz zugrunde liegt. Es liegt mit den Komponenten von Helmke und Schrader (1987) zwar ein Vorschlag zur Struktur der Urteilsgenauigkeit vor, diese darf jedoch nicht mit der Struktur der dazu notwendigen Kompetenz verwechselt werden. Schrader (2008) deutet zwar Dispositionen im Sinne notwendiger kognitiver Module an, differenziert diese jedoch nicht systematisch aus. Da auch keine standardisierten Kriterien zur Bewertung der gezeigten Beurteilungsgenauigkeit vorliegen, bleiben die Anforderungssituation und mögliche Kompetenzniveaus ebenfalls unterdefiniert.

Studien zu Urteilsgenauigkeit von Lehrkräften und Lehramtsstudierenden weisen oftmals Merkmale auf, die auch aus methodologischer Sicht kritisch zu reflektieren sind. Lehrkräften werden im Großteil der bisher durchgeführten Studien die einzuschätzenden Merkmale vorgegeben (beispielsweise Karing, 2009; Spinath, 2005; Schrader, 1989). Diese Erhebungsmethode scheint zunächst notwendig und im Hinblick auf die Datenauswertung vielleicht auch ökonomisch, klammert aber eine ganze Anzahl kognitiver und sozialer Prozesse aus, die im Vorgang diagnostischen Handelns und der Urteilsbildung eine zentrale Rolle spielen (z. B. Wahrnehmung und Auswahl eines einzuschätzenden Merkmals in Bezug zu einer Unterrichtssituation oder einem Entwicklungsverlauf). Die Validität der Ergebnisse im Hinblick auf diagnostische Kompetenz ist dadurch maßgeblich eingeschränkt. Insbesondere, wenn diagnostisches Handeln den Anspruch hat, über ad hoc Einschätzungen ohne weiterführende Zielstellung hinauszugehen, können spontane Urteile über vorgegebene Merkmale nur sehr eingeschränkt ein Indikator für Kompetenz sein. Weiterhin stellt Artelt (2016) fest, dass Lehrkräften in einer größeren Anzahl an Studien zur Einschätzung von Merkmalen andere Kriterien vorliegen als den Schülerinnen und Schülern (z. B. anderes Instrument oder alternative Itemformulierung für dasselbe Konstrukt). Darüber hinaus bleibe oft unklar bzw.

undefiniert, welcher Zeitraum und welche Situationen der Einschätzung zugrunde liegen sollen (beispielsweise Artelt & Karing, 2014; Helmke, Hosenfeld, & Schrader, 2004; Praetorius, Greb, Lipowsky, & Gollwitzer, 2010; Stang & Urhahne, 2016; Spinath, 2005).

Aber auch empirische Forschung spricht gegen eine exhaustive Modellierung diagnostischer Kompetenz von Lehrkräften durch die Urteilsgenauigkeit. Spinath (2005) stellt in einer empirischen Untersuchung zur Genauigkeit der Beurteilung fachbezogener Schüler- und Schülerinnenfähigkeiten sowie deren Motivation und Selbstkonzept fest, dass weder zwischen den Genauigkeitskomponenten noch zwischen der Urteilsgenauigkeit verschiedener Merkmale ein substanzieller Zusammenhang besteht. Sie empfiehlt daran anschließend von der begrifflichen Fassung von Urteilsgenauigkeit als diagnostische Kompetenz abzusehen. Zu einem ähnlichen Ergebnis kommen Praetorius, Greb, Lipowsky und Gollwitzer (2010) wie auch Praetorius, Karst, Dickhäuser und Lipowsky (2011). Meta-Analysen zeigen für kognitive fachspezifische Fähigkeiten mittlere bis hohe durchschnittliche Korrelationen zwischen Lehrkräfteurteil und Schülerinnen- oder Schülermerkmal $.43 < r < .63$ (Hoge & Coladarci, 1989; Südkamp, Kaiser, & Möller, 2012; Machts, Kaiser, Schmidt & Möller, 2016). Dagegen liegen die Werte für affektiv-motivationale Konstrukte Lehrkräften deutlich niedriger, bspw. für Prüfungsangst. Boehnke, Silbereisen, Reynolds und Richmond (1986) finden für Prüfungsangst eine durchschnittliche Korrelation von $r = .21$. Spinath (2005) berichtet $r = .15$, Zhou und Urhahne (2013) $r = .21$, Artelt und Karing (2014) finden je nach Angstkomponente $.21 < r < .46$ und Urhahne (2015) berichtet $r = .17$. Diese Ergebnisse sind ein Hinweis darauf, dass Urteile von Lehrkräften gegenstandsspezifisch sind. Kompetenz ist jedoch als Konstrukt definiert, das über inhaltlich verschiedene aber strukturell ähnliche Anforderungssituationen hinweg stabil ausgeprägt sein sollte. Unter dieser Annahme wäre sodann der Schluss nötig, dass Urteilsgenauigkeit kein Indikator für diagnostische Kompetenz sein kann. Wenn dagegen zu jedem einzuschätzenden Merkmal eine eigene Kompetenz existierte (z. B. „Prüfungsangstdiagnosekompetenz“), wäre das Kompetenzkonstrukt überflüssig und es könnte auf inhaltspezifische Konstrukte (z. B. Fähigkeiten) zurückgegriffen werden.

Weiterhin bestehen zwischen Lehrkräften unter Kontrolle von Alter, Geschlecht und Berufserfahrung unerwartet große Unterschiede in ihrer Urteilsgenauigkeit. In der Meta-Analyse von Südkamp et al. (2012) schwankt die Korrelation von Lehrkräfteurteil und Schülerinnen- oder Schülermerkmal in Bezug auf fachspezifische Fähigkeiten zwischen $-.18 < r < .72$. In der Studie von Urhahne (2015) schwankt die Korrelation für Prüfungsangst ebenfalls $-.13 < r < .47$. Zur Erklärung dieser Unterschiede wurden verschiedene Prädiktoren von Urteilsgenauigkeit auf den Ebenen der Lehrkräfte (z.B. Wissen, Motivation und Fähigkeiten bei Klug et al., 2016), der Schülerinnen und Schüler (z. B. uneinheitliches Leistungsprofil bei Kaiser, Retelsdorf, Südkamp, & Möller, 2013), der einzuschätzenden Konstrukte bzw. Aufgaben (z. B. Domänenspezifität bei Südkamp et al., 2012) und der Klasse (z. B. Klassengröße bei Paleczek et al., 2017) untersucht. Die Ergebnisse sind jedoch nicht eindeutig, so dass bisher nicht abschließend geklärt ist, warum die Unterschiede zwischen Lehrkräften und Inhaltsbereichen so gravierend ausfallen (Machts et al., 2016; Paleczek et al., 2017; Südkamp et al., 2017).

Einschlägige Messtheorien im Bereich der Diagnostik, wie etwa die Signalentdeckungstheorie (Brunswik, 1940), legen nahe, dass die Genauigkeit einer Beurteilung insbesondere von der Merkmalspräsentation, d. h. von der Verfügbarkeit eines einzuschätzenden Hinweisreizes abhängt. Da

dieser Faktor nicht durch Lehrkräfte steuerbar ist, wäre eine alleinige Attribution von Urteilsgenauigkeit auf die Lehrkräfte verfehlt. Interessant sind hierzu insbesondere die Ergebnisse von Mehrebenen-Analysen, die den Effekt von Lehrkräftemerkmalen unter Kontrolle anderer Varianzkomponenten schätzen können (z. B. Effekt der Klasse oder der Schule). Praetorius, Koch, Scheunpflug, Zeinz und Dresel (2017) stellen mit eben diesem Studiendesign fest, dass Lehrkräftemerkmale (Berufserfahrung und Selbstkonzept) einen kleineren Effekt auf die Urteilsgenauigkeit haben als Merkmale der Schülerinnen und Schüler. Ähnlich kommen Paleczek et al., 2017 zu dem Ergebnis, dass der Effekt der Lehrkräftemerkmale auf die Urteilsgenauigkeit kleiner ist als der Effekt der Klassenzusammensetzung. Auch diese Ergebnisse sprechen gegen die Verwendung von Urteilsgenauigkeit als Indikator für diagnostische Kompetenz.

Das bisher im Zusammenhang mit Urteilsgenauigkeit am häufigsten untersuchte Merkmal von Lehrkräften ist die Berufserfahrung. Zwar ist Berufserfahrung kein direkter Indikator für Kompetenz oder Expertise, zumindest auf Aggregatsebene sollten aber mittlere Zusammenhänge gegeben sein, wenn Urteilsgenauigkeit ähnlich wie andere Kompetenzen ein Ergebnis akkumulierter, reflektierter und integrierter Erfahrung wäre. Bisher wurden jedoch maximal kleine Effekte der Berufserfahrung auf die Urteilsgenauigkeit gezeigt (McElvany et al., 2009; Praetorius, Karst, Dickhäuser & Lipowsky, 2011; Schrader, 1989). Weit weniger untersucht ist der Zusammenhang von Urteilsgenauigkeit mit kognitiven Basisfähigkeiten von Lehrkräften. Dazu existieren zwei Studien, die einen positiven Effekt von Intelligenz auf die Urteilsgenauigkeit zeigen, einerseits im simulierten Klassenraum (Kaiser, Helm, Retelsdorf, Südkamp, & Möller, 2012) und andererseits in einem Rekognitions- und Verifikationstest (Studie in Kapitel 10). Der Zusammenhang mit kognitiven Basisfähigkeiten spricht nach der in Kapitel 2 entwickelten Kompetenzdefinition zwar nicht gegen eine Existenz zugehöriger Kompetenz, es sollten jedoch andere Konstrukte mit einem stabilen und größerem Effekt auf die Urteilsgenauigkeit identifiziert werden, bevor die Existenz einer Kompetenz postuliert wird, für die Urteilsgenauigkeit ein verlässlicher Indikator ist. Zusammengenommen mit der dargestellten Befundlage stellt sich die Frage, inwiefern Urteilsgenauigkeit nicht lediglich ein Effekt verschiedener Faktoren in einer Beurteilungssituation ist, der sich nur sekundär auf eine diagnostische Kompetenz von Lehrkräften zurückführen lässt. Trotz der unklaren Bedeutung des Konstrukts der Urteilsgenauigkeit für die diagnostische Kompetenz von Lehrkräften hat die Zusammenfassung und Verarbeitung von Informationen hin zu einem akkuraten Urteil jedoch durchaus eine große Bedeutung für Lern- und Entwicklungsprozesse von Schülerinnen und Schülern.

Eine zutreffende Einschätzung hat etwa positive Effekte auf das akademische Selbstkonzept (Urhahne, Chao, Florineth, Luttenberger, & Paechter, 2011), die Lern- und Leistungsmotivation (Jussim, 1989; Urhahne, 2015) und das Leseverstehen (Schwab, Seifert, & Gasteiger-Klicpera, 2014). Urteile der Lehrkräfte haben darüber hinaus einen Effekt auf die schulische Leistung (Rubie-Davies, Hattie & Hamilton, 2006) und sind häufig Entscheidungsgrundlage für institutionelle Bildungswege (Südkamp, Kaiser, & Möller, 2012). Stang und Urhahne (2016) vergleichen daher Urteile von Lehrkräften mit ihren schülerinnen- und schülerbezogenen Erwartungen und stellen die Hypothese auf, dass diese Effekte der Beurteilung u. a. auf selbsterfüllende Prophezeiungen zurückzuführen sind. Es zeigte sich etwa, dass eine Überschätzung der Schulleistung zu einer höheren Ausprägung von Erfolgserwartung und einer niedrigeren Ausprägung von Leistungsangst führte, welche wiederum

einen positiven Effekt auf die Schulleistung hatte. Vor diesem Hintergrund verstehen andere Modelle diagnostischer Kompetenz das Konstrukt der Urteilsgenauigkeit als einen von mehreren Qualitätsindikator diagnostischen Handelns (Leuders et al., 2018; Streit, Rüede, Weber, & Graf, 2019).

Wissensbasierte Modellbildung als Prozess der Informationsverarbeitung. Die Expertiseforschung zeigt disziplinenübergreifend, dass Expertinnen und Experten auf ein Wissen zurückgreifen, das fall- oder kategorienbezogen vernetzt ist. Dadurch sind sie in der Lage, sehr wenige und ggf. heterogene Informationen mit hoher Geschwindigkeit in verschiedenen Interpretationsrahmen auf ihre Kohärenz zu prüfen, darauf bezogene Hypothesen zu entwerfen und deren Wahrscheinlichkeit abzuschätzen (Bromme, 2008; Raufaste, Eyrolle, & Mariné, 1998). Zentrale Ansätze zur Definition diagnostischer Kompetenz stellen die domänenspezifische Wissensorganisation und -verwendung in den Mittelpunkt ihrer Modelle, sowohl in der Lehrkräftebildung (Hellmann, 2015; Ostermann, Leuders et al., 2018; Ostermann, 2018; Streit et al., 2019) als auch in der Medizin (Brush et al., 2017; Elstein & Schwarz, 2002; Monteiro & Norman, 2013; Nystrom et al., 2016).

Für die diagnostische Kompetenz von Lehrkräften greifen Ostermann (2018) und Hellmann (2015) auf das kognitionspsychologische Modell von Nickerson (1999) zurück, das die Ankerheuristik auf den Prozess der Exploration fremder Wissensbestände und -strukturen überträgt. Hellmann (2015) beschreibt diagnostische Kompetenz von Lehrkräften als „Fähigkeit, die Perspektive der Schüler im Sinne Nickersons (1999) adäquat einzunehmen und Urteile über Aufgabenschwierigkeiten und Lösungshäufigkeiten aus deren Sicht zu fällen“ (S. 54). Dieser Prozess der Perspektivenübernahme wird jedoch nicht im Sinne der sozial-kognitiven Fähigkeit der Empathie verstanden (Preston & de Waal, 2002; Singer & Lamm, 2009), sondern als systematischer Aufbau eines kognitiven Modells des fach- bzw. aufgabenspezifischen Wissens von Schülerinnen und Schülern.

Das Modell von Nickerson (1999) postuliert mehrere Schritte, durch die sich eine Person das Wissen einer anderen Person erschließt, sowie mögliche Fehler während dieser Schritte (siehe Abbildung 1). Die eigene Wissensbasis stellt das primäre mentale Modell dar, von dem die einschätzende Person ausgeht. Dieses wird im ersten Schritt um Merkmale der Einschätzungssituation korrigiert, die als untypisch eingeschätzt werden, und dann als Standardmodell für alle in der Situation befindlichen Personen verwendet. In der anschließenden Spezifikation des Modells es in einem zweiten Schritt erneut korrigiert, sowohl um kategoriales Wissen über die Personengruppe, der die fremde Person angehört, als auch um Erfahrungen mit dieser Personengruppe aus dem Langzeitgedächtnis. Dieses Initialmodell über das Wissen einer anderen Person wird in einem dritten Schritt kontinuierlich durch hinzukommende Informationen über die spezifische Person verfeinert und in ein Arbeitsmodell überführt.

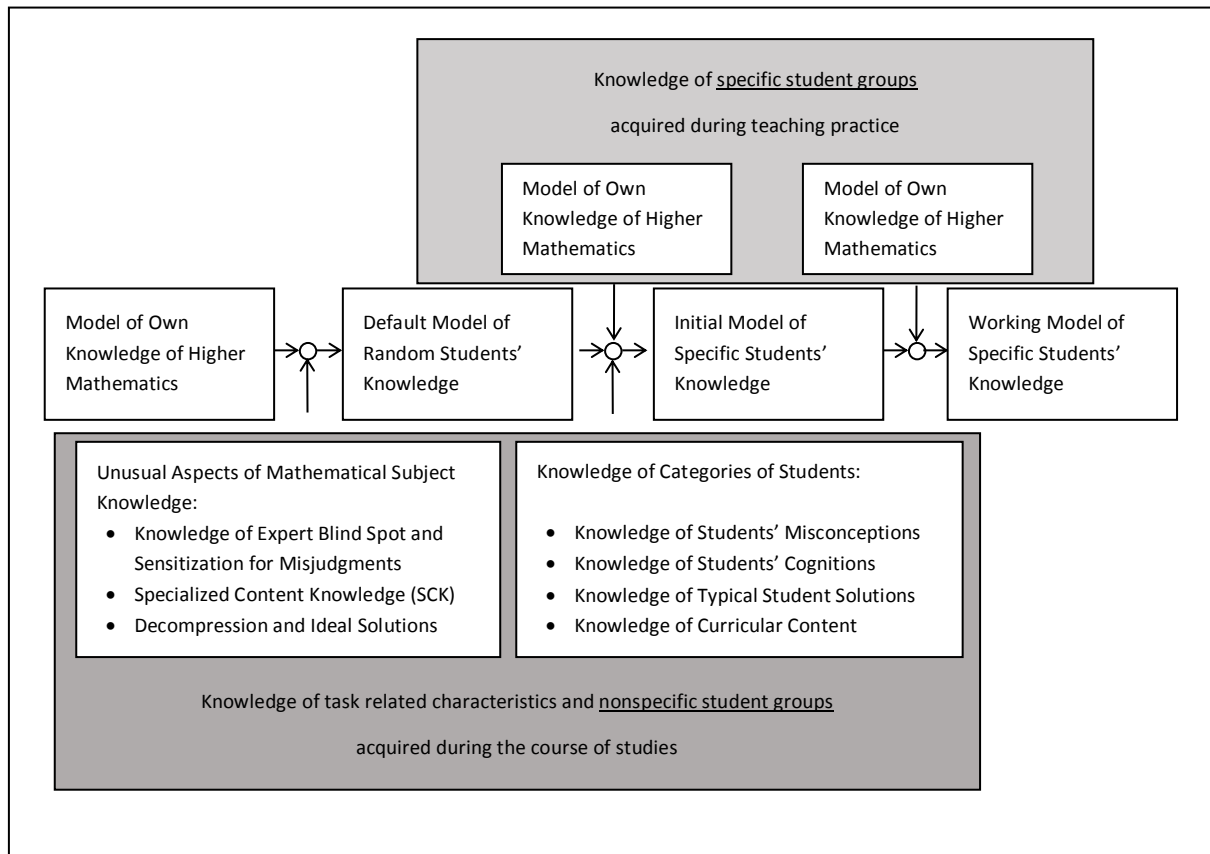


Abbildung 1. Wissensbasiertes Modell fachdidaktischer Diagnose (übernommen von Ostermann, 2018).

Hellmann (2015) und Ostermann (2018) passen dieses Modell auf die Vermittlung von Fachwissen im Unterricht an. Beide gehen davon aus, dass Lehrkräfte zunächst ihr eigenes Fachwissen aktivieren (Was weiß ich über die geforderten Fähigkeiten?) um eine generalisierte Einschätzung über die didaktische Eignung einer Aufgabe zu treffen (Ist das eine schwere Aufgabe?), anschließend daraus ein Basismodell abzuleiten (Was wissen andere über das Thema?) und schließlich schülerinnen- und schülerspezifische Wissensmodelle zu entwerfen (Welche Wissenslücken hat eine Schülerin oder ein Schüler?). Beide Modelle beschreiben damit die Anforderungssituation diagnostischer Kompetenz (adaptives didaktisches Handeln) und den Kontext, in dem diese zur Anwendung kommt (der Unterricht). Während beide Modelle als Voraussetzung der Handlungsfähigkeit in der unterrichtlichen Einschätzung von Merkmalen der Aufgaben und Fähigkeiten der Lernenden das fachdidaktische Wissen (Shulman, 1987) benennen, geht Ostermann (2018) darüber hinaus und benennt das Wissen über fachbezogene Fehlvorstellungen von Schülerinnen und Schülern sowie die Fähigkeit zur Dekomposition von Unterrichtsaufgaben als notwendige Dispositionen für diagnostische Kompetenz. Hellman (2015) findet in einer empirischen Studienreihe erste Belege, die für die Anwendbarkeit von Nickersons (1999) Modell sprechen. Ein zentrales Ergebnis ist der Befund eines negativen Zusammenhangs von Berufserfahrung mit der Tendenz, in einer Einschätzung bei dem eigenen Wissensmodell zu bleiben. Darüber hinaus konnten Lehrkräfte, die ihnen bekannte Schülerinnen und Schüler einschätzten, ein genaueres Urteil zur Lösungshäufigkeit von Aufgaben treffen, als Lehrkräfte, die die einzuschätzenden Schülerinnen und Schüler nicht kannten (Hellmann, 2015).

Streit et al. (2019) knüpfen ebenfalls an eine fachdidaktische Perspektive auf diagnostische Kompetenz an, genauer wird eine „eine didaktisch ausgerichtete Diagnose, die Kinder in ihrem fachlichen Lernprozess fördert“ (S. 39) modelliert und mit dem Begriff der Lernstandseinschätzung zusammengefasst. Im Zentrum des Modells steht die „Rekonstruktion der individuellen Denk- und Lernwege, insbesondere der handlungsleitenden Gründe“, die entlang „systematischer, interpretativer Vorgänge (z. B. Beschreiben, Bewerten, Hypothesen bilden, Maßnahmen ableiten)“ erfolgt (Streit et al. 2019, S. 41). In ihrer empirischen Studie stellen Streit et al. (2019) fest, dass didaktische Fachkräfte mit Expertise (d. h. Wissenschaftlerinnen und Wissenschaftler aus der Mathematikdidaktik mit Unterrichtserfahrung) bereits während des Diagnostizierens auf ihr Wissen zu pädagogischen Anschlusshandlungen zurückgreifen und dieses in ihr diagnostisches Vorgehen einbinden. Auf Novizinnen und Novizen (d. h. Lehramtsstudierende) traf dies nicht zu. Das vorausschauende Denken der Expertinnen und Experten stand darüber hinaus in einem stärkeren Zusammenhang mit der Verwendung von Informationen zu spezifischen Produkten der Schülerinnen und Schüler und deren Fehlvorstellungen (z. B. eine bearbeitete Aufgabe), als bei den Novizinnen und Novizen. Hellmann (2015) kommt dagegen zu dem Ergebnis, dass kein Zusammenhang zwischen Expertise und der Verwendung schülerinnen- und schülerspezifischer Informationen besteht.

Es ist noch unklar, inwiefern sich das Modell von Nickerson (1999) auch auf andere Domänen als die Fachdidaktik anwenden lässt. Jedoch beinhalten die beschriebenen Modelle eine klare Anforderungssituation, beschreiben die zur Bewältigung der Anforderungen notwendigen kognitiven Module und Strukturen, sowie deren funktionelles Zusammenwirken. Lediglich die Definition von Kriterien zur inhaltlichen Bestimmung von Kompetenzniveaus bleiben aus – möglicherweise, weil Urteilstgenauigkeit als einziger Indikator für kontextspezifische Leistung verwendet wird.

Während Modelle aus der Lehrkräftebildungsforschung zwar auf die Möglichkeit einer systematischen und heuristischen Form der Informationsverarbeitung beim Diagnostizieren hinweisen, diese aber nicht systematisch begründen (z. B. Hellmann, 2015; Leuders et al., 2018; Ostermann, 2018), gehen theoretische Modelle aus der Medizin über diesen Stand hinaus. Vor dem Hintergrund empirischer Befunde aus kognitions- und sozialpsychologischen Studien schlagen sie im Rahmen von Zwei-Prozess Modellen konkrete kognitive Mechanismen in Verbindung mit spezifischen Wissensstrukturen vor, die auch an die Expertiseforschung zu Lehrkräften anschlussfähig sind.

Brush et al. (2017) unterscheiden auf Grundlage eines Reviews über diagnostische Kompetenz von Ärztinnen und Ärzten zwischen formalisiertem und regelbasiertem Wissen auf der einen Seite und kontextualisiertem und erfahrungsbasiertem Wissen auf der anderen Seite. Die Besonderheit erfahrungsbasierten Wissens liegt nach Brush et al. (2017) in der Organisation der gespeicherten Informationen nach individuellen und eigensinnigen Kriterien im Sinne eines illness scripts. Die skriptförmige Speicherung von Informationen über diagnostische Situationen bzw. Fälle integriert Informationen über die zu diagnostizierende Person, den Ablauf der Diagnostik, den weiteren Behandlungsverlauf und den konkreten Kontext des diagnostischen Handelns zu einer komplexen und zeitlich strukturierten kognitiven Repräsentation im Langzeitgedächtnis. Für die systematische Form der Informationsverarbeitung nehmen Brush et al. (2017) die Aktivierung kognitiver Algorithmen bzw. regelgeleiteter Ablaufschemata an, wie etwa hypothesenbasiertes Abwägen von Wahrscheinlichkeiten und kriteriengeleitete Entscheidungsfindung. Für die heuristische Form der Informationsverarbeitung

gehen sie von einem iterativen Abgleich zwischen dem aktuellen Wahrnehmungsprozess und den im Langzeitgedächtnis gespeicherten Skripten aus. Der Grad der Übereinstimmung zwischen aktuellen und gespeicherten Modellen drückt dann die Wahrscheinlichkeit einer Kategorisierung aus. Wie diese beiden Prozesse zusammenwirken, beschreiben Brush et al. (2017) jedoch nicht.

Dieses Zusammenwirken greifen dagegen Monteiro und Norman (2013) in einem Review auf, das ebenfalls diagnostische Kompetenz im medizinischen Kontext thematisiert. Sie gehen nicht von einer Entkopplung der systematischen und heuristischen Informationsverarbeitung und deren separater Aktivierung aus. Denn dies würde bspw. implizieren, dass heuristische Verarbeitungsprozesse prinzipiell als Voreinstellung ablaufen, da sie sowohl lokal als auch temporal weniger kognitive Kapazität bzw. Ressourcen beim Diagnostizieren beanspruchen (bspw. gemessen durch den relativen Blutsauerstoffgehalt in differenziell aktivierten Hirnregionen wie dem präfrontalen Kortex; ausführlich siehe Rotgans et al., 2019). Stattdessen nehmen sie ein parallel-kompetitives Modell an, in dem beide Prozesse gleichzeitig ablaufen und eine Wechselwirkung (z. B. kompetitive Hemmung) aufeinander haben (siehe Abbildung 2).

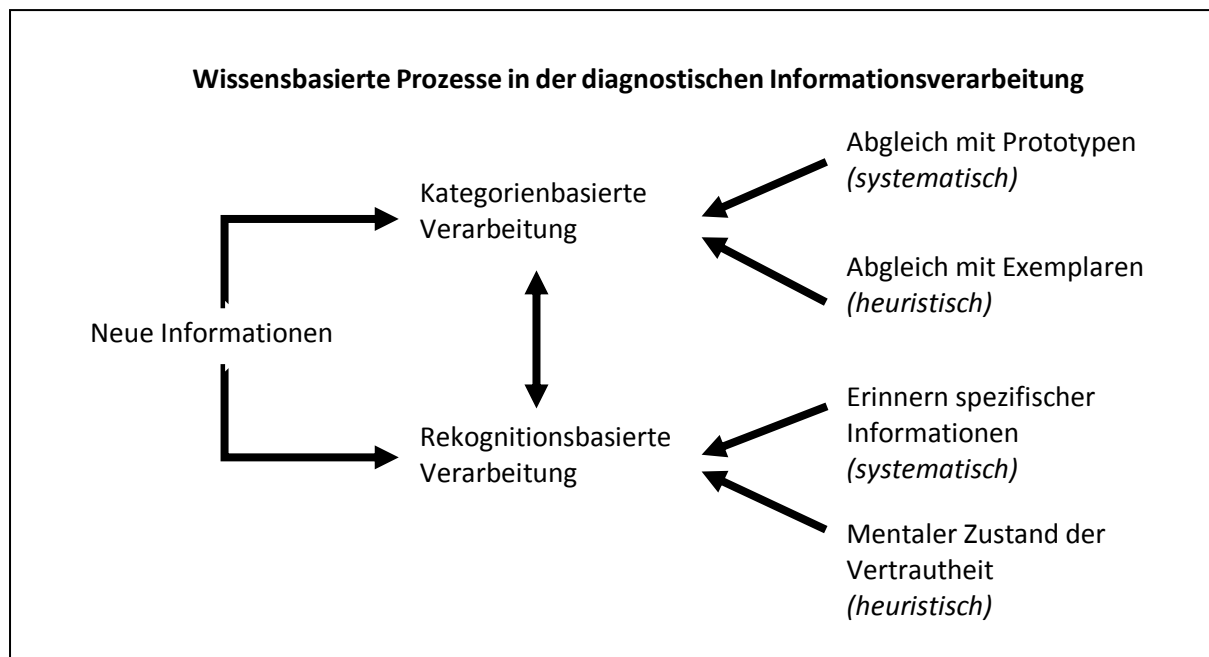


Abbildung 2. Modell wissensbasierter diagnostischer Informationsverarbeitung nach Monteiro und Norman (2013). Eigene Darstellung.

Monteiro und Norman (2013) nehmen an, dass in zwei verschiedenen Verarbeitungsmechanismen sowohl heuristische als auch systematische Informationsverarbeitung mit unterschiedlichen Wissensformen stattfindet: in kategorienbasierter und rekognitionsbasierter Verarbeitung. Die kategorienbasierte Verarbeitung beschreibt das Abrufen gespeicherter Prototypen und Exemplare mit großer Ähnlichkeit zu neu wahrgenommenen Informationen. Auf Wissen, das in Form von Prototypen im Sinne einer analytischen Abstraktion vorliegt (bspw. durch Durchschnittsbildung), greift primär das System der systematischen Informationsverarbeitung zu. Wissen, das in Form von Exemplaren im Sinne individualisierter Einzelfälle vorliegt, wird eher in die heuristische Verarbeitung eingebunden.

Während einer Diagnostik wird Wissen beider Formen verarbeitet. Die rekognitionsbasierte Verarbeitung beschreibt das Wiedererkennen bzw. Assoziieren neuer mit vorhandenen Informationen auf zwei Arten. Entweder durch das systematische Erinnern von Übereinstimmungen spezifischer Details bzw. Informationen oder als Bewusstsein um einen eigenen mentalen Zustand einer vagen Vertrautheit (heuristische Verarbeitung). Monteiro und Norman (2013) gehen davon aus, dass sich die kategorien- und rekognitionsbasierte Verarbeitung wechselseitig beeinflussen, etwa indem das kategorienbasierte System eine basale Struktur zum Abgleich von Informationen zur Verfügung stellt und das rekognitionsbasierte System eine Struktur zur Veränderung, Stabilisierung und Gewichtung von Prototypen und Exemplaren. Elstein (2002, 2009) und Nystrom (2016) konzipieren den diagnostischen Prozess auf Basis von Zwei-Prozess Modellen und Forschung zum Professionswissen von Expertinnen und Experten auf ähnliche Weise, indem sie die kognitiven Vorgänge beim Diagnostizieren als induktive und deduktive Prozesse zur Ausweitung und Ausdifferenzierung mentaler Modelle von Diagnosen bzw. Krankheitsbildern im Sinne von miteinander assoziierten Informationen zu Symptomen und deren Kausalbeziehungen beschreiben. Sie nehmen an, dass eine diagnostische Informationsverarbeitung auf Basis von Prototypen eher zur Identifikation noch benötigter Informationen führt und heuristische Verarbeitungsformen wesentlich zur initialen Strukturierung der diagnostischen Situation beitragen.

Zusammenfassend bieten Modelle aus der medizinischen Diagnoseforschung eine kohärente und weitreichend empirisch gestützte Grundlage zur Formulierung von Modellen diagnostischer Kompetenz. Im Modell von Monteiro und Norman (2013) hat Diagnostizieren einen konstruktiven Charakter, d. h. es bewirkt auch Veränderungen im kognitiven Apparat der Diagnostizierenden. Für Kompetenzmodelle ist das eine attraktive Eigenschaft, da so auch eine Erklärung bereitsteht, wie es durch eine kontinuierliche Praxis des Diagnostizierens und zugehöriger Erfahrungsverarbeitung zu den spezifischen Veränderungen kommt, die eine kontextspezifische Expertise mit sich bringen. Die Anforderungssituationen von Lehrkräften und Ärzten unterscheiden sich jedoch. Sowohl in der Gruppensituation als auch den einzuschätzenden Merkmalen und den dafür verfügbaren technologischen Möglichkeiten. Die Modelle sind jedoch in einem Abstraktionsgrad formuliert, der einen Transfer auf diagnostische Situationen von Lehrkräften prinzipiell ermöglicht. Insbesondere die Annahme einer heuristischen Informationsverarbeitung durch den Abgleich mit Fällen und die zugehörige Konzeption des Kompetenzniveaus als mentaler Zustand sind konzeptuelle Bausteine, die für Theorien zur diagnostischen Kompetenz von Lehrkräften auf Basis wissensbezogener Informationsverarbeitung eine sinnvolle Ergänzung darstellen.

Sozial-kognitive Ansätze

Der Gegenstand sozial-kognitiver Theorien ist die Wechselwirkung individueller Denk- und Verarbeitungsprozesse mit der sozialen Umwelt von Personen, wie etwa in der Lerntheorie von Bandura (1991). Um das Zusammenwirken von intra-individuellen und inter-individuellen Strukturen und Prozessen zu erklären, liegt ein Fokus dieser Theorien auf der Bildung kognitiver Modelle durch und für soziales Verhalten sowie auf der Beschreibung dafür relevanter Umweltstrukturen.

Dementsprechend greift insbesondere die Interventionsforschung auf entsprechende Theorien zurück, u. a. um relevante Stimuli für die Steigerung kognitiver Fähigkeiten zu identifizieren. Daher sind im Rahmen sozial-kognitiver Ansätze bzgl. diagnostischer Kompetenz insbesondere Konzeptionen der Schnittstellen zwischen Lehrkräften und ihrer Umwelt interessant.

Evidenzbasierte Förderung als diagnostisches Handeln. Klug (2011) stellt in einer Studienreihe ein Modell diagnostischer Kompetenz vor, das einen besonderen Fokus auf die Verknüpfung informationsgenerierender Schritte und Förderziele für Schülerinnen und Schülern setzt. Die Definition diagnostischer Kompetenz, die Klug, Bruder, Keller und Schmitz (2012) vorschlagen, richtet sich dabei auf die Beurteilung und Förderung selbstregulierten Lernens von Schülerinnen und Schülern:

Beim Diagnostizieren des Lernverhaltens geht es darum zu erkennen, welche Lernstrategien der Schüler bereits anwenden kann und wo er sich im Lernprozess befindet. Insbesondere seine Fähigkeit selbstreguliert zu lernen wird diagnostiziert mit dem Ziel, diese zu fördern. Die Diagnostik von Lernverhalten schließt eine Diagnostik von Lernstörungen aus, da diese nicht im Handlungsspielraum einer Lehrkraft liegt. Vielmehr geht es darum Schwierigkeiten und Verbesserungspotenzial im Lernverhalten der Schüler zu erkennen, so dass anschließend passende Lernstrategien durch die Lehrkraft vermittelt werden können. (S. 5)

Klug (2011) legt dazu ein auf empirischen Studien über Lehrkräfte basierendes Modell vor, das zentrale Konstrukte im Ablauf einer Diagnostik benennt. Das Modell unterscheidet drei aufeinander aufbauende Phasen diagnostischen Handelns, die Lehrkräfte in einem zyklischen Prozess durchlaufen (siehe *Abbildung 3*).

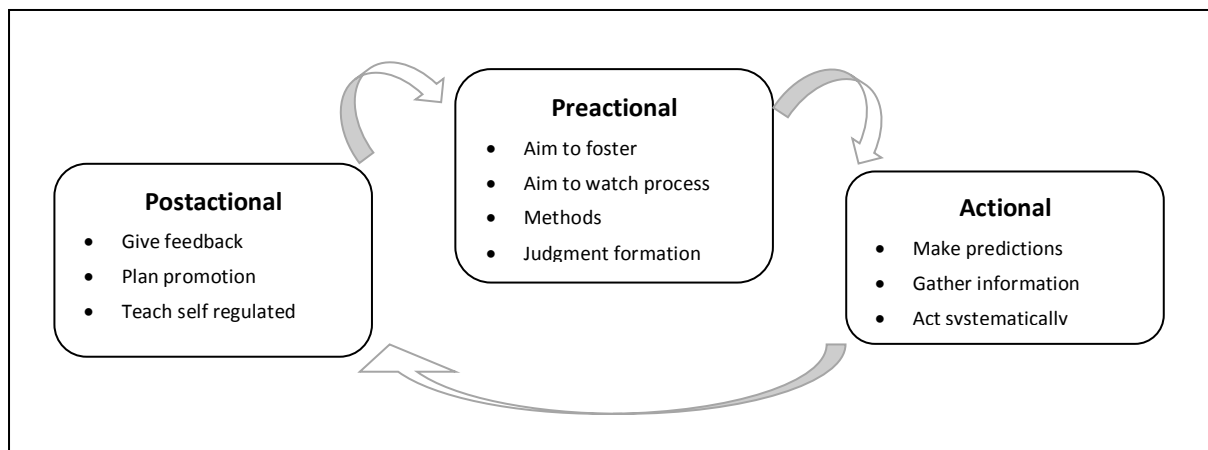


Abbildung 3. Prozessmodell diagnostischen Handelns von Lehrkräften nach Klug (2011).

Die präaktionale Phase umfasst die kognitiven Prozesse der Zielsetzung förderrelevanten Handelns und der Beobachtung von Lernverläufen sowie der Aktivierung theoretischen Wissens zu Diagnosemethoden, Urteilsbildung und -fehlern sowie Qualitätskriterien im Kontext von Diagnostik (z. B. Testgütekriterien). Die daran anschließende aktionale Phase beschreibt das hypothesenbasierte Sammeln und Interpretieren von Informationen mit dem Ziel eines vorübergehenden diagnostischen Urteils. In der postaktionalen Phase folgen die Kommunikation mit Beteiligten (z. B. Feedback) und die

Förderplanung und -durchführung. Die dreidimensionale Struktur des Modells wurde mittels konfirmatorischer Faktoranalyse eines schriftlichen Szenariotests bestätigt ($\chi^2 = 47.704$, $df = 36$, $p = .092$, $SRMR = .045$, $RMSEA = .033$, $CFI = .954$) und Lehrkräfte sowie Lehrkräfte im Vorbereitungsdienst erreichten in dem Test höhere Werte als Studierende (Klug et al., 2012), so dass zwei belastbare Hinweise für die Passung des Modells vorliegen. Für die Struktur diagnostischer Kompetenz orientiert sich das Modell an den allgemeinen Komponenten professioneller Kompetenz der COACTIV-Studie (Wissen, Fähigkeiten, Überzeugungen, Motivation). Als relevante Prädiktoren der Leistung im Szenariotest wurden Wissen ($.23 < \beta < .40$), Motivation ($.25 < \beta < .30$) und Einstellung zu Diagnostik ($.32 < \beta < .38$) sowie das professionelle Selbstkonzept ermittelt (Klug et al., 2016; Klug et al., 2012). Das Regressionsgewicht β bezeichnet die semipartielle Korrelation der Prädiktoren mit der erreichten Testleistung.

Kritisch zu hinterfragen ist die definitorische Abgrenzung zur Diagnostik von Lernstörungen. Laut psychologischer und psychiatrischer Diagnosestandards sind Lehrkräfte bspw. in die Diagnostik klassischer Lernstörungen wie Lese-Rechtschreibschwäche, Dyskalkulie, AD(H)S und Hochbegabung in den diagnostischen Prozess zu involvieren und müssen daher in entsprechenden Situationen handlungsfähig sein (z. B. Bundesärztekammer, 2005; Falkai & Wittchen, 2015). Prinzipiell sind Lehrkräfte per Gesetz der Bundesländer dazu berechtigt, Lese-Rechtschreibschwäche, Rechenschwäche und Hochbegabung eigenständig zu diagnostizieren, sowohl in der Grundschule (GsVO §16(2)), als auch in der Mittelstufe (Sek I-VO §16-19) und in der gymnasialen Oberstufe (VO-GO §31(1)).

Da auf dieser Kopplung mit dem Gesundheitssystem auch Fördermaßnahmen basieren, wie z. B. Nachteilsausgleiche oder die Kostenübernahme von Lerntherapie durch Krankenkassen, ist eine weichere Auslegung der Definition in diesem Punkt angebracht. Dies beschreibt Klug (2011) auch implizit, wenn sie die diagnostischen Planungsprozesse ausdifferenziert: "In these plans, (a) the characteristics of the student that should be fostered, for example, motivation, concentration, creativity, ability to think abstractly, and so forth; (b) the student's actual skill level; (c) the goals to be reached; and (d) the measures that will be used to reach these goals should be written down" (S. 19). Auch Klug, Bruder, Kelava, Spiel und Schmitz (2013) erweitern die Definition diagnostischer Kompetenz in aufgezählten Beispielen inhaltlich um Konstrukte, die klassischerweise in die Diagnostik von Lernstörungen eingebunden sind (z. B. Aufmerksamkeit). Als Lernverhalten bezeichnen sie „observable behavioral patterns that children display as they approach and undertake school learning tasks. Such learning behaviors include observable problem-solving strategies, flexibility, attention, and responses to learning situations (competition, novelty, error, etc.) in classroom settings" (S. 39). In dieser Definition erfolgt auch eine kontextuelle Eingrenzung auf den Unterricht, sodass Situationen wie die Pausen- und Hofaufsicht, in der ebenfalls Lernverhalten beobachtet werden kann (z. B. Hausaufgaben vor dem Unterricht abschreiben), nicht einbezogen werden. Auch Beratungssettings, z. B. von Vertrauenslehrerinnen und -lehrern, werden damit ausgeklammert. Da Klug et al. (2012) selbst einen mittleren Zusammenhang zwischen diagnostischer Kompetenz und Beratungskompetenz von Lehrkräften feststellen, scheint auch diese kontextuelle Abgrenzung nicht empirisch haltbar zu sein.

Zusammengefasst bildet das Modell von Klug (2011) grundlegende Strukturen und Prozesse diagnostischen Handelns ab und übernimmt zur Modellierung der Kompetenzstruktur den Vorschlag

von Weinert (1990) zur Abbildung durch Wissen und (kognitiven) Fähigkeiten. Durch die Operationalisierung über einen Szenariotest ist auch die inhaltliche Beschreibung von Niveaustufen möglich, bislang jedoch noch nicht erfolgt. Das Modell greift zentrale Kritik am Modell der Urteilsgenauigkeit auf und setzt einen Fokus auf die Förderung von Schülerinnen und Schülern als Zweck diagnostischen Handelns. Dadurch rückt die diagnostische Zielsetzung stärker in den Fokus, die auch in allgemeinen Motivations- und Handlungstheorien ein notwendiger Baustein zur stabilen Volitionsbildung ist (Heckhausen & Gollwitzer, 1987). Auch für Klauer (1982) gelten die Spezifikation eines Förderbedarfs und die Verknüpfung mit pädagogisch-didaktischen Maßnahmen als Ziel von Diagnostik im Schulkontext. Darüber hinaus bildet das zirkuläre Phasenmodell die funktionelle Verbindung der involvierten kognitiven Module und Strukturen ab und ist darin wegweisend. Weiterhin zeigt die Kritik, dass sich das Modell entgegen der angenommenen Definition durchaus bereits auf allgemeine Lernvoraussetzungen bezieht und anwenden lässt. Eine Ausweitung auf weitere diagnostische Anforderungssituationen scheint daher sinnvoll. Das Modell regt auch zu einer inhaltlichen Klärung der Art und des Umfangs der Förderung von Schülerinnen und Schülern an, die im Verantwortungsbereich von Lehrkräften liegt.

Informationsbasierte Urteilsbildung als diagnostisches Handeln. Der wohl konsistente Fokus von Modellen zu diagnostischem Handeln ist der entscheidungsgenerierende Umgang mit personenspezifischen Informationen. Ein diesbezügliches und im wissenschaftlichen Diskurs stark wahrgenommenes Modell haben Herppich et al. (2018) konzipiert. Auf Grund seiner potenziellen Bedeutung für den theoretischen Diskurs zu diagnostischer Kompetenz wird es an dieser Stelle genauer diskutiert. Herppich et al. (2018) stellen ein Rahmenmodell zu „assessment competence“ vor, dass sich sowohl auf fachdidaktische wie auch auf allgemein pädagogische und psychologische Fragestellungen im Schulkontext bezieht (z. B. schulfachbezogene Leistung feststellen, Übergangsempfehlungen erstellen, Lernstörungen identifizieren). Das Modell soll eine Art Blaupause darstellen, „a template to sample relevant assessment situations, and a model of assessment as a process that can be tailored to given boundary conditions“ (Herppich et al., 2018, S. 1). Kern des Modells ist die systematische Informationssammlung mit dem Ziel einer Urteilsbildung. Die Autorinnen und Autoren gehen dabei von einer Gleichsetzung mit diagnostischer Kompetenz aus und definieren Assessmentkompetenz wie folgt: „we defined educational assessment as the process of assessing school students with respect to those characteristics that are relevant for learning, in order to inform educational decisions“ (Herppich et al., 2018, S. 2f). Die daraus abgeleitete Kompetenzstruktur umfasst zum einen kognitive Dispositionen in Form von pädagogischem Wissen, Fachwissen und fachdidaktischem Wissen. Zum anderen umfasst sie einen Komplex anderer Dispositionen, zu denen eine generellen Verarbeitungsfähigkeit, Motivation, das Selbstkonzept, Überzeugungen sowie subjektiven Theorien gehören. Darüber hinaus schlagen Herppich et al. (2018) ein verhaltensbasiertes Prozessmodell der Urteilsbildung vor, das ausgehend von einer diagnostischen Zielstellung den Fokus auf zentrale Entscheidungen zur Informationssammlung und daran anschließende formale Schritte setzt (z. B. weitere Informationssammlung notwendig? Ja/Nein, Nein: Urteilsbildung, Ja: Hypothesenbildung notwendig? usw.). Herppich et al. (2018) beschreiben die Entscheidung für oder gegen eine fallbezogene Informationssammlung, als eine zielbezogene Wahl zwischen heuristischer und

systematischer Informationsverarbeitung, wobei nur die systematische Informationsverarbeitung mit einer zusätzlichen Informationssammlung einhergeht. Zur Bestimmung von Kompetenzniveaus werden verschiedene Performanzmessungen diskutiert (z. B. Urteilsgenauigkeit, Einhalten formaler Schritte, verwendete Informationstypen in der Urteilsbildung) und es wird sich für multiple Qualitätskriterien ausgesprochen.

Das Rahmenmodell wirft eine große Anzahl an Fragen auf, die vor einer Weiternutzung in Forschung und Praxis beantwortet werden müssten. Im Rahmen der vorliegenden Arbeit werden einige zentrale Punkte im Detail diskutiert, da eine vollständige Auflistung zu umfangreich wäre.

Dies betrifft erstens die Klärung des verwendeten Kompetenzbegriffs. So beschreiben die Herppich et al. (2018), dass sie in ihrer Definition von Kompetenz sowohl Blömeke et al. (2015) als auch Koeppen et al. (2008) folgen. Diese beiden Ansätze unterscheiden sich jedoch in einigen Punkten kategorial. Blömeke et al. (2015) schlagen nach einer Auseinandersetzung mit theoretischen Ansätzen zur Bestimmung des Verhältnisses von Kompetenz und Performanz vor, dass Kompetenz ein Konstrukt ist, das kognitive Dispositionen, situationsspezifische Fähigkeiten und Performanz im Sinne beobachtbaren Verhaltens umfasst. Darüber hinaus betrachten sie Motivation als Teil von Kompetenz. Koeppen et al. (2008) konzipieren Kompetenz jedoch sehr klar als rein kognitives Konstrukt: „we define competencies as context-specific cognitive dispositions that are acquired and needed to successfully cope with certain situations or tasks in specific domains“ (S. 8). Weitere zentrale Eigenschaften des Kompetenzkonstrukts bei Koeppen et al. (2008) sind Domänen- bzw. Kontextspezifität, der explizite Ausschluss von Motivation aus dem Konstrukt und die theoretische und methodologische Abgrenzung zu Performanz. Herppich et al. (2018) klären nicht auf, wie sie diese gegensätzlichen theoretischen Annahmen in ihrem Modell miteinander vereinbaren.

Zweitens müsste die Konzeptualisierung von Zwei-Prozess Modellen an den Stand der Forschung angeschlossen werden. Herppich et al. (2018) beschreiben diesen wie folgt:

Our model of assessment is based on dual process models of social judgment formation (Ferreira et al., 2006; Fiske & Neuberg, 1990). In line with the assumptions of these models, research on teachers' judgment processes showed that teachers form judgments about students along two information processing modes that constitute the endpoints of a continuum (Dünnebier et al., 2009; Glock & Krolak-Schwerdt, 2014). The first mode is rather simple: It requires little cognitive effort as little information is considered. As this mode involves the activation and processing of heuristic knowledge about the person to be judged, it is assumed efficient and in some cases sufficiently accurate (Gigerenzer & Goldstein, 1996) but also susceptible to subjectivity and bias (Krolak-Schwerdt et al., 2013). (S. 7)

Die Forschung zu Zwei-Prozess-Modellen der Informationsverarbeitung zeigt zwar, dass heuristische Informationsverarbeitung besonders unter Zeitdruck auftritt (siehe die Meta-Analyse von Phillips, Fletcher, Marks, & Hine, 2016). Die Ursache dafür ist jedoch nicht, dass *weniger* Informationen verarbeitet werden. Vielmehr ist mit heuristischer Informationsverarbeitung die Aktivierung einer Reihe *anderer* kognitiver Prozesse und meta-kognitiver Strategien assoziiert, als solche, die während einer systematischen Verarbeitung aktiv sind. Chaiken (1980) diskutiert diesen Unterschied sehr eindrücklich in Bezug auf den Vorgang der Urteilsbildung aus verbalen Botschaften. Bei systematischer Verarbeitung bilden primär konkrete Gesprächsinhalte die Urteilsgrundlage. Bei heuristischer

Verarbeitung bilden primär als relevant eingeschätzte Hinweisreize aus dem Kontext des Gesprächs die Urteilsgrundlage. Daran anschließend beschreiben Gigerenzer und Goldstein (1996), dass in heuristischer Verarbeitung fehlende Informationen durch nach bestimmten Regeln ausgewählte und verfügbare Informationen ersetzt werden. Aber auch eine quantitative Reduktion von Informationen in der Urteilsbildung führt nicht zwingend zur qualitativen Verschlechterung von Urteilen. Gerade die von Herppich et al. (2018) zitierte Studie von Gigerenzer und Goldstein (1996) stärkt diese Aussage. Es handelt sich um eine Überblicks- und Simulationsstudie aus dem Bounded Rationality Paradigma zur statistischen Effizienz von Heuristiken im Vergleich zur Verarbeitung durch aller verfügbaren Informationen. Ein Hauptergebnis ist, dass drei verschiedene Heuristiken selbst bei eingeschränkter Informationsmenge mindestens zur selben Urteilsgenauigkeit führen, wie die optimale Integration aller verfügbaren Informationen und eine getestete Heuristik sogar zu genaueren Urteilen führte. Auch im Bereich der Diagnostik entspricht es nicht der empirischen Befundlage, dass heuristische Informationsverarbeitung fehleranfälliger ist, als systematische Verarbeitung (Brush et al., 2017). So finden Sherbino et al. (2012) einen Anstieg der Urteilsgenauigkeit bei kürzerer Verarbeitungszeit und Monteiro et al. (2015) keine Veränderung der Urteilsgenauigkeit, wenn systematische Informationsverarbeitung experimentell verhindert wird.

Weiterhin gehen Herppich et al. (2018) davon aus, dass heuristische und systematische Verarbeitungsvorgänge getrennt ablaufen. Empirische Forschung zeigt jedoch, dass dies zwar der Fall sein kann, jedoch auch eine gleichzeitige Verarbeitung im heuristischen und systematischen System erfolgen kann (Parallel Operating Framework; Evans, 2008; Smith & DeCoster, 2000). Zuletzt setzt das Modell voraus, dass eine bewusste Entscheidung über den Modus der Informationsverarbeitung getroffen werden kann. Dies ist jedoch keinesfalls konsensueller Stand der Forschung. Beispielsweise zeigt die Meta-Analyse über Interventionen zur Stimulation der beiden Verarbeitungsmodi von Lambe, O'Reilly, Kelly und Curristan (2016): „instructions to use some style of reasoning at the point of diagnosis may amount to little more than exhortations to ‘think harder’, insufficient to alter cognitive behaviour” (S. 10).

Drittens stellt sich die Frage, was Herppich et al. (2018) mit Kontext meinen. Diese Frage ist theoretisch gewichtig, da der Kontextbezug das Kompetenzkonstrukt von generalisierten kognitiven Fähigkeiten unterscheiden soll. Er ist ein konstitutives Merkmal des Konstrukts. Das Modell von Herppich et al. (2018) nimmt Gültigkeit für fachspezifische und fachübergreifende Fragen in Anspruch. Koeppen et al. (2015) definieren den Kontextbegriff jedoch über die Domänenspezifität, d. h. in Anlehnung an Inhaltsbereiche (z. B. Mathematik, Literalität). Blömeke et al. (2015) greifen eben diesen Kontextbegriff auf und präzisieren ihn auf psychometrischer Ebene über Latent-State-Trait Modelle. Herppich et al. (2018) grenzen sich davon trotz ihres ausdrücklichen Anschlusses an obige Kompetenzbegriffe explizit davon ab: “Here, we propose a first generic approach to classifying assessment situations in varying contexts” (S. 5). Vorgeschlagen wird eine empirische Klärung des Kontextes für die Gültigkeit des Modells über Situationen mit Variation in folgenden Dimensionen: formativ/summativ, planbar/spontan, high-stakes/low-stakes und ein-/mehrdimensionale Beurteilung. Es ist natürlich eine offene Frage, wie sich eine solche strukturelle Konzeption zu der fächerspezifischen Organisation der schulischen Praxis und der Lehrkräftebildung und -weiterbildung empirisch bewährt. Auf Ebene des Wissens als kognitive Dispositionen im Modell von Herppich et al.

(2018) könnte als Hinweis zur Klärung dieser Frage die verfügbare empirische Forschung zur faktoriellen Struktur des professionellen Wissens von Lehrkräften herangezogen werden. Diese findet jedoch mehr oder weniger eine Struktur in Anlehnung an das Modell von Shulman (1987) – an dem sich auch die institutionelle Struktur der Lehrkräftebildung inhaltlich orientiert (z. B. König et al., 2018).

Weitere Punkte, die einen kritischen Umgang mit dem Modell nahelegen, betreffen u. a. die Modellierung des Diagnoseprozesses über ausschließlich verhaltensbasierte Elemente. Dies steht im Gegensatz zu der kognitiv orientierten Kompetenzdefinition, den als unterstützende Hinweise für das beschriebene Modell zitierten kognitionspsychologischen Studien, der Verwendung von kognitiven Zwei-Prozess Modellen als Element des theoretischen Rahmens und der Beschränkung auf ein (kognitives) Urteil als Ergebnis des Diagnoseprozesses. Aber auch die Auswahl der drei Entscheidungspunkte im Prozessmodell (Bedarf an zusätzlichen Informationen, Notwendigkeit der Hypothesenformulierung, Notwendigkeit expliziten bzw. systematischen Vorgehens) ohne Diskussion entscheidungsleitender Kriterien sowie der Einschluss kognitiver Basisfähigkeiten in das Kompetenzkonstrukt entgegen der gewählten Definitionen sind Punkte, die eine Nutzung des Modells in weiteren Forschungs- oder Anwendungszusammenhängen erschweren könnten.

Zusammenfassend birgt das Rahmenmodell von Herppich et al. (2018) zunächst theoretische Inkonsistenzen, deren Klärung einer weiteren Verwendung vorausgehen sollte. Sowohl das vorgeschlagene Prozess- als auch das Strukturmodell widersprechen z. T. dem Stand der Forschung zu Diagnostik in und außerhalb der Lehrkräftebildung. Innerhalb der Lehrkräftebildung betrifft dies sowohl fachdidaktisch orientierte Modelle (Hellmann, 2015; Leuders & Prediger, 2016; Edelenbos & Kubanek-German, 2004), als auch fachübergreifende Modelle (Klug, 2011; Ingenkamp & Lissmann, 2008). Es stellt sich daher die Frage, ob der Gültigkeitsanspruch des Modells nicht zu weit gesteckt ist und Kompetenzmodelle weiterhin eine mittlere Reichweite verfolgen sollten. Eine Abgrenzung zum Modell von Herppich et al. (2018) durch zumindest eine Co-Autorin existiert bereits: Klug, Schultes und Spiel (2018) sehen pädagogische Anschlusshandlungen im Gegensatz zum Rahmenmodell von Herppich et al. (2018) als Teil diagnostischer Kompetenz. Das Modell birgt aber auch Erkenntnispotenzial in der Darstellung verschiedener Facetten der Qualität professionellen diagnostischen Handelns. So wird vorgeschlagen, intersubjektiv nachvollziehbare Prozessdokumentationen und die Validität von Informationen in der Urteilsbildung als Qualitätsindikatoren zu verwenden. Diese Kriterien wurden zum einen auch in psychologischer Forschung zu medizinischer Diagnostik in den 70er Jahren produktiv genutzt (Shulman, Elstein, & Sprafka, 1975). Zum anderen sind diesbezüglich bereits Befunde aus der kognitionspsychologischen Forschung zur Urteilsbildung von Lehrkräften verfügbar (Bilz, Steger, & Fischer, 2016; Marksteiner, Reinhard, Dickhäuser, & Sporer, 2012; Marksteiner, Reinhard, Lettau, & Dickhäuser, 2013; Oudman, van de Pol, Bakker, Moerbeek, & van Gog, 2018; van de Pol, de Bruin, van Loon, & van Gog, 2019).

Informationsbasierte Steuerung als diagnostisches Handeln. Diagnostisches Handeln findet in institutionellen Kontexten statt und ist eng mit deren Funktionsweise verzahnt. Ingenkamp & Lissmann (2008) stellen eine Definition und ein Modell diagnostischen Handelns von Lehrkräften vor, das einen besonderen Fokus auf Steuerungsprozesse legt. Die Autoren verorten diesen Ausschnitt professionellen Handelns trotz der begrifflichen Fassung als pädagogische Diagnostik nicht

disziplinspezifisch: „Es hat den Anschein, dass Erziehungswissenschaftler lieber den Begriff Pädagogische Diagnostik verwenden, während Psychologen den Begriff Pädagogisch-psychologische Diagnostik bevorzugen. Auf diese Weise entsteht jedoch eine neue Verunsicherung und „Fachegoismus“ dominiert über interdisziplinäres Denken“ (Ingenkamp & Lissmann, 2008, S. 19). Damit schließen die Autoren an Shulmans (1981) Einschätzung über die (eingeschränkte) Interdisziplinarität der Forschung zu Lehrkräften an. Sie legen eine Definition diagnostischen Handelns vor, die die Individualebene umfasst aber auch darüber hinaus geht:

Pädagogische Diagnostik umfasst alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren. Zur Pädagogischen Diagnostik gehören ferner die diagnostischen Tätigkeiten, die die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen ermöglichen, sowie die mehr gesellschaftlichen verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder der Erteilung von Qualifikationen zum Ziel haben. (Ingenkamp & Lissmann, 2008, S. 13)

Der Tätigkeitsbegriff umfasst dabei die Informationssammlung, -interpretation und -kommunikation und wird zwischen Forschungs- und Evaluationstätigkeiten verortet. Die Definition umfasst die Individual-, Gruppen-, Institutions-, und Gesellschaftsebene und ist daher die bislang umfangreichste Definition diagnostischen Handelns. Zwar arbeiten Ingenkamp und Lissmann (2008) nicht mit dem Kompetenzbegriff, legen aber dennoch ein Prozessmodell diagnostischen Handelns und ein darin integriertes Strukturmodell zu dessen Voraussetzungen vor. Das zentrale Element des Modells ist, ähnlich wie bei Klug (2011), ein zyklischer Interaktionsprozess zwischen Lehrkräften und Schülerinnen und Schülern, für den auf beiden Seiten sowohl kognitive Prozesse als auch Handlungen benannt sind (siehe *Abbildung 4*).

Das entscheidende Moment dieses Modells ist die Initiation kommunikativen Handelns zu Diagnose- und Interventionsergebnissen:

Wenn für uns hinreichend erklärt ist, was wir beobachtet haben, dann ordnen wir es nach unseren Normen ein und kommen zu einem Urteil. Dieses kann mehr oder weniger formell mitgeteilt werden und führt gleichzeitig zu Konsequenzen für weitere pädagogische Aktionen. (Ingenkamp & Lissmann, 2008, S. 17)

In einem zusätzlichen und detaillierteren Prozessmodell sind mehrere Ergebnisse des diagnostischen Prozesses vorgesehen. Dazu gehören Resignation bzw. Einschaltung von Expertinnen und Experten, die Mitteilung des diagnostischen Urteils und die Wirkung auf die Determinanten des Diagnoseprozesses (siehe *Abbildung 5*).

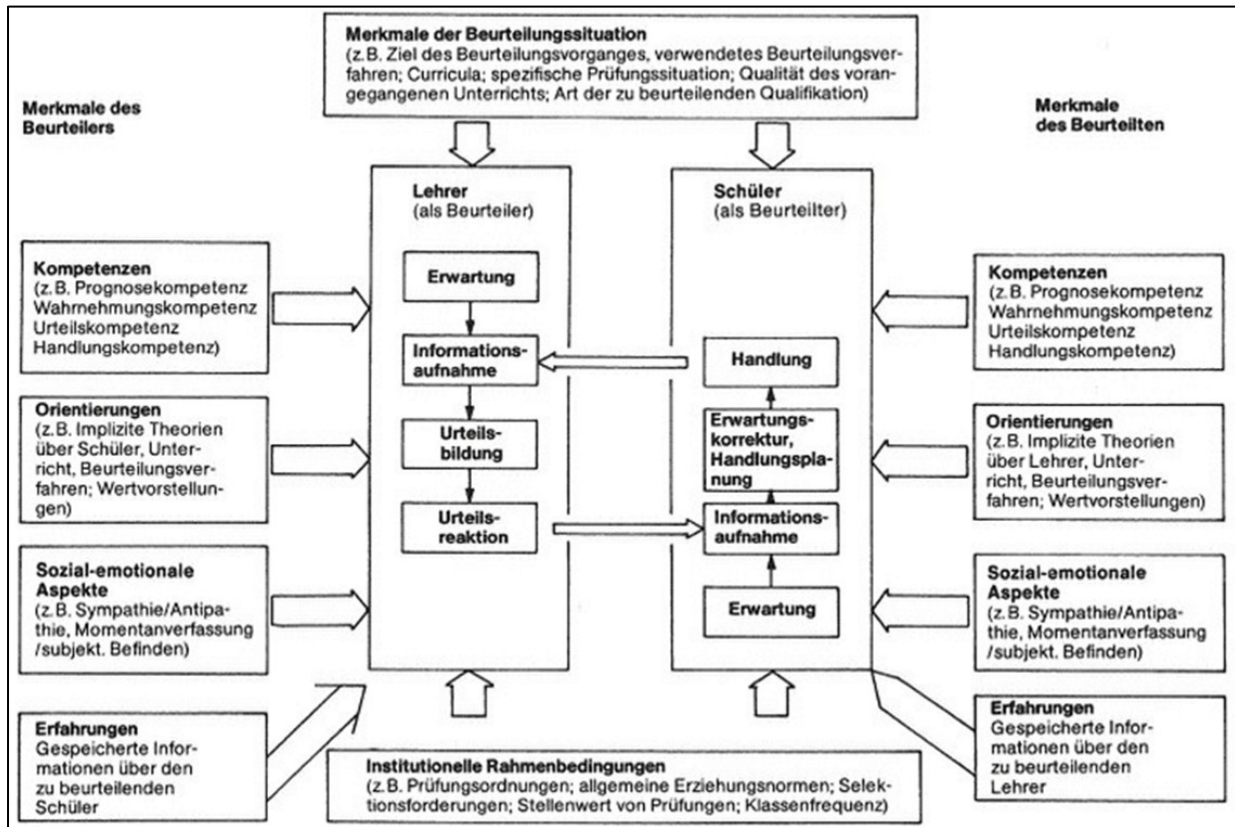
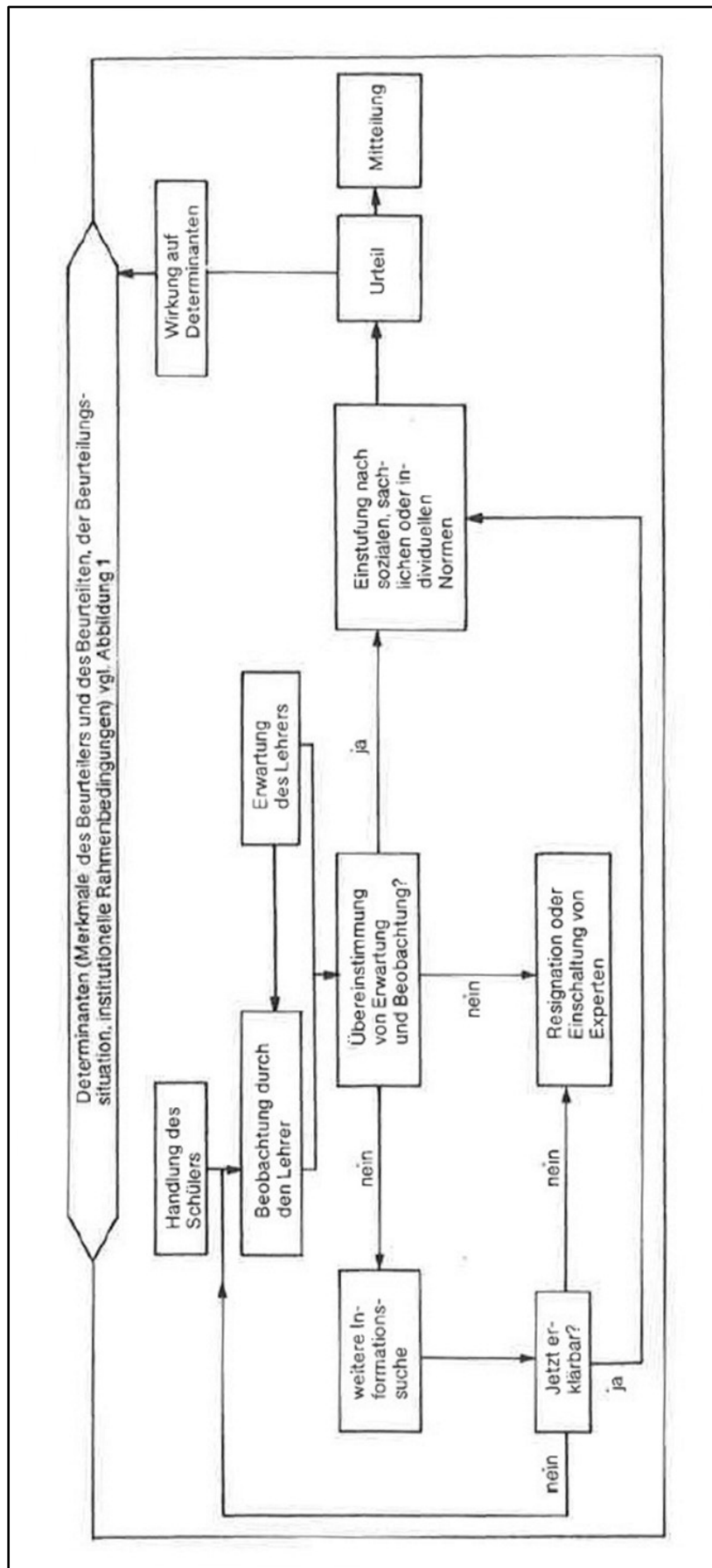


Abbildung 4. Integriertes Prozess- und Strukturmodell der Diagnostik von Ingenkamp und Lissmann (2008).



individueller Ebene der Lehrkraft von In

Das Strukturmodell benennt Merkmale der Beurteilenden und der Beurteilten (Kompetenzen, Orientierungen, sozial-emotionale Aspekte, Erfahrungen) sowie Merkmale der Beurteilungssituation (Situationsmerkmale, institutionelle Rahmenbedingungen). Zu den vorgeschlagenen Kompetenzen gehören Wissen und Fähigkeiten im Vergleichen, Analysieren, Prognostizieren, Interpretieren, Kommunizieren und Testen, d. h. sowohl kognitive als auch handlungspraktische Aspekte. Das Wissen umfasst dabei methodisches Wissen zum Messen und Auswerten mithilfe verschiedener Verfahren (z. B. Tests, Beobachtungen), Besonderheiten der Schulleistungsmessung, aber auch Wissen zur Diagnostik von Lernstörungen im fachbezogenen Bereich (Lese-Rechtsschreibstörung, Intelligenzminderung) und im sozialen und emotionalen Bereich (z. B. Aggression, Klassenklima, Motivation, Selbstkonzept, Prüfungsangst), sowie zu institutionsspezifischen Statusveränderungen in Form von Übergangsempfehlungen. In Bezug auf den kommunikativen Aspekt des Modells wird auch Wissen zu Beratung im Bildungswesen, zu repräsentativen Schulleistungstudien (z. B. PISA) und zu administrativen Prozessen (z. B. Erstellung eines Gutachtens mit Sonderpädagoginnen und -pädagogen) spezifiziert.

Das theoretische Modell von Ingenkamp und Lissmann (2008) ist nicht als Kompetenzmodell angelegt, beschreibt aber alle notwendigen Elemente zur Definition einer Kompetenz. Im Sinne einer kritischen Auseinandersetzung und Zusammenfassung kann es daher auch nicht an Kompetenzmodellen gemessen werden. So sind etwa weder eine systematische empirische Begründung oder Prüfung des Modells erfolgt noch eine Ausführung zu Kompetenzentwicklung oder Niveaustufen vorhanden. Dennoch bietet es für die Modellierung des (Erlernens) diagnostischen Handelns von Lehrkräften bedeutende theoretische Anregungen.

Die Besonderheit des Modells besteht in der Konzeption diagnostischen Handelns über die Individualebene von Lehrerinnen und Lehrern und über den Unterrichtskontext hinaus. Dass die individuelle Kompetenz einer Lehrkraft nicht immer den größten Effekt auf Ergebnisse diagnostischen Urteilens- und Handelns hat, geht bspw. auch aus Mehrebenen-Analysen und Meta-Analysen zu Urteilsgenauigkeit hervor (Machts et al., 2016; Südkamp et al., 2012; Paleczek et al., 2017). Die institutionellen Effekte auf Diagnostik liegen jedoch nicht zwingend außerhalb individueller Kompetenzmodelle. Diese implizite theoretische Annahme des Großteils bisher entwickelter Modelle diagnostischer Kompetenz ist eine Folge der Fokussierung auf die diagnostische Beziehung zwischen Lehrkraft und Schülerin oder Schüler (Edelenbos & Kubanek-German, 2004; Hellmann, 2015; Herppich et al., 2018; Hesse & Latzko, 2017; Horstkemper, 2004; Jäger, 2012; Jäger, Frey, Jäger-Flor, Lissmann, & Riebel, 2010; Karst, 2012; Klug, 2011; Leuders et al., 2018; Ostermann, 2018; Schrader, 2009; Streit et al., 2019; Südkamp & Praetorius, 2017; van Ophuyzen, 2006). Es existieren jedoch auch Anforderungssituationen auf anderen Beziehungsebenen, bspw. zwischen Lehrkräften untereinander, zwischen Lehrkräften und Fachkräften anderer Disziplinen (z. B. Psychologie, Sonderpädagogik, Sozialarbeit) sowie zwischen Lehrkräften und Eltern. Diese werden bei Ingenkamp und Lissmann (2008) nicht systematisch beschrieben aber durchaus angedeutet (z. B. S. 219, S. 225, S. 238). Es besteht daher eine dringende Notwendigkeit, diesen Aspekt für die Modellierung und Entwicklung diagnostischer Kompetenz aufzuarbeiten.

3.2 Erweiterung diagnostischer Kompetenz um kooperatives Handeln

Definitionen diagnostischen Handelns von Lehrkräften konzentrieren sich mehrheitlich auf die Interaktion von Lehrkräften mit Schülerinnen oder Schülern, die Unterrichtssituation und die Tätigkeiten der Vorbereitung und Auswertung von Unterrichts- und Diagnosematerial (Edelenbos & Kubanek-German, 2004; Hellmann, 2015; Herppich et al., 2018; Hesse & Latzko, 2017; Horstkemper, 2004; Jäger, 2012; Jäger, Frey, Jäger-Flor, Lissmann, & Riebel, 2010; Karst, 2012; Klug, 2011, Leuders et al., 2018; Ostermann, 2018; Schrader, 2009; Streit et al., 2019; Südkamp & Praetorius, 2017; van Ophuysen, 2006). Handlungsziele in diesen Definitionen beziehen sich auf die Erfassung und Förderung von Lernverhalten (Klug et al., 2013), die Einschätzung und Erfassung fachspezifischen Wissens (Leuders et al., 2018) und breiter gesehen auf die Identifikation und Modifikation von Lernvoraussetzungen und -umwelten (Ingenkamp & Lissmann, 2008). Kooperatives Handeln von Lehrkräften im Kontext von Diagnostik wird bislang jedoch weder in theoretischen Kompetenzmodellen noch in empirischer Lehrkräftebildungsforschung systematischer berücksichtigt. Ein erster Schritt der theoretischen und psychometrischen Modellierung dieses Aspekts diagnostischer Kompetenz ist eine Beschreibung der Anforderungssituationen kooperativen Handelns im Kontext von Diagnostik, um daraus die kognitiven Module und Strukturen abzuleiten, die die diesbezügliche Handlungsfähigkeit von Lehrkräften möglicherweise steigern.

Berliner (1985) beschreibt eine professionell Handelnde Lehrkraft als “[...] a decision maker who handles a complex set of interacting variables in a dynamic social environment” (S. 6). Die soziale Umwelt, in der Lehrerinnen und Lehrer Entscheidungen treffen, beschränkt sich keinesfalls nur auf den Unterricht. Spezifisch im Kontext der Diagnostik gibt es viele Situationen, in denen Arbeitsbündnisse auf gleichrangiger Ebene für Lehrkräfte notwendig sind wie z. B. Elterngespräche (Klug, Bruder, Keller, & Schmitz, 2012), Zusammenarbeit mit Sonderpädagoginnen und -pädagogen oder Psychologinnen und Psychologen (Jurkowski & Müller, 2018) sowie Fallkonferenzen in multiprofessionellen Teams (Hesjedal, Hetland, & Iversen, 2015). Letztere sind auch auf institutioneller Ebene von Bedeutung für den Transfer von Ressourcen und Informationen zwischen Bildungs- und Gesundheitssystem (Byrne, 2017). Zwar ist die Kommunikation über Handlungen und Ergebnisse im diagnostischen Prozess, die Ingenkamp und Lissmann (2008) in der Dimension professioneller Kooperation andeuten, in Bezug auf Frequenz, Gestaltungsformen und Ergebnisse nach wie vor nicht ausreichend untersucht, um bereits mit ausreichender Sicherheit von positiven Effekten auf Lern- und Leistungsmerkmale von Schülerinnen und Schülern sowie auf die Qualität einer Diagnostik ausgehen zu können (Brinkmann & Twiford, 2012). Behrmann und van Ophuysen (2017) vertreten jedoch bereits die Einführung von Multiperspektivität im Sinne einer fallbezogenen Validierung als Qualitätsmerkmal von Diagnostik.

Der Bedarf, Fähigkeiten der effektiven Leistungserbringung in kooperativen Beziehungen im Rahmen der Lehrkräftebildung zu entwickeln, ist aus einer Reihe empirischer Studien ableitbar. Für bereits beruflich aktive Lehrkräfte dokumentiert der Teaching and Learning International Survey (TALIS) einen seit mehreren Jahren unveränderten Bedarf an Anregung und Unterstützung für Kooperation im Kontext diagnostischen Handelns. So geben 3.5 % ($SD = 4.3$, $x_{min} = 0.4$, $x_{max} = 25.0$) der befragten Lehrkräfte an, nie in Diskussion über die Lernentwicklung ihrer Schülerinnen und Schüler zu treten und 8.8 % ($SD = 4.9$, $x_{min} = 1.3$, $x_{max} = 20.4$) geben an, nie mit anderen Lehrkräften an ihrer Schule

an dem Ziel zusammenzuarbeiten, gemeinsame Standards für die schulische Leistungsmessung herzustellen (OECD, 2014). Während 65 % der befragten Schulleiterinnen und Schulleiter die Zusammenarbeit beginnender und erfahrener Lehrkräfte in Mentoringbeziehungen als wichtigen Initiator von Kooperationsbeziehungen sehen, geben nur 22 % der beginnenden Lehrkräfte an, eine solche Beziehung in Anspruch nehmen zu können (OECD, 2018). Generell werden kooperativ angelegte schulinterne Lernarrangements weit weniger implementiert als externe Weiterbildungen im Kursformat, obwohl Lehrkräfte erstere als wirksamer in Bezug auf ihre professionelle Entwicklung einschätzen (OECD, 2018). Richter und Pant (2016) stellen in einer repräsentativen Befragung von 1015 Lehrerinnen und Lehrern der ersten Sekundarstufe in Deutschland fest, dass zwar eine sehr positive Einstellung zu kooperativer Zusammenarbeit im Schulkontext besteht, das Ausmaß von Kooperationsaktivitäten aber unterhalb des Durchschnitts der OECD-Länder liegt. In der Forsa-Studie der Deutschen Schulakademie (Pant, 2018) geben 57 % der 1016 befragten Lehrerinnen und Lehrer explizit an, in der Unterrichtsgestaltung eng mit anderen Kolleginnen und Kollegen zusammenarbeiten zu wollen. Von den verbleibenden 43 % der Lehrkräfte geben 13 % an, dass Teamsitzungen nicht konstruktiv verlaufen würden und 9 %, dass die Zusammenarbeit auf Grund unterschiedlicher Sichtweisen schwierig sei. Moldavsky, Pass and Sayal (2014) stellen fest, dass Lehrkräfte die Kooperation mit anderen Lehrkräften der Kooperation mit psychologischen Fachkräften vorziehen.

Eder, Dämon und Hörl (2011) weisen in einer Studie mit 537 Lehrkräften und 1113 Lehramtsstudierenden zu Prädiktoren von Kooperationsbereitschaft darauf hin, dass ein unter Lehrkräften gehäuft auftretendes Denk- und Handlungsmuster, die Vermeidung von Kooperation und der Ausbau von Autonomie, bereits im Studium vertreten wird. Dieses Autonomie-Paritäts-Muster (Lortie, 1972) bezeichnet die berufstypische Verhaltensnorm unter Lehrkräften, dass die professionelle Anerkennung untereinander durch Unabhängigkeit im Unterrichtshandeln entsteht. Kooperative Entscheidungen und Handlungen können vor dem Hintergrund dieses Handlungs- und Deutungsmusters zu einer Bedrohung der professionellen Identität führen. Eder, Dämon, & Hörl (2011) stellen fest, dass dieses Muster „bereits bei Studierenden als eine Bewältigungsstrategie fungiert, die Personen mit weniger guten personalen oder situativen Voraussetzungen dazu dient, sich im beruflichen Umfeld von anderen abzugrenzen“ (S. 15). Ein zentraler Befund ihrer Studie ist, dass ein Zusammenhang zwischen geringer Kooperationsbereitschaft mit einem ungünstigen Leistungsselbstkonzept und niedrigen Selbstwirksamkeitserwartungen besteht. Zu ähnlichen Ergebnissen kommen Rothland, Biderbeck, Grabosch und Heiligt (2018). In ihrer Längsschnittstudie mit 397 Lehramtsstudierenden besteht ein Zusammenhang zwischen der Nutzung praktisch-kooperativer Lerngelegenheiten und einer niedrigen Ausprägung des Autonomie-Paritäts-Musters.

Ein weiterer Befund bezieht sich auf die Stabilität universitär erworbener Kompetenzen im Übergang in reguläre Beschäftigung als Lehrerinnen oder Lehrer. Lehramtsstudierende, die sich in dieser Übergangssituation befinden, nehmen etablierte Normen und Routinen in Schulen als normative Handlungsrichtlinien wahr, was mitunter zu einer retrospektiven Entwertung und Einkapselung bereits erworbener Kompetenzen führt (Ghousseini & Herbst, 2016; Hoth et al., 2016). Ghousseini und Herbst (2016) stellen fest, dass ein diesbezügliches Motiv in der Konfliktvermeidung im Lehrkräftekollegium liegt, auf dessen Unterstützung Lehramtsstudierende tendenziell angewiesen sind. Auch Main (2007) identifiziert die frühe Auseinandersetzung mit Kooperationsformen in der

Lehrkräftebildung als Gelingensbedingung für den Aufbau effektiver Kooperationsbeziehungen an Schulen. Das Problem einer Nivellierung formal erworbener Kompetenzen gegenüber informellen und dadurch authentischen Lerngelegenheiten beschrieb auch L. Shulman (1992) in seinem Entwurf fallbasierten Lernens in der Lehrkräftebildung. Lehrkräfte stehen durch ihre Tätigkeit im Bildungssystem kontinuierlich im Spannungsfeld aktueller und angestrebter Bildungspraxis: „This tension becomes particularly acute when students are sent out as observers, interns, student teachers, or apprentices to work under real conditions“ (L. Shulman, 1992, S. 8).

Auch zur Nutzung und zum Nutzen kooperativen Handelns im Bereich der Diagnostik legen empirische Ergebnisse eine Erweiterung von Kompetenzmodellen um Kooperationsaspekte nahe. In der Studie von Richter und Pant (2016) stimmen 52 % der Lehrerinnen und Lehrer der Aussage voll und ganz zu, Hinweise auf Stärken und Schwächen einzelner Schülerinnen und Schüler von Kolleginnen und Kollegen zu erhalten. Auch geben 31 % bzw. 33 % an, fachbezogene bzw. fachübergreifende Abstimmungsprozesse zur Förderung einzelner Schülerinnen und Schüler durchzuführen. Die deskriptive Analyse von Richter und Pant (2016) zu Zusammenhängen zwischen diesen Verhaltensweisen und individuellen sowie strukturellen Variablen lässt positive Effekte diagnostikbezogener Kommunikations- und Kooperationsaktivitäten auf Lern- und Leistungsmerkmale von Schülerinnen und Schüler sowie auf die Gesundheit von Lehrkräften vermuten. So ist eine gesteigerte Häufigkeit dieser Verhaltensweisen auf individueller Ebene mit höheren Werten in der Selbstwirksamkeit, dem Enthusiasmus für das Unterrichten und der Berufszufriedenheit aber mit niedrigeren Werten in der emotionalen Erschöpfung verbunden. Zudem steigt die Häufigkeit des Kooperationsverhaltens im Kontext von Diagnostik mit zunehmendem Anteil an Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf. Auf struktureller Ebene bestehen sehr deutliche positive deskriptive Zusammenhänge mit einem inklusiven Schulangebot, einem Ganztagsangebot, dem Vorhandensein von Koordinationsstrukturen und der Unterstützung durch die Schulleitung. In der TALIS Studie geben 74 % derjenigen Lehrkräfte, die ihre Weiterbildungen als wirksam einschätzten, an, dass darin Gelegenheiten zum Aufbau von Kooperationsbeziehungen gegeben waren (OECD, 2018). Weitere Studien finden ebenfalls einen vergleichsweise starken Effekt kooperativ angelegter Lerngelegenheiten auf die professionelle Entwicklung (Avalos, 2011; Jensen, Sonnemann, Roberts-Hull, & Hunter, 2016; Kraft, Blazar, & Hogan, 2018; Instance & Paniagua, 2018).

Auf kognitiver Ebene sind ebenso Hinweise für die prinzipielle Effektivität kooperativen Vorgehens im diagnostischen Handeln vorhanden. Kognitionpsychologische Studien zur Urteils- und Entscheidungsbildung zeigen, dass eine Außenperspektive die Überschätzung der eigenen Fähigkeiten und andere Arten kognitiver Voreingenommenheit (Bias) reduziert (Gigerenzer, Hoffrage, & Kleinbölting, 1991). Weiterhin stellen zwei Meta-Reviews im Kontext medizinischer Diagnostik fest, dass kooperatives Handeln in Form von Fallbesprechungen einerseits die Rate korrekt-richtiger Urteile, aber andererseits auch die Rate falsch-positiver Urteile erhöht (Graber et al., 2012; McDonald et al., 2013). Letzteres dürfte jedoch dann ein akzeptabler Trade-Off sein, wenn Diagnostik einen expliziten Bezug zu Förderung aufweist. Darüber hinaus hat die sozialpsychologische Forschung eine Reihe förderlicher Effekte sozialer Gruppen auf Prozesse der Urteilsbildung identifiziert. Dazu gehören u. a. gegenseitige kognitive Anregung (cognitive stimulation: Nijstad & Stroebe, 2006) und eine erhöhte Anstrengung bzw. individuelle Investition (Hertel, Kerr, & Messé, 2000; Williams & Karau, 1991).

In Anschluss an diesen Forschungsstand lassen sich fünf zentrale Anforderungssituationen in Bezug auf Kooperation im diagnostischen Handeln ableiten. Da das Modell von Klug (2011) das bisher einzige zusammenhängend empirisch geprüfte Modell diagnostischen Handelns ist, werden die hier entwickelten kooperativen Anforderungssituationen systematisch darauf bezogen (siehe *Abbildung 6*).

Einer Diagnostik vorgeschaltet, aber dieser nicht zwangsläufig zugehörig, sind Prozesse des universellen Screenings, d. h. formeller und informeller Verfahren zur Identifikation von Personen, die von einer erwarteten Lern- und Leistungsentwicklung mit einer bestimmten Wahrscheinlichkeit abweichen. Dazu gehören etwa curriculumsbasierte Lernstandserhebungen in Schulklassen, die spontane Wahrnehmung von Lehrkräften inner- und außerhalb des Unterrichts wie auch klassen- oder schulweite Befragungen zu lern- und leistungsrelevanten Zuständen und Verhaltensweisen von Schülerinnen und Schülern. Kriterien zur Bestimmung eines diagnoserelevanten Zustands liegen für Lehrkräfte bislang nicht vor, sodass Lehrerinnen und Lehrer diesen vor dem Hintergrund ihrer professionellen und erfahrungsbezogenen Wissensbestände und Überzeugungen feststellen. Ist dieses individuelle anlassbezogene Kriterium erreicht, gehen Lehrkräfte die präaktionale Phase über. Diese wird hier in Anschluss an die in Kapitel 2 entwickelte Definition von Kompetenz als kognitiver Zustand konzipiert, der mit spezifischen diagnostischen Handlungen korrespondiert.

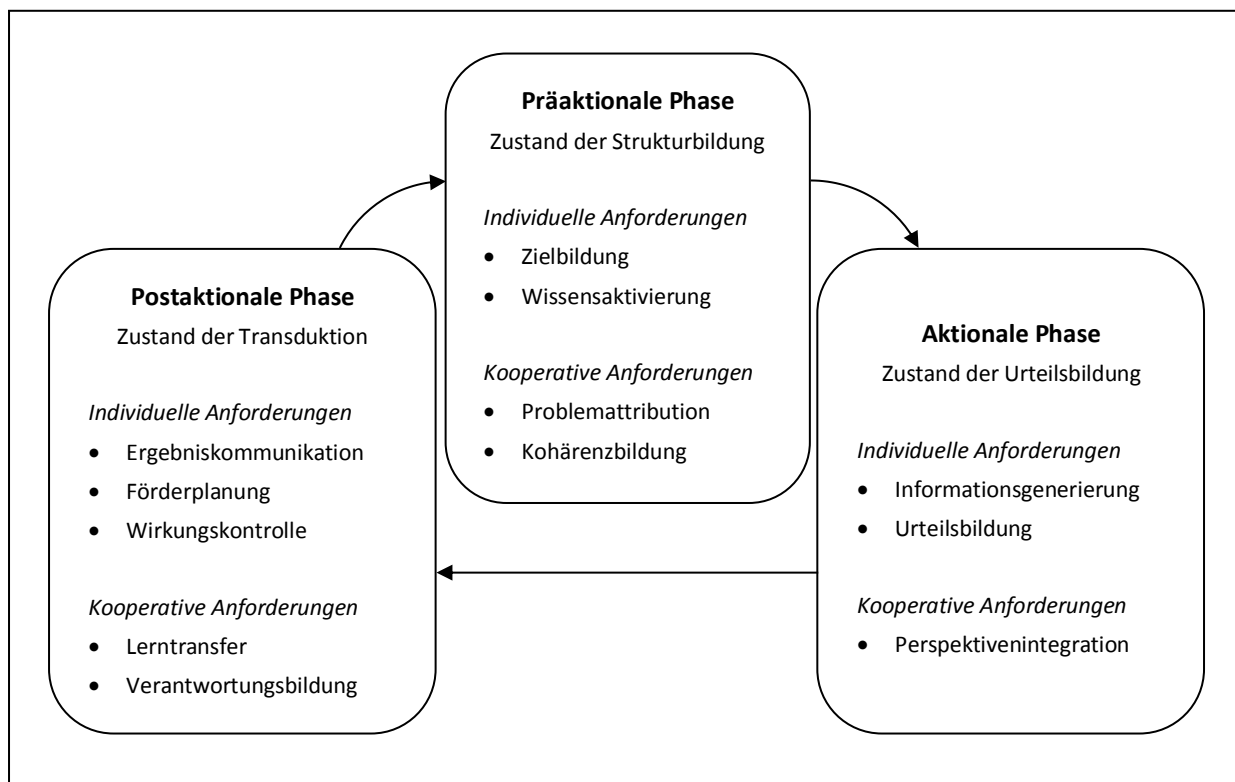


Abbildung 6. Diagnostische Anforderungssituationen für Lehrkräfte. Eigene Darstellung.

Die präaktionale Phase beinhaltet Planungs- und Entscheidungsprozesse zur Strukturierung der Problemlage und zur Festlegung erster Schritte der Informationsgenerierung. Eine erste notwendige kooperative Anforderung besteht dabei in der Eingrenzung des Problems auf eine kontextuelle oder

personenbezogene Symptomatik. Manifestierte Einschränkungen im Lern- und Leistungsverhalten können ihre Ursache in der schulischen Lernumwelt oder in den persönlichen Merkmalen bzw. dem sozialen Hintergrund von Schülerinnen und Schülern haben. Da ein Kontextfaktor die pädagogische Beziehung zwischen Lehrerin oder Lehrer und Schülerin oder Schülern ist, kann diese Strukturierungsleistung nur durch Informationen erbracht werden, die außerhalb dieser pädagogischen Beziehung liegen. Daher besteht eine Anforderung darin, diagnoserelevante Informationen anderer Fachkräfte einzuholen und im Umkehrschluss eigene Informationen über Schülerinnen und Schüler für das diagnostische Handeln anderer Lehrkräfte nutzungsbereit zur Verfügung zu stellen. Dies kann in unterschiedlichem Format erfolgen, bspw. über Fallkonferenzen oder aktenförmige Dokumentation. Zweck dieses kooperativen Handelns ist die Reduktion falscher Attributionen von Problemen auf Schülerinnen und Schüler (Problemattribution).

Eine weitere mögliche, aber nicht notwendige, Anforderung in der präaktionalen Phase betrifft die Planung kontextübergreifender diagnostischer Schritte. Liegt ein personenbezogenes Problem vor, das für das Lernen in verschiedenen Unterrichtsfächern von Bedeutung ist, kann die Abstimmung zwischen Lehrkräften über ein kohärentes Vorgehen zu Effizienz und Erfolg der Diagnostik führen. So können Lehrkräfte etwa aufeinander abgestimmte Handlungsschritte zur Informationsgenerierung einplanen, um den Zeitraum der Diagnostik zu verkürzen sowie die Lehr- und Lernprozesse in einzelnen Unterrichtsfächern nicht zu sehr durch die Diagnostik einzuschränken. Notwendig dafür ist die Festlegung einer gemeinsamen Zielstellung, die Identifikation jeweiliger Ressourcen und Verantwortungen zur Informationssammlung sowie die Festlegung von Verfahren zur Integration von Informationen und zur Entscheidungsfindung. Für die organisatorischen Aspekte kann dies bspw. in Form eines Plenums oder einer autoritativen Struktur erfolgen und für die inhaltlichen Aspekte in Form normbezogener Kriterien. Wenn Diagnostik auf pädagogische Anschlusshandlungen zielt, beziehen diese kommunikativen Prozesse bereits Fördermöglichkeiten ein. Zweck dieses kooperativen Handelns ist die Herstellung eines kohärenten Diagnoseprozesses und ggf. die Herstellung von Voraussetzungen einer multiperspektivischen Urteilsbildung (Kohärenzbildung).

Ist das diagnostische Handeln abschließend geplant, tritt mit der aktionalen Phase der Zustand der Informationssammlung und Urteilsbildung ein. Die Forschung zu kognitiven Ansätzen der Modellierung diagnostischer Kompetenz zeigt, dass diese Prozesse auf individueller Ebene sowohl bei heuristischer als auch bei systematischer Informationsverarbeitung eng miteinander verbunden sind (Monteiro & Norman, 2013). Sind mehrere Personen in das diagnostische Handeln involviert oder wird eine multiperspektivische Urteilsbildung (Behrmann & van Ophuysen, 2017) verfolgt, entsteht eine weitere Anforderung. Diese betrifft die Herstellung eines diagnostischen Urteils, das die verschiedenen Perspektiven so integriert, dass eine kollektiv akzeptierte Handlungsgrundlage entsteht. Dafür ist die erneute Organisation eines Entscheidungsprozesses notwendig. Zweck dieses kooperativen Handelns sind die Reduktion von Urteilsfehlern und die Herstellung der Voraussetzungen einer kontextübergreifenden und -verzahnenden Förderung (Perspektivenintegration).

Nach der Urteilsbildung tritt mit der postaktionalen Phase der Zustand der Transduktion ein. Transduktion wird zur Beschreibung von Prozessen genutzt, die durch Übertragung von Informationen eine Aktivitätskette auslösen (causal-state construction; Shalizi & Crutchfield, 2002). Je nach Modellierung des Diagnoseprozesses umfasst der Anforderungsbereich in der postaktionalen Phase

verschiedene Prozesse, wie etwa individuelle Förderung (Hellmann, 2015; Klug 2011; Streit et al., 2019) oder Wirkungskontrolle, Überweisung an weitere Expertinnen und Experten oder Gutachtenerstellung (Ingenkamp & Lissmann, 2008). Während Förderung, Wirkungskontrolle und Gutachtenerstellung nicht zwingend kooperativ erfolgen müssen, ist Zusammenarbeit für eine nachhaltig erfolgreiche Überweisung notwendig. Der Vorgang des Überweisens scheint zunächst trivial, birgt aber neben der Kontaktvermittlung weitere Voraussetzungen für die Inanspruchnahme der zusätzlichen Expertise und für die rückführende Integration der mit dieser Expertise generierten Informationen. Dazu gehören die Auswahl einer passenden Kontaktstelle, die Motivierung zur weiterführenden Kontaktaufnahme und die Organisation eines Informationsrücklaufs. Zweck dieses kooperativen Handelns in der postaktionalen Phase ist die Herstellung einer Situation, in der eine Verantwortungsübernahme für die handlungspraktischen Konsequenzen des diagnostischen Urteils erfolgen kann. Dies kann sich auf Schülerinnen und Schüler selbst beziehen aber auch auf Eltern oder andere pädagogische, psychologische oder medizinische Fachkräfte (Verantwortungsbildung).

Eine weitere kooperative Anforderung bezieht sich auf die professionelle Weiterentwicklung. Die Forschung zu dieser Facette des Berufsfeldes zeigt, dass Lehrerinnen und Lehrer in ihrer Kompetenzentwicklung Settings bevorzugen und häufig wahrnehmen (z. B. in Form von Mentoring), in denen sie von ihren direkten Kolleginnen und Kollegen lernen können (OECD, 2018). Es ist somit durchaus vertretbar, die Vermittlung professioneller Lernergebnisse aus Diagnoseprozessen als notwendige kooperative Anforderungssituation zu konzipieren, mit der Lehrkräfte konfrontiert sind (Lerntransfer).

Brownell, Adams, Sindelar, Waldron und Vanhover (2006) weisen darauf hin, dass für gelingende Kooperationen kontextspezifische Fähigkeiten und Überzeugungen notwendig sind. Diesen Befund unterstützen auch Meta-Reviews zu professionellen Zusammenschlüssen von Lehrkräften, die sowohl positive Effekte auf die Entwicklung professioneller Handlungskompetenzen feststellen als auch negative Effekte im Sinne einer erhöhten Arbeits- und Stressbelastung (Vangrieken, Dochy, Raes, & Kyndt, 2015; Vangrieken, Meredeith, Packer, & Kyndt, 2017). Darüber hinaus weisen sozialpsychologische Studien auch auf Risiken der Zusammenarbeit in Gruppen hin, wie etwa die Begrenzung während der Urteilsbildung auf eine einzige dominante Perspektive (cognitive restriction: Kerr & Tindale, 2004; Produktionsblockierung: Diehl & Stroebe, 1987) oder die Verwendung ausschließlich allen Gruppenmitgliedern bekannter Informationen (information-pooling effect: Stasser & Titus, 1985). Um diese empirischen Befunde produktiv zu nutzen, benötigt die Lehrkräftebildung Modelle diagnostischer Kompetenz, die auch Fähigkeiten, Wissensbestände und Überzeugungen für kooperative diagnostische Handlungssituationen einbeziehen.

Da Kooperation auf konzeptueller Ebene eine neue diagnostische Anforderungssituation ist, sind spezifische Dispositionen diesbezüglich kompetenten Handelns noch nicht empirisch erforscht und es lassen sich an dieser Stelle lediglich theoretische Überlegungen dazu anstellen. Da kooperatives Handeln ein inter-individueller Realitätsausschnitt ist, kommen zur Modellierung relevanter individueller Dispositionen für diese Kategorie an Anforderungen nicht nur kognitionspsychologische Konstrukte wie Wissen oder Überzeugungen infrage, sondern auch sozial-kognitiv orientierte Fähigkeiten. Dazu gehören Konstrukte aus Theorien zur Identifikation mit sozialen Gruppen und zur Gruppenwahrnehmung wie Entitativität (Campbell, 1958) oder kollektive Selbstwirksamkeit

(Schwarzer & Schmitz, 1999), aus Theorien zur sozialen Wahrnehmung und Urteilsbildung wie Empathie (Preston & De Waal, 2002) oder Selbstregulation (Bandura, 1991) sowie aus Theorien zu interpersonaler Kommunikation und Persuasion wie Vertrauen (Krampen, Viebig & Walter, 1982) oder Konkordanz (Sepucha & Ozanne, 2010).

3.3 Zusammenfassung und Definition diagnostischer Kompetenz

Kompetentes diagnostisches Handeln ist notwendig, um lernwirksame Umwelten herzustellen und Schülerinnen und Schüler in ihren Lern- und Entwicklungsprozessen zu unterstützen. Die bisherige Forschung hat theoretische Modelle zu diesem Bereich entwickelt, die verschiedene Aspekte der funktionellen Verbindung kognitiver Module und Strukturen sowie deren Nutzen in diagnostischen Anforderungssituationen beschreiben. Als Gemeinsamkeiten der Modelle lässt sich eine Modellierung von Prozessen der Informationsverarbeitung, Urteilsbildung und Handlungssteuerung identifizieren. Fixpunkt aller Modelle ist die Vermittlung zwischen den individuellen Bewältigungsmöglichkeiten von Schülerinnen und Schülern und den strukturellen Bewältigungsanforderungen in Schule und Unterricht. Es bestehen jedoch Unterschiede zwischen den Modellen, die sich aus der inhaltlichen Bestimmung des Gegenstands der Diagnostik ableiten. Bei einem Fokus auf den fachlichen Lern- und Leistungsstand von Schülerinnen und Schülern zielt diagnostische Kompetenz auf die handlungspraktische Verbindung mit der fachdidaktischen Unterrichtsgestaltung (Edelenbos & Kubanek-German, 2004; Hellmann, 2015; Leuders et al., 2018; Ostermann, 2018; Streit et al., 2019). Liegt der Fokus auf allgemeinen Lernvoraussetzungen zielt das Konstrukt auf die Initiation oder Förderung von Prozessen der Persönlichkeitsentwicklung von Schülerinnen und Schülern (Klug, 2011). Diese Spaltung in theoretischen Ansätzen zur Modellierung diagnostischer Kompetenz ist auch Ausdruck der Verteilung dafür benötigter Dispositionen auf verschiedene Standardgruppen in den KMK-Standards der Lehrkräftebildung (KMK, 2004). Einen umfassenden theoretischen Entwurf legen lediglich Ingenkamp und Lissmann (2008) vor.

Auch die psychologische Forschung zu medizinischer Diagnostik hat theoretische Modelle hervorgebracht, die einige Anregung für die Erforschung und Entwicklung diagnostischer Kompetenz von Lehrkräften bereithalten. Neben einer Modellierung der Urteilsbildung in Zwei-Prozess Modellen sind auch zahlreiche Erkenntnisse zur Reduktion von Urteilsfehlern entstanden (Brush et al., 2017; Monteiro & Norman, 2013). Eine Facette, die trotz empirisch nachweislichem Bedarf in der Forschung zu diagnostischer Kompetenz von Lehrkräften nicht berücksichtigt wurde, ist das kooperative Handeln.

Diagnostisches Handeln mit verschiedenen Gruppen von Akteurinnen und Akteuren zu verknüpfen, kann die Qualität der Diagnostik und pädagogischer Anschlusshandlungen steigern. Es ist eine Frage des Standards bzw. der Qualitätskriterien im diagnostischen Prozess, inwiefern die Voraussetzungen zu kooperativem Handeln als notwendiger Bestandteil diagnostischer Kompetenz von Lehrkräften zu konzipieren sind. Die Forschung zum Lernverhalten von Lehrkräften legt nahe, dass schulinterne Kooperation beim Diagnostizieren ein wichtiger Schritt zur kontinuierlichen professionellen Weiterentwicklung in diesem Bereich ist. Auch programmatisch kann eine Ausweitung von Diagnostik um kooperatives Handeln einen Gegenpol zum sehr prävalenten Autonomie-Paritäts-Muster in den Denk- und Handlungsorientierungen der Lehrkräfte bilden. Daran schließt sich jedoch

eine Forschungslücke an. Dispositionen, die zu Kompetenz in kooperativen diagnostischen Anforderungssituationen führen sind bislang nicht untersucht. Auch, ob diese angenommenen Dispositionen eine Schnittfläche mit Dispositionen zur Handlungsfähigkeit in nicht-kooperativen Anforderungssituationen bilden, ist unbekannt. Dies betrifft insbesondere die Verarbeitung personenbezogener Informationen, die sowohl auf Ebene der pädagogischen Beziehung zwischen Lehrkräften und Schülerinnen und Schülern als auch auf Ebene der professionellen Beziehung zu anderen Fachkräften Teil des diagnostischen Prozesses ist. Die zentrale Bedeutung entsprechender Dispositionen unterstützt auch die Expertiseforschung. Krolak-Schwerdt und Rommer (2005) aber auch Krolak-Schwerdt, Böhmer und Gräsel (2012) zeigen, dass sich Lehrerinnen und Lehrer je nach Zielstellung eher an spezifischen Personeninformationen oder an sozialen Kategorien in der Urteilsbildung orientieren. Da diese Studien keine kompetenzbasierten Prädiktoren messen, bleibt die Erklärung dieses Performanzunterschieds durch Dispositionen diagnostischer Kompetenz noch offen.

Die folgende Definition diagnostischer Kompetenz, die auch Grundlage der vorliegenden Forschungsarbeit ist, greift die kognitiven und sozial-kognitiven Definitionsansätze auf und integriert sie mit Bezug auf die zuvor festgestellten theoretischen Eigenschaften des Kompetenzkonstrukts.

Diagnostische Kompetenz von Lehrkräften bezeichnet die funktionelle kognitive Konfiguration, die eine an Gütekriterien orientierte und effektive Messung individueller und gruppenbezogener Lern- und Leistungsmerkmale im Schulkontext, daran anschließende individuelle und kooperative Urteilsbildung sowie die darauf basierende Initiation und Wirkungskontrolle von Förder- und Selektionsmaßnahmen ermöglicht. Grundlage dieser Konfiguration sind kognitive und sozial-kognitive Module und Strukturen zur heuristischen wie auch systematischen Konstruktion merkmalsbezogener mentaler Entwicklungsmodelle von Schülerinnen und Schülern, zur zielbezogenen Auswahl und Vermittlung von Förder- und Selektionsmaßnahmen sowie zur kooperativen und problemorientierten Entscheidungsfindung im Einzelfall.

Diese Definition umfasst inhaltlich wie im Modell von Ingenkamp und Lissmann (2008) die Dimensionen der fachbezogenen und fachübergreifenden Lernvoraussetzungen, knüpft an die Prozessstrukturen des diagnostischen Handelns von Klug (2011) und der Urteilsbildung (Ingenkamp & Lissmann, 2008; Schrader, 2009) an und unterlegt letztere im Anschluss an Ostermann (2018) sowie Monteiro und Norman (2013) auf kognitiver Ebene mit mentalen Modellen zu Schülerinnen und Schülern. Damit liegt eine Definition vor, die sowohl den Prozess- als auch Produktcharakter diagnostischer Kompetenz in individuellen und kooperativen Anforderungssituationen auf kognitiver Ebene abbildet. Sie ist daher auch Ausgangspunkt der Konzeption einer Intervention zur Entwicklung diagnostischer Kompetenz im Lehramtsstudium.

4 Entwicklung diagnostischer Kompetenz mit problemorientiertem Lernen

Diagnostisches Handeln stellt Lehrkräfte vor eine Vielzahl komplexer Anforderungen. Diese bestehen unter anderem darin, Informationen aus unterschiedlichen Quellen wie Beobachtungen und Tests zu integrieren, parallel mehrere Hypothesen zu Ursachen von Lernschwierigkeiten zu verfolgen und gewonnene Erkenntnisse an Unterrichtshandeln anschlussfähig zu machen. Lernschwierigkeiten und -störungen haben eine recht stabile Prävalenzrate. Schwankungen können Fischbach et al. (2013) größtenteils durch die Veränderung diagnostischer Kriterien erklären. In ihrer für Deutschland repräsentativen Stichprobe von 2195 Mädchen und Jungen zwischen der zweiten und dritten Klassenstufe zeigten 23 % eine Lernschwäche, d. h. eine schulische Leistung unterhalb der Klassennorm und 13 % eine Lernstörung, d. h. zusätzlich eine Diskrepanz zwischen Minderleistung und Intelligenz von mindestens 1.2 Standardabweichungen (üblicherweise 1.5 Standardabweichungen). Diese Werte entsprechen bisherigen Studien im nationalen Kontext (Thomas, Schulte-Körne, & Hasselhorn, 2015; Moll, Kunze, Neuhoﬀ, Bruder, & Schulte-Körne, 2014), sowie im internationalen Kontext (Pullen, 2017; Shalev, Auerbach, Manor, & Gross-Tur, 2000). Die Stabilität von Lernschwierigkeiten ist u.a. auf eine soziale Heritabilität kognitiver Basisfähigkeiten zurückzuführen, bspw. von über 60 % für Lesefähigkeit (Schulte-Körne, 2014; Moll, Wallner, & Landerl, 2012; Linkersdörfer, Lonnemann, Lindberg, Hasselhorn, & Fiebach, 2012). Verhaltensstörungen wie oppositionelles Trotzverhalten, Störungen des Sozialverhaltens und AD(H)S treten mit 7-9 % etwas geringer auf als oben benannte Entwicklungsstörungen schulischer Fähigkeiten (Linderkamp & Grünke, 2007), Angststörungen mit etwa 20 % jedoch deutlich häufiger (Fehm & Fydrich, 2011; Linderkamp & Grünke, 2007).

Lernschwierigkeiten und -störungen können gravierende Folgen für die Bildungsbiografie haben. So fanden Esser und Schmidt (1993) eine Übergangsquote Betroffener von Grundschule zu Gymnasium von nur 3 %, von 25 % für den Übergang an eine Realschule, von über 50 % an eine Hauptschule und etwa 17 % an eine Förder- bzw. Sonderschule. Ein segregierender Umgang des Schulsystems mit Personen, die außerhalb der ‚normalen‘ Lernentwicklung liegen (bspw. Aufteilung auf leistungshomogene Schultypen), ist bereits umfassend kritisiert worden (im Überblick: Solga & Dombrowski, 2009). Im deutschen Bildungssystem ist insbesondere der Übergang von der Grundschule zur ersten Sekundarstufe ein kritischer Punkt in der Bildungsbiografie, da die Schulform als differenzielles Entwicklungsmilieu ein wesentlicher Prädiktor für Lernzuwächse ist (Baumert, Köller, & Schnabel, 2000; Neumann et al., 2007; Szczesny & Watermann, 2011; Tischler, Daseking, & Petermann, 2013), da Lehrkräfte in ihrem übergangsbezogenen Urteil anfällig für nicht leistungsbezogene Merkmale von Schülerinnen und Schülern sind (z. B. soziale Herkunft) (Gräsel, Krolak-Schwerdt, Nölle, & Hörstermann, 2010; Riek & van Ophuysen, 2016) und ein späterer Wechsel zwischen den Schulformen mit großem Aufwand verbunden ist (Ditton, 2013). Esser, Wyschkon und Schmidt (2002) zeigten eine erhöhte Quote späterer Arbeitslosigkeit für Schülerinnen und Schüler mit umschriebener Entwicklungsstörung schulischer Fähigkeiten und auch Thomas et al. (2015) sowie Petermann (2018) gehen von langfristigen negativen Effekten auf Schulerfolg und berufliche Integration aus. Eine diagnosebezogene Förderung kann diese Effekte deutlich abpuffern. Meta-Analysen und Reviews zeigen, dass sich ein Interventionserfolg insbesondere in den Leistungsbereichen Lesen, Schreiben und

Rechnen dann einstellt, wenn zuvor eine individuelle Diagnostik erfolgt ist (Ise, Dolle, Pixner & Schulte-Körne, 2012; Ise, Engel & Schulte-Körne, 2012; Lindberg, Hasselhorn, & Lonnemann, 2018). Weitere positive Effekte akkurater Diagnostik auf verschiedene Lern- und Leistungsmerkmale sind in Kapitel 3.1 aufgeführt. Eine nicht frühzeitig erkannte Lernschwierigkeit oder -störung ist dagegen ein starker Risikofaktor für eine Chronifizierung (Linderkamp & Grünke, 2007).

Ein Modell zur Einordnung der Befundlage zu Lernschwierigkeiten und -störungen sowie positiven Effekten kompetenten diagnostischen Handelns bietet das Vulnerabilitäts-Stress-Modell aus der klinischen Psychologie (Hankin & Abela, 1999). Die Entwicklung psychologischer Einschränkungen ist darin als stufenförmiger Prozess konzipiert, dessen Eskalation (manifestierte Symptomatik) durch Schutzfaktoren (z. B. soziale Unterstützung, meta-kognitive Strategien, Emotionsregulation) verhindert oder verlangsamt, und durch Risikofaktoren (z. B. sozialer Stress, selbstwertschädlicher Attributionsstil, Reagibilität) ausgelöst oder beschleunigt werden kann. Demnach kann Diagnostik dazu beitragen, Risiko- und Schutzfaktoren zu identifizieren, um Stress zu verringern oder Stressursachen aufzulösen.

Mit dem Bewusstsein um die Befunde zur alltäglichen und biografischen Bedeutung differenzieller Lernvoraussetzungen zeichnet sich in den letzten Jahren ein schulkultureller Wandel hin zu einem inklusivem Bildungssystem ab. Dieser äußert sich nicht nur in Schulstrukturen (z. B. Einrichtung der Integrierten Sekundarschule), sondern wird von einer bildungspolitisch motivierten Qualitätsentwicklung von Unterricht begleitet, wie etwa durch nationale und internationale Vergleichsarbeiten zum Erreichen von Bildungsstandards (z. B. PISA, VERA) und andere wirkungsbasierte Evaluationsverfahren (zusammenfassend: Klieme et al., 2003; KMK, 2006; Koch & Gräsel, 2004). Diese Veränderungen ziehen einen immensen Weiterbildungsbedarf für Lehrkräfte nach sich, wie auch einen Bedarf an nachweislich wirksamen Möglichkeiten zur Entwicklung diagnostischer Kompetenz im Lehramtsstudium (Ricken, 2017). Die diagnostischen Anforderungen an Lehrkräfte spitzen sich darüber hinaus durch die spektrumsorientierte Neuausrichtung des psychologischen Klassifikationssystems im DSM-V zu, das nicht mehr zwischen Lernstörungen und -schwierigkeiten unterscheidet und nicht länger das Intelligenz-Diskrepanz Kriterium zur Vergabe einer Diagnose beibehält, wohl aber das Klassennorm-Diskrepanz Kriterium (Schulte-Körne, 2014). Informationen über Schülerinnen und Schüler seitens der Lehrkräfte bekommen dadurch ein deutlich größeres Gewicht in diagnostischen und lerntherapeutischen Entscheidungen. Damit steigt die Relevanz einer akkuraten, effektiven und verantwortungsvollen Diagnostik durch Lehrkräfte deutlich an.

Die Expertiseforschung zeigt, dass ein hohes Kompetenzniveau durch die Integration von einerseits theoretischem und fallbezogenem Wissen und andererseits deklarativem und prozeduralem Wissen erreicht wird (Brush et al., 2017). D. h. zugleich, dass funktionale kognitive Strukturen bei hohem Kompetenzniveau hochgradig individuell sind. Wenngleich zum Erreichen dieses Niveaus in aller Regel eine jahrelange gezielte Auseinandersetzung mit Diagnostik in verschiedenen Situationen notwendig ist, hat die Lehrkräftebildung die Chance, diesen Prozess nicht nur in Form von Weiterbildungen zu begleiten, sondern auch bereits frühzeitig auf die individuelle Entwicklung zugrundeliegender kognitiver Strukturen hinzuwirken. Die Herausforderung besteht dabei in der Bereitstellung einer Umwelt, die eine Ausbildung grundlegender kognitiver Strukturen so anregt, dass innerhalb dieser Rahmenbedingungen selbstgesteuerte Lern- und Entwicklungsprozesse erfolgen können, die eine

individuelle Verknüpfung verschiedener kognitiver Dispositionen, wie z. B. Wissen und Überzeugungen zulassen. Darüber hinaus können Lerngelegenheiten, bzw. Interventionen dann auch nachhaltig wirken, wenn die Form der Kompetenzentwicklung an spätere professionelle Lernprozesse anschlussfähig ist, d. h. eine Ähnlichkeit zu weiteren Lerngelegenheiten im Arbeitsleben besteht (Zlatkin-Troitschanskaia, Shavelson, & Kuhn, 2015).

In der Konzeption von Interventionen bzw. Lernumwelten für die Entwicklung diagnostischer Kompetenz kann das Prinzip der Evidenzbasierung viele Designentscheidungen erleichtern. Evidenz im Sinne wissenschaftlicher Forschungsergebnisse hat nach dem theoretischen Modell von Sackett, Rosenberg, Gray, Haynes und Richardson (1996) die Funktion, ‚klinische‘, d. h. durch kontextspezifisches Handeln erworbene Expertise zu ergänzen und zu kontrollieren. Die Lehrkräftebildung kann durch Umsetzung dieses Prinzips nicht nur auf institutioneller Ebene die Qualität und Effektivität der Vermittlung von Kompetenzen erhöhen, sondern gleichermaßen auf individueller Ebene als Modell für angehende Lehrkräfte wirken (Darling-Hammond, 2017). Gerade im Bereich diagnostischen Handelns, in dem Evidenzbildung eine Handlungsmaxime ist, liegt eine Orientierung am Modell von Sackett et al. (1996) nahe. Für die Entwicklung von Interventionen bedeutet das, Konzepte sowohl auf lernbezogenen Theorien und Forschungsergebnissen aufzubauen als auch hinsichtlich ihrer Lerneffekte empirisch zu prüfen. Diese Form der empirisch begründeten Interventionsentwicklung kann so zur Herstellung einer belastbaren Entscheidungsgrundlage für konkrete Bildungsplanung beitragen.

Nach wie vor weisen eine größere Anzahl bisher entwickelter evidenzbasierter Interventionen zur Entwicklung diagnostischer Kompetenz jedoch einige Merkmale auf, die das Zusammenspiel aus Anforderungen der Schulpraxis und Bedingungen des Lehramtsstudiums nicht ausreichend berücksichtigen. Beispielsweise basieren entgegen empirischen Befunden zum notwendigen Einbezug kooperativen Handelns in diagnostische Tätigkeits- und Lernprozesse (siehe Kapitel 3.2) bisherige Interventionen größtenteils auf individuell-kognitiven Lernmodellen mit Fokus auf die Entwicklung von Dispositionen für den diagnostischen Kontext der Beurteilung von Schülerinnen und Schülern im Unterricht (Klug, 2011; Krammer, 2014; Santagata & Guarino, 2011; Seidel, Blomberg, & Stürmer, 2010; Trittel, Gerich, & Schmitz, 2014; kritisch: Stahnke, Schueler, & Roesken-Winter, 2016; Shavelson, 2013). Forschung und Entwicklung zu stärker konstruktivistisch und kooperativ basierten Interventionen wurde bereits vielfach eingefordert (Buholzer & Zulliger, 2013; Ghousseini & Herbst, 2016; Hoth et al., 2016; Santagata & Yeh, 2016; Schrader, 2017). Dagegen zeichnet sich ein grundlegender Fallbezug in Interventionen als wirkungsvolles Element für Kompetenzentwicklung ab (Krumschick et al., 2017; J. H. Shulman, 1992; Syring et al., 2015; Yadav et al., 2011), was an die Expertiseforschung und das Modell evidenzbasierter Praxis von Sackett et al. (1996) anschließt. In bisher entwickelten Interventionen werden zwar ‚Fälle‘ verwendet, diese unterscheiden sich jedoch drastisch voneinander. Da häufig keine theoretische Bestimmung dessen erfolgt, was ein Fall ist, bleibt offen, inwiefern das verwendete Lernmaterial nicht nur diagnoserelevante Inhalte anbietet, sondern auch Anforderungen bereithält, deren Struktur eine Ähnlichkeit zu kognitiven und kommunikativen Handlungen im diagnostischen Prozess hat. Die vorliegenden Studien greifen ebenfalls auf Fälle als Lern- und Testmaterial zurück. Daher erfolgt zunächst eine problemtheoretische Bestimmung dessen, was einen diagnostischen Fall ausmacht.

4.1 Diagnostische Fälle aus problemtheoretischer Perspektive

In der psychologischen Forschung haben Probleme eine soziale und eine epistemische Funktion. Einerseits sind Personen in ihrem Alltag beständig mit Problemen konfrontiert. Dieser Zustand ist so ubiquitär, dass ein gesamtgesellschaftliches Interesse an der Erforschung von Möglichkeiten zur Problemlösung besteht (Gagné, 1980). Das Gesamt an therapeutischen, d. h. helfenden Professionen geht in der Praxis von wahrgenommenen Problemen aus und nutzt dabei bestimmte Fähigkeiten zur Problemlösung (Oevermann, 1996; Staub-Bernasconi, 1983). Andererseits eröffnen Problemsituationen die Möglichkeit, grundlegende kognitive und soziale Strukturen zu erforschen. Letztere werden im Sinne eines soziologischen Krisenexperiments (Garfinkel, 1967) besonders dann evident, wenn ihre gesellschaftstypische Funktion und ihr alltäglicher Ablauf unterbrochen werden. Im Kontext der Diagnostik überschneiden sich analog dazu soziales und epistemisches Handeln, da es beim Diagnostizieren nicht nur darum geht, eine Problemlösung anzustoßen, sondern auch darum, kausale Zusammenhänge in der Problemstellung aufzudecken, um letztlich präventiv handeln zu können.

Psychologische Definitionen von Problemen stellen zwei konstitutive Merkmale dieses Untersuchungsgegenstands heraus (Jonassen, 2000; Newell & Simon, 1972; van Merriënboer, 1997). Erstens ist ein Problem ein Konstrukt, dass die Differenz eines gegenwärtigen Zustands und eines Zielzustands beschreibt. Zweitens muss ein überzeugungs- oder wertgesteuertes Interesse an der Erreichung des Zielzustands bestehen. Im Schulkontext kann damit bspw. die Symptomatik einer Lernschwierigkeit als Problem beschrieben werden, aber auch fehlendes kooperatives Handeln zur Förderung von Schülerinnen und Schülern. Der ontologische Charakter des Konstrukts, d. h. ob Probleme auch ohne ihre Wahrnehmung existieren, ist eine philosophische oder evolutionsbiologische Frage und damit nicht Gegenstand dieser Forschungsarbeit. Die vorliegende Arbeit nimmt an dieser Stelle eine kognitionspsychologische Perspektive ein, in der Wahrnehmungs- und Verarbeitungsprozesse genuiner Bestandteil von Problemen sind und bezieht sich daher auch nur auf Probleme, für die diese Annahme zutrifft.

Newell und Simon (1972) gehen in ihrer Auseinandersetzung mit Problemen und Prozessen der Problemlösung von den Begrenzungen des kognitiven Apparates von Personen aus. Da im Arbeitsgedächtnis nur eine bestimmte Menge von Informationen bearbeitet werden kann (auch wenn diese als ‚chunks‘ in hierarchischer Struktur gespeichert sind) haben Probleme nicht zwangsweise einen definitiven Zustand. Vielmehr sind Probleme in erster Linie eine ausschnittshafte kognitive Repräsentation einer komplexeren Realität in einer bestimmten Situation. Diese kognitive Repräsentation bezeichnen Newell und Simon (1972: 788) als „problem space“, in dem Hinweisreize der Umwelt und Gedächtnisinhalte miteinander verknüpft sind. Diese können unterschiedlicher Art sein, wie z. B. physikalische Eigenschaften, Wertvorstellungen, Wissensbestandteile oder Handlungsabläufe. Elstein, Shulman und Sprafka (1978) sowie Goldschmidt (1997) präzisieren dieses Modell dahingehend, dass im Problemraum sowohl der gegenwärtige als auch mehrere zukünftige Zustände in Wahrscheinlichkeitsbeziehungen zueinanderstehen.

Die kognitive Konstruktion eines Problemraumes erfolgt in mentalen Modellen (Newell & Simon, 1972; Kim, 2015). Mentale Modelle sind multimodale kognitive Artefakte des Gedächtnisapparates zur

Speicherung situationsbezogener Informationen (Jonassen & Henning, 1999). In der Forschung zu mentalen Modellen bzw. zu kognitiven Repräsentationen wird konsistent eine Drei-Ebenen-Struktur festgestellt (Kintsch & van Dijk, 1978; Johnson-Laird, 1983). Auf der ersten Ebene, der Oberfläche, sind saliente bzw. relevante Aspekte und Objekte einer Situation gespeichert, die als deklaratives Wissen abrufbar sind. Auf der strukturellen bzw. propositionalen Ebene sind diese Elemente durch kausale und assoziative Verbindungen zu einem kohärenten Netzwerk verbunden. Auf der semantischen Ebene bzw. im Situationsmodell wird dieses Netzwerk mit Informationen aus dem Langzeitgedächtnis angereichert, so dass persönliche Bedeutungen generiert werden können. Ein für die Kompetenzentwicklung bedeutsamer Befund ist, dass mentale Modelle von Expertinnen und Experten unabhängig von der Domäne deutlich stärker mit Kausalbeziehungen und relevanten Situationsabläufen angereichert sind, d. h. umfangreichere Situationsmodelle konstruiert und zu fallbezogenem Wissen aggregiert werden (Bradley, Paul, & Seeman, 2006; Brush et al., 2017; Burkhardt, Détienne, & Wiedenbeck, 1997; Doane, Sohn, & Jodlowski, 2004; Hirschfeld & Gelman, 1994). Daran anschließend stellt die vorliegende Forschungsarbeit auch eine Studie zum Effekt von textbasierten mentalen Modellen Lehramtsstudierender auf die diagnostische Urteilsgenauigkeit vor (Kapitel 10).

Vor diesem Hintergrund lässt sich der kognitive Prozess des Problemlösens als eine Veränderung mentaler Modelle und damit als Veränderung von States bzw. kognitiven Zuständen beschreiben (Goldschmidt, 1997; Jonassen, 2000). Goldschmidt (1997, S. 442) beschreibt diesen Prozess als diskrete Veränderung, in der vom Initialzustand ausgehend mehrere zwischengelagerte Zustände durchlaufen werden, bis ein Zielzustand erreicht ist. Dazu bilden Personen einerseits Kriterien mit Schwellenwerten, die den Wechsel zu einem nächsten Zustand und das Erreichen des Zielzustands markieren (z. B. Kohärenz, kognitive Geschlossenheit). Andererseits verfügen Personen über Operatoren, die sie auf ihre mentalen Modelle anwenden können, um einen Zustandswechsel auszulösen (z. B. Entscheidungsregeln, Heuristiken, Bewertungen). Neben diesen internalen Konstrukten ist auch die Anwendung externaler Operatoren, wie etwa Dokumentation oder Kommunikation möglich. Jonassen (2000, S. 65) geht ebenfalls davon aus, dass Fähigkeiten zur Problemlösung an ihrem Effekt auf Ausbildung und Manipulation des Problemraumes gemessen werden können. Er geht jedoch nicht von einem diskreten, sondern von einem kontinuierlichen Veränderungsprozess aus. Elstein, Shulman und Sprafka (1978, S. 21) beschreiben ein weiteres, alternatives Modell des kognitiven Problemlösens. Sie gehen davon aus, dass Personen im Verlauf der Manipulation des Problemraumes gleichzeitig mehrere Modelle möglicher Zielzustände entwerfen, deren Diskrepanz vom Ausgangszustand bewerten und darauf basierende Wahrscheinlichkeiten der Zielerreichung schätzen. Diese bilden dann die Grundlage weiterer Manipulationen des Problemraums. Dabei ist es eine Frage der Strategie, ob eine Wahrscheinlichkeit maximiert und andere gleichzeitig minimiert oder ob alle Wahrscheinlichkeiten möglichst hochgehalten werden. Mit dieser Optimierungsfrage beschäftigen sich Elstein, Shulman und Sprafka (1978) jedoch nicht weiter. Es ist vorstellbar, dass in verschiedenen Situationen jeweils andere Strategien effizient sind, bspw. abhängig vom zeitlichen Handlungs- und Entscheidungsdruck.

Dass Personen während des Problemlöseprozesses qualitativ unterschiedliche kognitive Zustände durchlaufen, ist experimentell bereits mehrfach gezeigt worden (Alexander & Murphy, 1998; Chen &

Siegler, 2013; Kim, 2015; Krawczyk, 2012). Welcher Art diese Zustände sind und wie ihre Veränderung charakterisiert werden kann, ist dagegen weniger gut untersucht (Kim, 2015). Elstein, Shulman und Sprafka (1978) identifizierten für den Kontext der medizinischen Diagnostik folgende kognitive Operatoren, die zu einer effizienten und akkuraten Diagnose führten: (1) Signalidentifikation und -extraktion aus der Umwelt, (2) frühe Hypothesenbildung, (3) Signalinterpretation und (4) Hypothesenevaluation. In ihrer empirischen Studie stellten sie jedoch ebenfalls fest, dass diese Fähigkeiten sowohl fallspezifisch (z. B. in Bezug auf bestimmte Symptomatiken) als auch kontextspezifisch (z. B. in Bezug auf die Umstände der Diagnostik) unterschiedlich ausgeprägt waren. Darauf aufbauend können Problemlösefähigkeiten als alle kognitiven und sozialen Operationen definiert werden, deren Aktivität in einer Manipulation des Problemraums resultiert. In Kapitel 3.1 wurde bereits diskutiert, dass kontextspezifisches und fallbezogenes Wissen die kognitive Modellierung diagnostischer Probleme durch die Bildung von ‚illness-scripts‘ erleichtern und damit auch aus problemtheoretischer Perspektive ein wesentlicher Bestandteil diagnostischer Kompetenz sind.

Die Anwendbarkeit und damit auch Entwicklung von Fähigkeiten zur kognitiven Modellierung hängen jedoch von der Art des Problems ab. Jonassen (2000) entwirft eine Typologie, die Probleme auf zwei Achsen verortet. Die Strukturiertheit gibt an, wie explizit definiert Initial- und Zielzustand im Problemraum sind. So zeichnen sich ‚ill-structured‘ Probleme durch die Unbestimmtheit und Ambiguität dieser Zustände aus und ‚well-structured‘ Probleme durch ihre vollständige Geschlossenheit. Problemräume unterscheiden sich im Grad ihrer Komplexität, d. h. durch die Anzahl konstitutiver Elemente, durch die Art und das Ausmaß ihrer Verbindungen untereinander sowie durch ihre jeweilige Stabilität. Für Diagnostik besonders relevante Problemtypen sind (1) Diagnose-Lösungs-Probleme, (2) Entscheidungsprobleme und (3) Fallanalyse-Probleme. Sie zeichnen sich durch einen geringen Grad an Strukturiertheit und eine hohe Instabilität der Problemelemente aus. Entsprechend hoch sind die Anforderungen an die Problemlösefähigkeiten. In Diagnose-Lösungs-Problemen und Fallanalyse-Problemen sind weder der Initial- noch der Zielzustand durch die Umwelt definiert und zur Konstruktion des Initialzustands müssen bereits situationsexterne Informationen einbezogen werden. In der vorliegenden Forschungsarbeit wurde ein Kategoriensystem zur Fallentwicklung konzipiert und validiert, das in der Studie in Kapitel 8 genauer vorgestellt wird.

Probleme mit hohem Komplexitätsgrad wurden in ihrer Eignung für die institutionelle Nutzung zu Lernzwecken im Bildungssystem auch kritisch betrachtet, besonders aus Perspektive der Cognitive Load Theorie (Sweller, 1994). Beispielsweise geht Renkl (2014) davon aus, dass es insbesondere in frühen Phasen der Kompetenzentwicklung zu informationeller Überlastung und damit eingeschränkten Lerneffekten kommen kann. Eine Meta-Analyse von Chernikova et al. (2019) stellt jedoch fest, dass Kompetenzentwicklung mit Fallanalyse-Problemen einen größeren Effekt auf den Lernzuwachs von angehenden Lehrerinnen und Lehrern wie auch Medizinerinnen und Medizinerinnen hatten, als ein nicht-problembasiertes Lernangebot. Insbesondere bei Lernenden mit geringem Vorwissen. Positive Effekte traten besonders dann auf, wenn Lernende zu interaktionsbasierten Lernaktivitäten und problembezogener Rollenübernahme angeregt wurden. Wie Hmelo-Silver, Duncan und Chinn (2007) anmerken, können fall- und problembasierte Lernformate durchaus für frühen Phasen der Ausbildung geeignet sein, solange sie didaktisch gerahmt werden.

Mit Blick auf fallbasiertes Lernen in der Lehrkräftebildung definiert L. Shulman (1992) einen Fall als konkretes Beispiel eines größeren theoretischen Konzepts, einer breiteren Kategorie oder einer Klasse von Ereignisabläufen. Aus problemtheoretischer Perspektive besteht an die Konstruktion diagnostischer Fälle daher die strukturelle Anforderung, mit dem Lernmaterial typische Merkmale von o. g. ‚ill-structured‘ Problemen abzubilden. Mit den Fällen muss die Konstruktion mentaler Modelle möglich sein, in denen Operatoren zur Manipulation des Problemraums angewendet werden können. In dieser Forschungsarbeit zeigt eine Studie (Kapitel 10), dass bereits die Rezeption konstruierter Fälle die Bildung mentaler Modelle anregt. Sie zeigt darüber hinaus, dass kognitiven Basisfähigkeiten einen Effekt auf die Konstruktion mentaler Modelle und die diagnostische Urteilsgenauigkeit hatten. Dies ist ein starker empirischer Beleg für die in Kapitel 2.4 formulierte allgemeine Kompetenzdefinition, die eine funktionelle Verbindung kontextspezifischer kognitiver Module und Strukturen mit allgemeinen kognitiven Grundfähigkeiten als Teil von Kompetenz betrachtet. Mit dem Blick auf diagnostische Fälle als eine kontextspezifische Form von ‚ill-structured‘ Problemen steht ein Lerngegenstand zur Verfügung, der bei einer passenden didaktischen Rahmung voraussichtlich die Entwicklung diagnostischer Kompetenz unterstützen kann.

4.2 Problemorientiertes Lernen als Methode der Fallarbeit

Vor dem Hintergrund der entwickelten Definition diagnostischer Kompetenz ist die kooperative Zusammenarbeit ein wesentlicher Bestandteil diagnostischen Handelns. Eine dementsprechende didaktische Rahmung des Lernens mit Problemfällen sind Kooperationsskripte (Fischer et al., 2013; Vogel, Wecker, Kollar, & Fischer, 2017). Damit sind soziale Interaktions- und Ablaufschemata gemeint, die typisch für einen Kontext sind. Diese können auch als Teil der Operatoren zur Manipulation des Problemraumes konzipiert werden, so dass eine Entwicklung dieser Fähigkeiten ebenfalls zur Effizienz von Problemlösungen beitragen sollte.

Eine didaktische Methode, deren Kernelement kooperative Fallarbeit ist, ist problemorientiertes Lernen (POL) (Barrows, 1986). Diese im medizinischen Ausbildungskontext entwickelte Methode hat einen starken Fokus auf situiertes Lernen in Kleingruppen, die ein realitätsnahes Problem bearbeiten und sich dabei selbstgesteuert Wissen und Fähigkeiten mit Bezug zum Kontext der Problemstellung aneignen (Hmelo-Silver, 2004). Eine besondere Rolle spielen Lernziele, auf die sich die Kleingruppe gemeinsam festlegt, dann jedoch individuell bearbeitet und im Anschluss wieder gemeinsam diskutiert. POL zielt damit über den Aufbau kontextspezifischen Wissens hinaus auf die Entwicklung epistemischer und sozialer Fähigkeiten zur Bewältigung diagnostischer Anforderungssituationen. Voraussetzung dafür ist das stringente Durchlaufen spezifischer Schritte in der Fallarbeit (siehe Kapitel 4.2), deren Wiederholung an jedem Fall eine konsistente Struktur zur Verfügung stellt. Vor dem Hintergrund einer angestrebten beständigen professionellen Weiterentwicklung kann die Entwicklung kontextspezifischer und kooperativer Lernfähigkeiten den Transfer der Kompetenzentwicklung in verschiedene institutionelle Praxisorte hinein unterstützen (Barrows, 2000).

Die Kooperation von Lehrerinnen und Lehrern hat ein großes Potenzial für die Gestaltung von Schulen als Lernumwelt. Bekannt sind etwa positive Effekte auf die Leistungen von Schülerinnen und Schülern und auf die Herstellung eines lernförderlichen Klassenklimas (Huber & Ahlgrimm, 2012;

Toom, Pietarinen, Soini, & Pyhältö, 2017; Vangrieken, Meredith, Parker, & Kyndt, 2017; Warwas & Helm, 2018). In der Forschung wurden bisher fast ausschließlich in der Arbeitswelt tätige Lehrerinnen und Lehrer hinsichtlich ihres kooperativen Handelns befragt und Ergebnisse vor dem Hintergrund situierten Lernens interpretiert (Gräsel, Fussangel & Parchmann, 2006; Gräsel, Fussangel & Pröbstel, 2006; Richter & Pant, 2016; OECD, 2018). Untersuchungen zum Effekt kooperativer Lernformen auf die Kompetenzentwicklung angehender Lehrkräfte sind dagegen seltener (Hemker, Prescher, & Narciss, 2017).

Die allgemeine Befundlage für Lernergebnisse mit POL ist dagegen sehr umfangreich und vielfach rezipiert, so dass an dieser Stelle eine sehr geraffte Darstellung erfolgt. Bekannt sind positive Effekte auf den Aufbau epistemischer Überzeugungen zur aktiven Wissenskonstruktion (Kuhn, 2007) und auf den langfristigen Erhalt erworbener Wissensbestände (Wirkala & Kuhn, 2011). Wenngleich Meta-Analysen und Meta-Synthesen zum Wissenserwerb uneindeutige Ergebnisse zum Erwerb deklarativen Wissens zeigen, weisen sie jedoch eindeutig auf positive Effekte auf den Erwerb prozeduralen und anwendungsorientierten Wissens sowie auf Argumentationsfähigkeiten hin (Dochy, Segers, van der Bossche, & Gijbels, 2003; Strobel & van Barneveld, 2009; Bergstrom, Pugh, Phillips, & Machlev, 2016) nach. Dazu kommt eine durchschnittlich höhere Lernmotivation als in lehrkräftezentrierte Lerngelegenheiten, u. a. durch eine höhere Nutzenerwartung (Sungur & Tekkaya, 2006; Hwang & Kim, 2006; Sangestani & Khatiban, 2013). Spezifisch für Lehramtsstudierende zeigen empirische Studien ähnliche Effekte (Chernikova et al., 2019; Hemker et al., 2017; Wilhelm & Brovelli, 2009).

Auch für die Erstellung von Lernmaterial, d. h. zur Fallkonstruktion, existiert eine Fülle an theoretischer und empirischer Literatur sowie einige Ratgeber. Zentral sind die Problemtypologie von Jonassen (2000) und Jonassen und Hung (2008) sowie das 3C3R-Modell (3 Core Components: Context, Content, Connection & 3 Processing Components: Researching, Reasoning, Reflecting) zur Fallkonstruktion von Hung (2006). Für die Lernumwelt bzw. Intervention, die drei der vorgestellten Studien diskutieren, wurden diese Modelle mit dem in der Lehrkräftebildung bestehenden Ansatz von Klug et al. (2016) zu einer Konstruktionsmatrix für diagnostische Fälle in der Lehrkräftebildung integriert. Eine genauere Diskussion dieser Matrix findet in der Studie in Kapitel 8 statt. In der im Rahmen dieser Forschungsarbeit entwickelten Intervention liegt der Fokus der Fälle auf individuellen Lernvoraussetzungen und -störungen mit hohen Prävalenzraten, die in handlungsauffordernden Unterrichtssituationen und diagnostischen Zusatzmaterialien ‚verpackt‘ sind.

4.3 Die Intervention POL-TF

Gagné (1965) stellt in seinem lernpsychologischen Grundlagenwerk fest, dass unterschiedliche Arten von Lernergebnissen auch unterschiedliche Lernbedingungen voraussetzen. Diesem lerntheoretischen Grundsatz folgend, ist es hochgradig unwahrscheinlich, dass Personen diagnostische Kompetenz lediglich in einer einzigen Form von Lernumwelt entwickeln können. Im Rahmen der vorliegenden Forschungsarbeit wurde eine Lernumwelt entwickelt, die ihren Fokus auf den problembezogenen Erwerb bestimmter Dispositionen legt. Dazu gehören diagnostische Wissensbestände und das diagnostische Selbstkonzept, das aus selbstbezogenen Überzeugungen zum

Diagnostizieren besteht. Damit bildet die Entwicklung zweier zentraler Bestandteile zur Konstruktion eines fallbezogenen Problemraumes den Schwerpunkt der Intervention.

POL erfreut sich bereits einer breiten Anwendung in der Medizindidaktik, in der noch stärker als in der Lehrkräftebildung die Arbeit an Einzelfällen im Vordergrund steht. Fallarbeit soll die Komplexität professioneller Handlungssituationen ohne den Erfolgsdruck der Berufspraxis einer Erfahrungsverarbeitung zugänglich machen (Reh & Schnelle, 2010). Dazu wird ein eingegrenzter Ausschnitt aus der beruflichen Praxis ‚aufgenommen‘, der dann als Lerngegenstand exploriert werden kann. Der betreffende Ausschnitt wird mit dem Ziel ausgewählt, zentrale Herausforderungen professionellen Handelns verdichtet abzubilden und zur multidirektionalen kognitiven Auseinandersetzung damit anzuregen. Idealerweise liegt dem Fall daher ein unabgeschlossenes Narrativ zu Grunde, welches die Möglichkeit zur Übernahme der Perspektive einer Entscheidungsträgerin bzw. eines Entscheidungsträgers bietet (Barrows, 1986). Die Gruppensituationen verlangt den einzelnen Personen häufig ab, individuelle Denk- und Entscheidungsvorgänge zu explizieren (Bergstrom et al., 2016). Typische Gesprächsinhalte sind Wahrnehmungen von Informationen im Fall, deren Interpretation und Erfahrungen zu ähnlichen Handlungssituationen. Im gemeinsamen Austausch können die Studierenden daher potenziell auch alternative kognitive Umgänge mit dem Fall kennenlernen. Da die Handlungssituation im Fall nicht abgeschlossen ist, sind mehrere Weiterführungen des dargestellten Praxisausschnitts möglich. Diese Eigenschaft der Fälle soll das Abwägen verschiedener Handlungsmöglichkeiten und die Konstruktion des diagnostischen Problemraums durch Identifikation und Manipulation des Initialzustands sowie möglicher Zielzustände unterstützen. Für diese kognitive Operation braucht es Annahmen über Gesetzmäßigkeiten und wahrscheinliche Handlungsverläufe, wie auch Inferenzbildung (Elstein, Shulman, & Sprafka, 1978). Auch Fischer et al. (2014) konzipieren epistemische Tätigkeiten als notwendige kognitive Aktivitäten erfolgreichen Diagnostizierens.

Die sinnvolle Anwendung und Übung dieser Fähigkeiten setzt jedoch nicht nur fallbezogenes Wissen über die konkrete Handlungssituation voraus, sondern auch grundlegendes Wissen über relevante Einflussgrößen in der Situation. Für Fälle pädagogischer Diagnostik besteht dieses kontextspezifische Grundlagenwissen insbesondere in Kenntnissen über Lernvoraussetzungen und -störungen. Durch die Aneignung dieses Wissens über das selbstständige Festlegen von Lernzielen können Studierende den Fall vertieft verstehen und sinnvolle Handlungsmöglichkeiten für die entscheidungstragende Person im Fall entwerfen. Dieses Ineinandergreifen von verschiedenen Ebenen und Wissensdomänen, sowie die Verbindung von theoretischem Wissen und dessen Anwendung in praxisnahen Situationen kann eine anregende und reichhaltige Lernumwelt bieten. Eine wichtige Voraussetzung dafür sind didaktische Reduktionen im Design dieser Lernumwelt (Jonassen, 1997; Hmelo-Silver, 2004).

Im typischen Verlauf ist eine Bearbeitung jeweils eines Falls in zwei Lerneinheiten und sieben Schritten vorgesehen, wobei zwischen den Lerneinheiten ein Zeitraum zur selbstgesteuerten Auseinandersetzung mit Lernzielen vorgesehen ist (siehe *Abbildung 7*). Dieser Verlauf wiederholt sich mit den nächsten Fällen.

Zur Implementation von POL-TF (Problemorientiertes Lernen mit diagnostischen Textfällen) auf Kursebene wurde das Four Component Instructional Design (4C-ID) Modell angepasst (van

Merrienboër, Clark & de Croock, 2002). Das Modell wurde in den 1990er Jahren von van Merrienboër entwickelt. Es systematisiert Designelemente für das Lernen in komplexen Lernumwelten und stellt eine Blaupause didaktischer Strukturelemente zur Verfügung, mit denen Lernvorgänge methodisch geleitet aufbereiten werden können. Nach dem 4C-ID Model sind vier Bestandteile essenziell für die Lernbegleitung.

--Beginn der ersten Lerneinheit--

1. Fall lesen und Verständnisfragen klären
2. Informationen im Fall zusammentragen und Problem definieren
3. Relevante Themen auswählen
4. Vorwissen aktivieren und visualisieren
5. Gemeinsame Lernziele festlegen

--Ende der ersten Lerneinheit--

Alle Gruppenmitglieder bearbeiten nun jeweils individuell dieselben Lernziele in Eigenarbeit bis zur nächsten Lerneinheit, die bspw. eine Woche später anschließt.

--Beginn der zweiten Lerneinheit--

6. Zusammengetragene Informationen diskutieren, vergleichen und visualisieren
7. Reflexion des Lernprozesses, ggf. Hypothesenbildung und Planung weiteren Vorgehens im Fall

--Ende der zweiten Lerneinheit--

Abbildung 7. Schritte der kooperativen Fallarbeit in POL-TF.

Erstens ist dies der Lerngegenstand. Vorgesehen sind Lerngegenstände, die zur Konstruktion kognitiver Schemata für Elemente im Lernprozess anregen (z. B. ‚illness-scripts‘) anregen. So fordern inhaltlich verschiedene diagnostische Fälle zwar die Auswahl oder Auswertung spezifischer Diagnoseinstrumente ein. Für den Auswahl- oder Auswertungsprozess selbst kann jedoch ein Schema gebildet werden, in dem verschiedene Kriterien gegeneinander abgewogen werden. Da dieses Schema immer wieder in der Fallarbeit angewendet wird, entsteht die Möglichkeit zur Bildung einer kognitiven Routine für die Schemaanwendung.

Zweitens ist dies die Inhaltliche Unterstützung. Ziel ist dabei die Verbindung von Vorwissen mit neuen Wissensbausteinen, die eine tiefgehende Bearbeitung der Fälle ermöglichen. In POL-TF sind das Grundlagen diagnostischen Wissens, mit denen das Verstehen diagnostischer Informationen vereinfacht wird. Dazu gehört Wissen über den typischen Verlauf einer Diagnostik, über Qualitätskriterien, verschiedene Diagnosemethoden sowie Lernvoraussetzungen und -störungen.

Drittens ist dies die methodische Unterstützung. Die jeweiligen Bearbeitungsschritte in den Lernvorgängen (z. B. Lernziele formulieren) können durch eine anfängliche Begleitung erleichtert werden. Da sie verhältnismäßig schnell gelernt und in abrufbare Routinen integriert werden können, erfolgt die Bereitstellung dieser Unterstützung erst an dem Punkt, an dem sie relevant wird. Für POL-TF sind das primär einfache Moderationsaufgaben (z. B. Zeitmanagement, Lernziel- und Hypothesenformulierung).

Viertens sind dies Teilübungen. Einzelne Fähigkeiten aus der Fallarbeit werden aus der komplexen Fallsituation isoliert und den Lernenden als wiederholbare Übung bereitgestellt, um die Bildung schnell abrufbarer Routinen zu unterstützen. In POL-TF sind das insbesondere die Hypothesenbildung und die Formulierung von Feedbacks für Schülerinnen und Schüler. Das Design der entwickelten Intervention ist in Abbildung 8 dargestellt.

Zusammengefasst zeigt sich nicht nur durch die konsistenten Prävalenzraten von Lernschwierigkeiten und -störungen im schulischen Kontext ein Bedarf an der theoriegeleiteten Entwicklung von Interventionen zur Steigerung diagnostischer Kompetenz angehender Lehrkräfte. Bisherige fallbasierte Interventionen weisen zwei Merkmale auf, die diesen Bedarf noch akzentuieren. Keine der o. g. Lernumwelten definiert theoretisch, was ein Fall ist. Zudem bleibt der kooperative Aspekt von Diagnostik der Kompetenzentwicklung außen vor. Damit sind zwei zentrale Bestandteile entsprechender Interventionen unterbestimmt. Die problemtheoretische Perspektive auf Diagnostik eröffnet dagegen einen integrativen Umgang mit dem diagnostischen Prozess und seiner fallbasierten Vermittlung. Die darauf basierende didaktische Methode des problemorientierten Lernens betont den kooperativen Aspekt diagnostischen Handelns, ohne dabei die individuelle Auseinandersetzung mit diagnostischen Fällen außer Acht zu lassen. Da das 4C-ID Modell auch auf die Entwicklung kognitiver Routinen zielt, knüpft die entwickelte Intervention direkt an die Erkenntnisse der Expertiseforschung zur Funktionalität von ‚illness-scripts‘ in diagnostischen Anforderungssituationen an. Vor diesem Hintergrund wurde eine Intervention zur Entwicklung diagnostischer Kompetenz ausgearbeitet, deren empirische Überprüfung ein zentraler Aspekt dieser Forschungsarbeit ist.

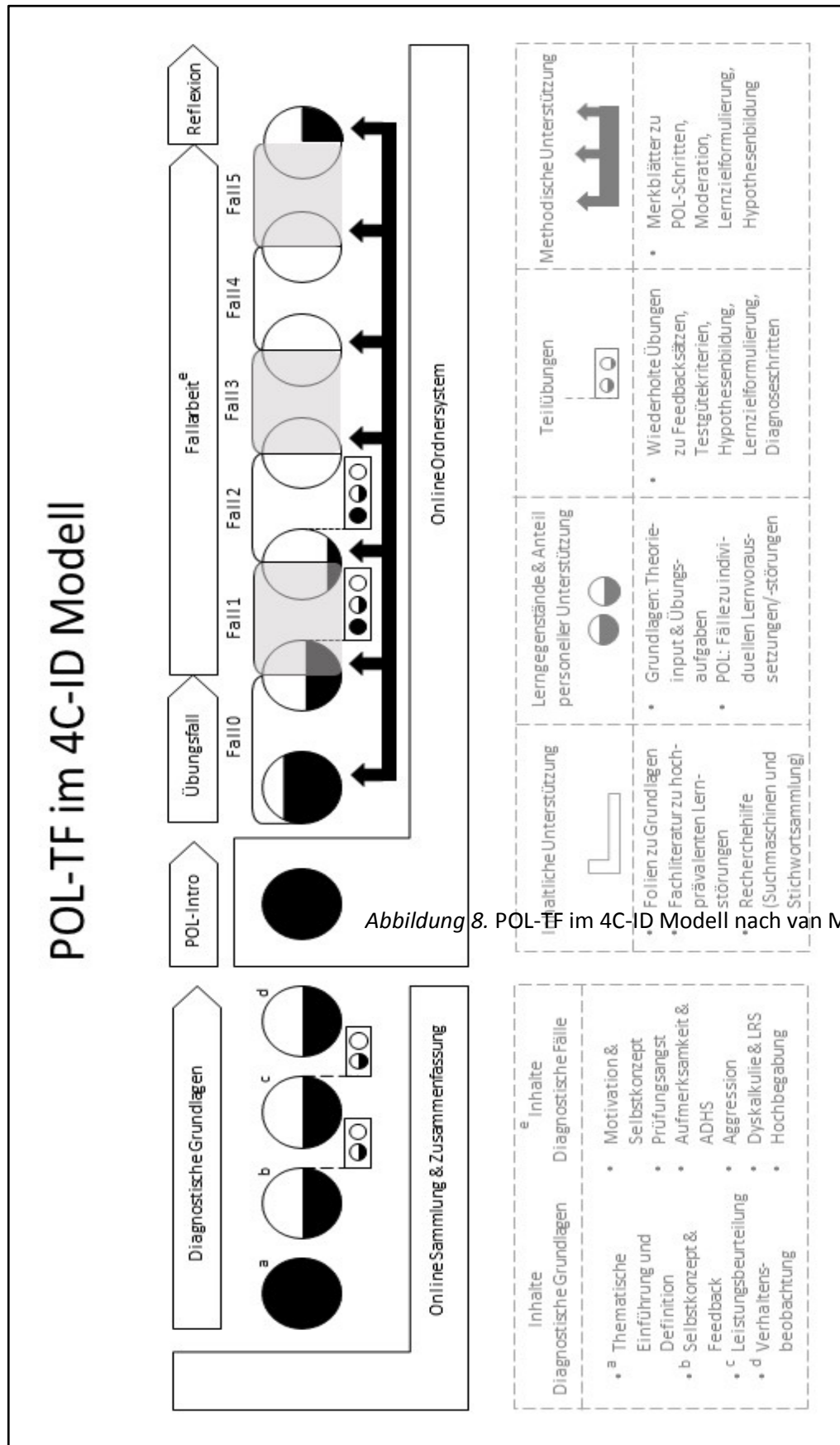


Abbildung 8. POL-TF im 4C-ID Modell nach van Merriënboër et al. (2002)

Methodischer Teil

5 Methoden: Veränderungsmessung in kleinen Stichproben

Interventionsforschung untersucht die Effekte einer gezielt hergestellten Umwelt auf die Entwicklung von Personen, Organisationen oder Institutionen. Da es zur direkten Messung intra-individueller psychologischer Veränderungen kaum geeignete technologischen Messinstrumente und -methoden gibt, wird die Wirksamkeit einer Intervention momentan hauptsächlich durch indirekte Messverfahren beurteilt. Die Psychometrie greift dazu auf die klassische und probabilistische Testtheorie zurück, die mit Hilfe statistischer Modellierung von numerischen Messwerten auf psychologische Merkmale schließt. Um über die Ausprägung gemessener Merkmale hinaus auch das Ausmaß ihrer Veränderung und den Effekt einer Intervention auf ihre Veränderung schätzen zu können, existieren eine Reihe statistischer Modelle (im Überblick: Dietrich, 2019).

Die Auswahl eines geeigneten statistischen Modells ist sowohl vom Studiendesign als auch von der Spezifität der aufgestellten Hypothesen abhängig. In dieser Entscheidung wird prinzipiell das Falsifikationsprinzip angewendet. Da sich ein einfacheres Modell auf eine größere Zahl an Fällen bzw. Stichproben anwenden lässt, ist seine Falsifizierbarkeit größer als die eines komplexeren Modells. Bei gleicher Aussage ist sodann das einfachere Modell zu bevorzugen. Der gewählte Komplexitätsgrad ist immer ein Trade-Off zwischen einerseits der Reduktion des Messfehlers durch eine möglichst umfangreiche Spezifikation des Modells und andererseits der dadurch entstehenden Bedingtheit der Interpretation in Bezug auf das Modell. Dieser Trade-Off betrifft zu aller erst die Entscheidung zwischen einer latenten oder manifesten Modellierung.

Latente Modelle zur Messung quantitativer Veränderungen haben die attraktive Eigenschaft der messfehlerbereinigten Parameterschätzung, was jedoch auch mit einigen methodologischen Voraussetzungen, Konsequenzen und Limitationen einhergeht. Für latente Veränderungsmessungen auf intra-individueller Ebene können neben einfachen Pfadmodellen auch Latent-Change Modelle, Wachstumskurvenmodelle oder autoregressive Zeitreihenanalysen geschätzt werden (Eid, Gollwitzer, & Schmitt, 2011). Zur expliziten Schätzung und ‚Herauspartialisierung‘ von Messfehlern wird in Strukturgleichungsmodellen eine probabilistische Beziehung zwischen beobachtbaren, d. h. manifesten Indikatoren (z. B. Ankreuzverhalten) und nicht beobachtbaren, d. h. latenten Merkmalen (z. B. kognitive Struktur) angenommen. Diese bilden das Messmodell, das in Regressionsgleichungen formuliert wird. Dabei geben die Regressionskoeffizienten an, mit welchem Gewicht die jeweiligen Indikatorwerte in die Schätzung des latenten Merkmals eingehen und welcher Anteil nicht zur Varianzerklärung beiträgt (Messfehler). Die Beziehung zwischen latenten Merkmalen bildet das Strukturmodell ab, für das ebenfalls Regressionsgleichungen formuliert werden. Diese Modellspezifikation hat zur Folge, dass die Gleichungen des Strukturmodells nur noch ‚bereinigte‘ Werte erklären. Die Formulierung der Gleichungen erfolgt mit dem Ziel, die Passung zwischen theoretisch angenommenen und empirisch gemessenen Beziehungen zu beurteilen. Eine Reduktion von Messfehlern ist dabei wünschenswert, um das Risiko von Fehlern zweiter Art zu minimieren. D. h. eine real vorhandene Beziehung zwischen latenten Merkmalen soll nach Möglichkeit auch zu identifizieren sein.

In der statistischen Messfehlerbereinigung liegt eine erste methodologische Konsequenz für ableitbare Interpretationen. Da die Parameterschätzung direkt von latenten Variablen abhängt, können Modellparameter ausschließlich unter Annahme der latenten Struktur, d. h. einer bedingten Verteilung der Merkmale interpretiert werden. Eine direkte Interpretation von Randverteilungen, d. h. der Ausprägung einzelner latenter Variablen (z. B. Intercept und Standardfehler), ist dadurch im Gegensatz zu manifesten Veränderungsmessungen nicht mehr ohne weiteres möglich. Wird bspw. in einem latenten Wachstumskurvenmodell zu einer Messung von einer abhängigen Variable und zwei unabhängigen Variablen zu mehreren Messzeitpunkten das Ausmaß der Veränderung der abhängigen Variable geschätzt, bildet die Ausprägung der abhängigen Variable nicht mehr den Mittelwert in der Population ab. Sie bildet diesen unter der Bedingung des angenommenen Strukturmodells ab. D. h. dessen Schätzung ist nur für die Personen gültig, für die auch das Strukturmodell gültig ist. Damit sind Vergleiche von Gruppen, in denen auf Grund der Einführung zusätzlicher (hypothetischer) Wirkbedingungen kein äquivalentes Strukturmodell angenommen werden kann (z. B. Experimental- und Kontrollgruppe) nicht ohne Vorliegen strikter Messinvarianz zulässig. Chen (2008) zeigt, dass mit dem Anteil nicht-messinvarianter Items zwischen Gruppen der Bias der geschätzten Mittelwertsunterschiede proportional steigt (siehe auch Guenole & Brown, 2014). Auch Steinmetz (2013) sowie Schmitt, Golubovich und Leong (2011) stellen fest, dass bereits ungleiche Intercepts zwischen Gruppen zu fehlerbehafteten Schätzungen von Mittelwertsdifferenzen führen. Systematische Reviews zur Verwendung von Messinvarianztestung berichten, dass in psychologischer Forschung verwendete Strukturgleichungsmodelle nur in allerseltensten Fällen strikte Messinvarianz aufweisen (Putnick & Bornstein, 2016; Schmitt & Kuljanin, 2008).

Eine zweite methodologische Besonderheit latenter Modellierung entsteht im Rahmen der Spezifikation, d. h. bei der Festlegung von Zusammenhangsregeln für manifeste Indikatoren und latente Variablen. Die geschätzten Parameter sind nur dann messfehlerfrei, wenn eine korrekte Spezifikation des Mess- und Strukturmodells vorliegt. Die Spezifikation erfolgt u. a. durch die Definition der Linkfunktionen (linear, logistisch, quadratisch etc.), realweltlich vorhandener Cluster (Orte, Messzeitpunkte, soziale Zusammenhänge etc.) und Auswahl oder Auslassung der einzuschließenden Variablen. Eine Fehlspezifikation des Modells (z. B. Ignorieren einer Clustervariable oder eines bestimmten Pfads) kann bei Verwendung von Algorithmen zur Schätzung der Gleichungen im gesamten Modell, d. h. die zur Schätzung auf die gesamte Kovarianzmatrix zugreifen (z. B. Maximum Likelihood), dazu führen, dass Parameter nicht mehr messfehlerfrei sind (Tomarken & Waller, 2005). In Forschungsbereichen ohne umfangreiche empirische Grundlage steigt das Risiko einer Fehlspezifikation und damit der Einführung von Messfehlern durch die latente Modellierung.

Eine dritte methodologische Konsequenz betrifft die notwendige Interpretation globaler Fit Indices (z. B. Standardized Root Mean Square Residual [SRMR], Root Mean Square Error of Approximation [RMSEA], Goodness-of-Fit Statistic [GFI]). Prinzipiell beschreibt die Passung bzw. der Fit eines aufgestellten Modells, dass es nicht falsifiziert wurde. Dies bedeutet jedoch nicht, dass es die empirischen Daten ‚wahrheitsgemäß‘ abbildet (Brannick, 1995). Da Strukturmodelle aus mehreren Regressionsgleichungen bestehen, ist nicht nur eine lokale Passung jeder einzelnen Gleichung auf den betreffenden Ausschnitt der Datenmatrix relevant, sondern auch die globale Passung des gesamten Gleichungssystems. Tomarken und Waller (2003) stellen jedoch fest, dass zwei verschiedener

Strukturgleichungsmodelle einen exakt gleichen globalen Fit haben können. Dies kann trotz unterschiedlicher lokaler Passungen der einzelnen Gleichungen zustande kommen, indem sich zwar die aufgestellten Gleichungen voneinander unterscheiden, nicht aber die durch sie bedingte Restriktion der Kovarianzmatrix gemessener Variablen (z. B. resultieren die Struktur-, bzw. Pfadmodelle $X \rightarrow Y \rightarrow Z$ und $Z \rightarrow Y \rightarrow X$ in einer identischen implizierten Kovarianzmatrix). Eine bessere Modellpassung auf globaler Ebene bedeutet daher nicht zwingend, dass ein Modell einen empirischen Sachverhalt ‚besser‘ abbildet oder ‚plausibler‘ ist.

Darüber hinaus ist die qualitative Interpretation globaler Fit Indices, d. h. ab wann eine ausreichende Modellpassung vorliegt, eine komplexe Frage. Hu und Bentler (1999) berichten auf Basis von Simulationsstudien einige Cut-Off Werte, die eine solche Aussage treffen sollen. Eine Interpretation entlang von Daumenregeln birgt jedoch auch Risiken. Der Großteil absoluter Fit Indices basiert etwa auf dem χ^2 Index und ist daher abhängig von der Stichprobengröße (Bearden, Sharma, & Teel, 1982). So berechnet sich der AIC (Akaike Information Criterion: Tanaka, 1993) etwa durch $\chi^2 + 2df$. Bei einer sehr großen Stichprobe besteht daher das Risiko, dass konkurrierende Modelle sich nicht mehr ausreichend in ihrem Modellfit unterscheiden. Weiterhin ist in Mehrgruppenmodellen mit unterschiedlichen Stichprobengrößen somit nicht eindeutig feststellbar, ob ein niedriger Fit-Wert auf das spezifizierte Modell zurückzuführen ist. Eine Alternative stellen relative Fit Indices dar (z. B. Tucker-Lewis Index [TLI]), die weniger von der Stichprobengröße abhängen. Relative Fit Indices geben Abweichungen von einem Nullmodell an, d. h. einem Strukturmodell ohne Korrelation zwischen den latenten Variablen. Da ein Nullmodell jedoch insbesondere für intra-individuelle Strukturen aus psychologischen Konstrukten empirisch nur in den seltensten Fällen wahrscheinlich ist, haben relative Fit Indices in diesem Kontext nur eine geringe Aussagekraft über die Adäquanz des aufgestellten Modells. Auch ein Vergleich verschiedener Modelle über relative Fit Indices ist an diese Einschränkung gebunden. Generell zeigen globale Fit Indices komplexe Abhängigkeitsstrukturen verschiedener, sich gegenseitig beeinflussender Parameter, wie z. B. Schätzalgorithmus, Stichprobengröße, Verteilung und Modellspezifikation (Fan, Thompson, & Wang; 1999). Hu und Bentler (1998) schlussfolgerten bereits: „it is difficult to designate a specific cutoff value for each fit index because it does not work equally well with various types of fit indices, sample sizes, estimators, or distributions“ (S. 449).

In der Praxis der Strukturgleichungsmodellierung sind Fit Indices ein gängiger Anlass zur Modifikation eines Modells. Hermida (2015) kommt etwa in einem systematischen Review zu dem Ergebnis, dass 37 % der 315 analysierten Studien Modellfit Indices zur Veränderung des getesteten Modells hinsichtlich Residualkorrelationen nutzen und nur 7 % bei Testung einer A-priori Spezifikation bleiben. Obwohl Strukturgleichungsmodellierung ein hypothesenprüfendes Verfahren ist, findet in der Forschungspraxis aus ökonomischen Gründen häufiger ein exploratives Vorgehen statt (Landis, Edwards, & Cortina, 2009). Die Spezifikation von Messfehlern stellt eine besondere Herausforderung dar, deren Bewältigung insbesondere bei komplexen Datenlagen und Modellen offenbar häufig in dieser Art „post hockery“ oder „fitishism“ (Nachtigall, Kroehne, Funke, & Steyer, 2003) resultiert.

Die wohl bekannteste Limitation latenter Modellierung ist ihre Anwendung in kleinen Stichproben. Da die Anzahl freier Parameter in latenten Modellen generell höher als die in manifesten Modellen ist, wird eine größere Informationsmenge benötigt, damit Modelle konvergieren, d. h. damit eine geschätzte Kovarianzmatrix gefunden wird, die nicht zu stark von den empirischen Daten abweicht.

Trotz der Abhängigkeit der Stichprobengröße von einzelnen Modellparametern (Komplexität, Schätzalgorithmus, Verteilung gemessener Werte, Linkfunktionen etc.), wird häufig ein Mindestumfang von 200 Personen angegeben (Barrett, 2007; Weston & Gore, 2006). Bei kleineren Stichproben verschärft sich die Bedeutung der aufgeführten Besonderheiten latenter Modellierung, die ihrerseits auch nur einen Ausschnitt aus weiteren Merkmalen von Strukturgleichungsmodellen darstellen (McCoach, Black & O'Connell, 2007). Strukturgleichungsmodelle haben jeweils sehr stringente statistische Voraussetzungen, die sich nicht nur je nach Modell, sondern auch nach angewandtem Algorithmus zur Schätzung von Parametern der aufgestellten Gleichungen unterscheiden (Bagozzi & Yi, 2012). Während bspw. die häufig eingesetzte Maximum Likelihood (ML) Methode eine multivariate Normalverteilung voraussetzt, ist die Weighted Least Squares Methode (WLS) robust gegenüber einer Verletzung dieser Annahme, setzt aber eine weit größere Stichprobe voraus, wenn kein Bias in der Parameterschätzung entstehen soll (Yung & Bentler, 1994). Weiterhin stellen Chen, Curran, Bollen, Kirby und Paxton (2008) fest, dass der häufig verwendete RMSEA Fit Index korrekte Modelle in Stichproben mit $n < 100$ größtenteils fehlerhaft ablehnt. Inwiefern die Komplexität latenter Modellierung und der damit einhergehenden Komplexität der Ergebnisinterpretation und -kommunikation zielführend ist, bleibt natürlich eine Entscheidungsfrage im konkreten Einzelfall. Die aufgeführten Besonderheiten latenter Modellierung verdeutlichen jedoch, dass jeder methodische Vorteil durch eine methodologische Einschränkung erworben wird.

Die Interventionsstudien in der vorliegenden Forschungsarbeit wurden mit verhältnismäßig kleinen und nicht randomisierten Stichproben durchgeführt. Dies entspricht der allgemeinen Lage in der Forschung zur Lehrkräftebildung, in der bereits verschiedentlich beschrieben wurde, dass in diesem Kontext kleine und nicht randomisierte Stichproben der Normalfall sind. Eine genauere Diskussion dieser kontextuellen Charakteristika des Forschungsfeldes findet in Kapitel 9.1 statt. Auch die Modellierung kleiner Stichproben weist einige methodische Besonderheiten auf (im Überblick: Savalei, 2010). Wenn sie grundlegende Modellannahmen klassischer und probabilistischer Testtheorie nicht erfüllen, ist bereits die Anwendung klassischer statistischer Modelle (z. B. einfache Regressions- und Varianzanalyse) eingeschränkt. Daher wurden zahlreiche robuste Schätzverfahren entwickelt, die bspw. transformierte Verteilungen oder alternative Umgänge mit Maßen zentraler Tendenz ausnutzen (Field & Wilcox, 2017). Wenn jedoch komplexere Fragestellungen an die Daten formuliert werden, bspw. die Schätzung ‚wahrer‘ Werte oder die Veränderung von Zusammenhängen über die Zeit, sind Möglichkeiten der Modellierung mit robusten klassischen Verfahren ebenfalls stark begrenzt.

Zusammengefasst führt eine latente Modellierung intra-individueller Veränderungen statistische und methodologische Annahmen ein, die eine Interpretation der Ergebnisse gegenüber manifesten Modellen deutlich erschweren. Darüber hinaus reicht die Messfehlerbereinigung in einfachen Strukturgleichungsmodellen auch nur bis an den Forschungskontext heran, d. h. Messfehler durch systematische Designparameter und Kontextmerkmale der Studien sind davon ausgenommen. Gerade in Interventionsstudien ist dies jedoch ein gewichtiger Faktor für die Beurteilung des Transfers vom Entwicklungskontext in Anwendungskontexte. Die latente Modellierung über Mehrgruppenmodelle kann dieses Problem zwar lösen, führt jedoch gleichzeitig andere methodische Anforderungen ein, wie etwa strikte Messinvarianz und eine erheblich größere Anzahl frei zu schätzender Parameter. Für

Interventionsforschung in Feldern, in denen regelmäßig nur kleine Stichproben gewonnen werden können, sind diese Anforderungen kaum oder gar nicht zu bewältigen.

Um der Pattsituation zwischen unzureichenden Modellierungsmöglichkeiten in ‚klassischen‘ manifesten Verfahren einerseits und zu hohen Anforderungen und interpretativen Einschränkungen in latenten Verfahren andererseits zu begegnen, greift die vorliegende Forschungsarbeit in zwei Studien (Kapitel 8 und 9) auf Auswertungsmethoden zurück, die nicht Bestandteil des klassischen Methodenkanons sind, und von denen eine spezifisch für diese Forschungsarbeit mathematisch korrigiert und statistisch weiterentwickelt wurde. Im Folgenden werden zentrale Grundlagen dieser Methoden vorgestellt.

5.1 Replikation und manifeste meta-analytische Regressionsmodelle

Eine Möglichkeit zur manifesten messfehlerbereinigten Schätzung von Interventionseffekten bietet der meta-analytische Ansatz. Dieser wendet das Reliabilitätsprinzip, d. h. die explizite Schätzung des Messfehlers durch mehrere aber leicht verschiedene Indikatoren auf die Ebene der Studien an. Die zentrale Annahme des Verfahrens ist, dass Messfehler aus (1) der Interaktion von Personen und Items und (2) systematischen Kontexteffekten entstehen. Letztere können nur durch die mehrfache Durchführung einer Studie in verschiedenen Kontexten und mit verschiedenen Teilnehmenden kontrolliert und damit explizit geschätzt werden. Hintergrund dessen ist wiederum das Falsifikationsprinzip. Wissenschaftliche Aussagen über einen Realitätsausschnitt sollten nicht nur möglichst wenig von einem konkreten Messinstrument abhängen, sondern auch einer wiederholten Überprüfung standhalten. Popper (1934) formulierte dazu: „unsere eigenen Beobachtungen pflegen wir wissenschaftlich nicht ernst zu nehmen, bevor wir sie nicht selbst durch wiederholte Beobachtungen oder Versuche nachgeprüft und uns überzeugt haben, dass es sich nicht nur um ein zufälliges Eintreffen handelt“ (S. 17).

Besonders in der psychologischen Interventionsforschung ist die Replikation von Befunden ein Goldstandard (Nosek et al., 2015) und konnte bereits zur Revision oder Spezifizierung sehr bekannter psychologischer Effekte führen. Beispielsweise zeigten Carney, Cuddy und Yapp (2010) in einer Interventionsstudie einen direkten positiven Effekt von Power Posing auf den Testosteronspiegel, das Machtgefühl und verhaltensbezogene Risikotoleranz sowie einen negativen Effekt auf den Cortisolspiegel. In einer Meta-Analyse von sieben prä-registrierten Studien zur direkten Replikation der Ergebnisse, d. h. einer Eins-zu-Eins Anwendung der Intervention mit anderen Teilnehmenden, stellten Gronau, van Erp, Heck, Cesario, Jonas und Wagenmakers (2017) einen schwachen Effekt auf das Machtgefühl aber keinen anderen Effekt fest. Ein zweites Beispiel betrifft den Bystander-Effekt (Darley & Latane', 1968), der eine Reduktion der Wahrscheinlichkeit individuellen Hilfeverhaltens bei Anwesenheit anderer potenziell hilfefähiger Personen beschreibt. In einer Meta-Analyse von 105 Studien zeigen Fischer et al. (2011), dass dieser Effekt in gefährlichen Notfällen deutlich reduziert bis gar nicht vorhanden ist. Insgesamt konnte ein umfangreicher Komplex an Moderatoren des Bystander-Effekts identifiziert werden. Dazu gehörten Geschlecht, Studienbedingung (experimentell vs. quasi-experimentell und Labor vs. Feld), persönliche Bekanntheit mit Bystandern und Anzahl der Bystander. Unter Kontrolle aller Moderatoren zeigte sich ein kleiner bis mittlerer negativer Effekt der

Anwesenheit von Bystander auf Hilfeverhalten, der jedoch deutlich unterhalb dem Ergebnis der Originalstudie lag. Die Meta-Analyse trug damit erheblich zur Präzision der Aussage des Bystander-Effekts bei.

Generell besteht in der psychologischen Forschung ein deutlicher Bedarf an Replikationsstudien. Nosek et al. (2015) zeigten im Rahmen der Open Science Collaboration in einer Replikation von 100 bekannten psychologischen Studien (genauer zur Auswahl in Nosek et al., 2015), dass sich nur in etwa 40 % der Fälle überhaupt ein Effekt entsprechend der jeweiligen Originalstudie zeigte, dieser jedoch etwa nur halb so groß war (Originalstudien: $M = 0.403$, $SD = 0.188$), Replikationsstudien: $M = 0.197$, $SD = 0.257$). Makel und Plucker (2014) stellten in der Psychologie einen Anteil von Replikationsstudien an insgesamt publizierten Studien von 1,07 % fest, was in Anbetracht des Falsifikationsprinzips, auf dem die betreffende Forschung basiert, mehr als inakzeptabel ist. Als Ursachen für die seltene Durchführung von Replikationsstudien gelten der geringe Reputationsgewinn, die Reduktion signifikanter Effekte, hohe Anforderungen an die Stichprobengewinnung, zu wenig detaillierte Studienprotokolle und sogenanntes p-Hacking (Baker, 2016; Brandt et al., 2014; Everett & Earp, 2015; Francis, 2012; Nosek et al., 2015; Earp & Trafimow, 2015; Tressoldi, 2012).

Aus methodischer Perspektive bedeutet der Einschluss einer Replikationsstudie in die statistische Analyse eine Erhöhung der Stichprobengröße bei gleichzeitiger Einführung neuer Prädiktoren. Verschiedene Stichproben additiv zu poolen, d. h. die Daten als eine einzige Stichprobe zu behandeln, erhöht das Risiko, den Standardfehler von Variablen künstlich zu unterschätzen, sodass es zur Überschätzung von Effektgrößen und deren Auftretenswahrscheinlichkeit kommt (Moeyaert, Ferron, Beretvas, & van den Noortgate, 2015). Selbst im Fall vollständig balancierter Datensätze führt eine fehlende Berücksichtigung der Studienidentität bei den meisten Schätzalgorithmen daher zu einem enormen Bias (Verma, Gagliardi, & Ferretti, 2009).

Meta-analytische Verfahren treffen zusätzliche methodologische Annahmen über die Verteilung von Messgrößen in der Population, die in den entsprechenden Regressionsgleichungen abgebildet sind. Gegeben sei eine multiple Regression mit der Gleichung:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\varepsilon} \quad (1)$$

wobei der Messfehler $\hat{\varepsilon} \sim N(0, \sigma)$ die Abweichung vom ‚wahren‘ Effekt in einer einzelnen Studie angibt und die Residualvarianz des Modells bindet. In der Regressionsanalyse einer einzelnen Studie geben die Regressionskoeffizienten $\hat{\beta}_j$ die geschätzten Effektgrößen der Prädiktoren x_j an. Im Unterschied dazu behandeln meta-analytische Modelle die Effektgröße von Studien als abhängige Variable, die nicht nur durch die originären Prädiktoren auf Variablenebene, sondern auch durch Prädiktoren auf Studienebene geschätzt wird (Viechtbauer, 2010). Dadurch ergeben sich große Ähnlichkeiten zu Multi-Level-Modellen in der Behandlung des Messfehlers. Meta-analytische Modelle differenzieren diesen genauer aus und zerlegen ihn je nach Modelltyp in weitere Varianzkomponenten.

Random-Effects Modelle treffen die methodologische Annahme, dass jede Studie $\hat{y}_i$ eine Effektgröße θ_i berichtet, die von einem einzigen ‚wahren‘ Wert der Effektgröße μ in der Population abweicht, sodass

$$\hat{\theta}_i = \mu + u_i \quad (2)$$

mit dem Messfehler $u_i \sim N(0, \tau^2)$. Bei minimaler Varianz der gefundenen Effekte $\tau^2 = 0$, entspricht der Durchschnitt der gefunden Effekte dem einen wahren Wert in der Population. Existiert Heterogenität der Effektgröße zwischen den Studien, sodass $\tau^2 \neq 0$, kann die zusätzliche Annahme getroffen werden, dass ein Teil des Messfehlers nicht zufallsbedingt, sondern systematisch ist. Dieser Teil des Messfehlers beschreibt einen moderierenden Faktor in der Varianzaufklärung und kann in die Regressionsgleichung aufgenommen werden. Dementsprechend ergibt sich ein Mixed-Effects Modell mit

$$\hat{\theta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_j x_{ij} + u_i + \varepsilon_i \quad (3)$$

wobei x_{ij} den Wert der j -ten Variable aus der i -ten Studie darstellt und der systematische Messfehler weiterhin durch $u_i \sim N(0, \tau^2)$, gegeben ist. Die Residualvarianz τ^2 gibt hier den nicht durch Gruppierungsvariablen erklärten Varianzanteil eines einzigen ‚wahren‘ Effekts in der Population an. In Matrix Notation ist damit ein multivariates Modell $\mathbf{T}_i \sim N(\boldsymbol{\theta}_i, \Sigma_i)$ mit $i = 1 \dots k$ Studien gegeben. Dessen verallgemeinerte Form lautet

$$\mathbf{T} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} + \boldsymbol{\varepsilon}. \quad (4)$$

mit dem Vektor der zu schätzenden Regressionsgewichte $\boldsymbol{\beta}$ mit $\beta = 1 \dots j$, der die Effekte der Prädiktor-Matrix $\mathbf{X}$ mit

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} \end{bmatrix} \quad (5)$$

beschreibt, sowie dem Random-Effects Vektor $\mathbf{U}$ mit $u = 1 \dots i$, der den zu schätzenden Intercept jeder definierten Gruppierungsvariable angibt. Die im Modell geschätzte Kovarianzmatrix $\hat{\mathbf{V}}$ der beiden Messfehlerterme $\mathbf{U} + \boldsymbol{\varepsilon}$ ist gegeben durch

$$\mathbf{V} = \text{Diag}[\Sigma_1, \dots, \Sigma_i] \quad (6)$$

was bspw. im GLS-Schätzer

$$\beta = [X'V^{-1}X]^{-1}X'V^{-1}T \quad (7)$$

resultiert (schrittweise Herleitung: Konstantopoulos, 2006). Fixed-Effects Modelle spitzen die Verteilungsannahme der Modelle mit Random-Effects Komponente weiter zu und gehen davon aus, dass es nicht einen einzigen ‚wahren‘ Wert der Effektgröße in der Population gibt, sondern verschiedene Effektgrößen existieren, die zwar schwanken, jedoch nicht um einen ‚wahren‘ Wert. Daher ist in Fixed-Effects Modellen

$$\theta_{\mu} = \frac{\theta_i}{k} \quad (8)$$

mit der Anzahl der Studien k als ungewichtetem Durchschnitt.

Die Auswahl des Regressionsmodells hat methodologische Konsequenzen für die angestrebte Reichweite der getroffenen Aussagen. Random- und Mixed-Effects Modelle gehen davon aus, dass die k Studien nur eine Zufallsstichprobe aus einer größeren hypothetischen Population von Studien sind. Sie zielen auf die prädiktive Schätzung einer Effektgröße in eben dieser hypothetischen und unbekannten Population. Das Modell impliziert damit, dass zukünftige Studien, die aus der hypothetischen Population gezogen werden, zu einem ähnlichen Ergebnis kommen müssten, wenn die Schätzung korrekt ist. Dagegen gehen Fixed-Effects Modelle davon aus, dass die k Studien die gesamte Population abbilden. Sie zielen auf die Schätzung einer durchschnittlichen Effektgröße dieser konkreten Verteilung. Heterogenität zwischen den Studien wird im Fixed-Effects Modell nicht als zufälliger Messfehler konzipiert, sondern als systematischer Unterschied. Das Modell impliziert damit, dass *verschiedene* ‚wahre‘ Effekte existieren. Eine zentrale methodische Konsequenz der Entscheidung für den Modelltyp betrifft die Erhöhung der Anzahl frei zu schätzender Parameter mit jeder definierten Random-Effects Komponente. Damit verbunden ist zwar eine größere Power von Mixed- und Random-effects Modellen, aber auch eine höhere Anforderung an die Stichprobe.

Mathematisch besteht augenscheinlich eine große Ähnlichkeit von multivariaten meta-analytischen Random-Effects Modellen zu multivariaten Multilevel-Modellen mit Random-Intercept für die gewählten Gruppierungsvariablen. In Abgrenzung zu hierarchischen linearen Regressionen (HLM) weisen meta-analytische Modelle jedoch zwei zentrale zusätzliche Merkmale auf. Erstens unterscheidet sich die Varianz der Prädiktoren zwischen den Studien, sodass diese als Gewichtungsfaktor in die Regressionsgleichung eingehen. Der modellübergreifende Intercept der Gleichung ergibt sich entsprechend aus

$$\widehat{\beta}_0 = \frac{\sum_i^k w_i \theta_i}{\sum_i^k w_i} \quad (9)$$

mit $w_i = \frac{1}{\sigma_i + \hat{\tau}^2}$ und $\beta_0 \sim N(0, \sigma.)$, sodass Studien mit geringerer Varianz einen größeren Effekt auf diesen Gleichungsterm haben. Die Varianz $\sigma.$ von $\widehat{\beta}_0$ beträgt dementsprechend das Reziproke der aufsummierten Regressionsgewichte mit

$$\sigma. = \frac{1}{\sum_i^k w_i} \quad (10)$$

Um die gruppenspezifische Varianz, d. h. die Heterogenität in einem meta-analytischen Regressionsmodell qualitativ einzuschätzen, existieren neben τ^2 auch weitere Maße (im Überblick: Higgins & Thompson, 2002).

Zweitens unterscheiden sich meta-analytische von einfachen Regressionsmodellen in der Aggregationsstrategie der Datenanalyse. Die Zusammenführung der individuellen Messwerte aus den verschiedenen Studien kann in einem Schritt durch eine direkte Erweiterung der Regressionsgleichungen um das gewählte Maß der Effektstärke erfolgen oder in zwei Schritten durch eine separate Berechnung der Effektstärke und deren anschließende numerische Übertragung in die Regressionsgleichung. Mit wenigen Ausnahmen, wie z. B. sehr große Datensätze, resultieren beide Strategien in quasi-identischen Werten (Kontopantelis, 2018). In der vorliegenden Forschungsarbeit wurde die Aggregation in zwei Schritten vorgenommen, um die zusätzlichen deskriptiven Kontrollmöglichkeiten auszuschöpfen. Für die Studie in Kapitel 8 wurden zur Schätzung der Effektstärken der entwickelten Intervention auf die jeweiligen Outcome-Variablen die jeweiligen Gruppenunterschiede der standardisierten Mittelwertsveränderungen (Standardized Mean Change Difference g) vom ersten Messzeitpunkt zum zweiten Messzeitpunkt berechnet (vor und nach der Intervention). Dazu wurde die Gleichung von Morris (2000) verwendet, in der g gegeben ist durch

$$g = c(n_T - 1) \frac{\mu_{post,T} - \mu_{pre,T}}{SD_{pre,T}} - c(n_C - 1) \frac{\mu_{post,C} - \mu_{pre,C}}{SD_{pre,C}} \quad (11)$$

mit $T = \text{Treatment-Group}$ und $C = \text{Control-Group}$ sowie dem Korrekturfaktor $c(n - 1)$ zur Reduktion von Bias in der Schätzung von Standardfehlern mit

$$c(n - 1) = \sqrt{\frac{2}{(n-1)}} \left(\frac{\Gamma\left[\frac{(n-1)}{2}\right]}{\Gamma\left[\frac{(n-2)}{2}\right]} \right) \quad (12)$$

in dem Γ die Eulersche Gammafunktion ist. Der Korrekturfaktor führt insbesondere in kleinen Stichproben zu einer genaueren Schätzung der Populationsvarianz als die einfache Standardabweichung (Morris, 2000). Die Verteilung der Gammafunktion ist eine Verallgemeinerung der Exponentialfunktion, so dass die Korrektur bei kleinen Fallzahlen größer ausfällt, als bei größeren Fallzahlen. Die Verwendung von g als Maß der Effektstärke ist besonders für Interventionsstudien geeignet, in denen für die genestete Struktur von Messwerten in Personen kontrolliert werden muss, um eine künstliche Erhöhung der Stichprobengröße und Bias durch die autokorrelative Datenstruktur auszuschließen.

Zusammengefasst bieten manifeste meta-analytische Regressionsmodelle in Kombination mit der Replikation von Studien eine pragmatische Möglichkeit zur messfehlerbereinigten Schätzung von Interventionseffekten in kleinen Stichproben. Die explizite Modellierung studienbedingter Variation schließt an das Falsifikationsprinzip im engeren Sinne an und löst den Trade-Off zwischen Spezifität und Komplexität auf der Designebene zum Teil auf. Im Gegensatz zu latenter Modellierung können die Analyseergebnisse bedingt durch den geringeren statistischen Komplexitätsgrad sehr direkt und intuitiv interpretiert werden. Gerade in der Forschung zur Lehrkräftebildung, in der Interventionseffekte über die Teilnehmenden in das schulische Bildungssystem hinein transportiert werden, sind Replikationsstudien vor einer systematischen Implementation unabdinglich. In der vorliegenden Forschungsarbeit berichtet die Studie in Kapitel 8 Replikationsergebnisse zur entwickelten Intervention.

Neben der Schätzung ‚wahrer‘ Effekte besteht in der Interventionsforschung auch Interesse an der Veränderung von Zusammenhängen über die Zeit. Eine entsprechende Modellierung längsschnittlicher Daten kann bspw. Aussagen darüber treffen, ob eine Intervention den Zusammenhang zwischen Dispositionen und Performanz fördert und somit zur Kompetenzentwicklung beiträgt. Aber auch zur Beschreibung von Entwicklungsverläufen und -verlaufsmustern in der Ausbildung kognitiver Strukturen ist eine längsschnittliche Modellierung von Zusammenhängen geeignet. Da in der Lehrkräftebildung nur selten Stichproben mit ausreichender Größe für eine latente Modellierung von Zusammenhangsveränderungen gewonnen werden kann, besteht ein besonderer Bedarf an Möglichkeiten manifester Modellierung.

5.2 Manifeste Modellierung von Zusammenhangsveränderungen

Eine zentrale Grundlage der Interventionsentwicklung sind quantitative Zusammenhangsanalysen, die potenziell kompetenzförderliche Prädiktoren oder Situationen identifizieren. Entsprechend der allgemeinen Kompetenzdefinition (Kapitel 2.4), die Kompetenz als ein kontextspezifisches funktionelles Zusammenwirken relevanter kognitiver Strukturen und Module beschreibt, sollten entsprechende Dispositionen einen kontextspezifischen Zusammenhang aufweisen. Interventionen zur Kompetenzentwicklung zielen daher aus methodischer Sicht unter anderem auf die Ausbildung intra-individueller Zusammenhänge zwischen psychologischen Konstrukten. Evidenz zur Veränderung von Zusammenhängen über die Zeit ist nur zuverlässig mit einer längsschnittlichen Modellierung von Daten herzustellen.

Die Forschung zur Lehrkräftebildung hat in Konsequenz der Besonderheiten des Forschungsfeldes (im Detail: siehe Kapitel 9.1) nicht nur einen generellen Bedarf an anwendbaren statistischen Modellen zur Zusammenhangsanalyse, dieser besteht insbesondere im Kontext diagnostischer Kompetenz. Mit den zunehmenden Anforderungen an inklusives Unterrichten besteht besonderer Bedarf an der längsschnittlichen Modellierung von Zusammenhängen möglicher Dispositionen diagnostischer Kompetenz (Casale, Hennemann, & Grosche, 2015; Hagiwara et al., 2019; Hillenbrand, Melzer, & Sung, 2014). Die intensive Interventionsforschung und Instrumentenentwicklung zu inklusionsbejahenden Einstellungen ist von den heterogenen Ausbildungsstrukturen für angehende Lehrkräfte und den damit verbundenen Einschränkungen in der Gewinnung großer und längsschnittlich angelegter Stichproben keinesfalls ausgenommen (Heinisch, Sonnleitner, & Rank, 2018; Hellmich & Görel, 2014; Merk, Cramer, Dai, Bohl, & Syring, 2018; Seifried & Heyl, 2016). Eine Folge dieser Konstellation ist der Rückgriff auf methodische Vorgehensweisen, die ein hohes Risiko für fehlerhafte Schätzungen tragen (s. u.).

Fragen nach der Veränderung von Zusammenhängen zielen auf die Schätzung der Korrelation intra-individuell korrelierter Variablenausprägungen bzw. auf die Schätzung des Anstiegs oder Abfalls dieser Korrelationen. Zusammenhangsanalysen treffen jedoch die Annahme unabhängiger Residuen (Eid, Gollwitzer und Schmidt 2011), die in dieser Datenstruktur nicht vorliegt. Dementsprechend müsste die Autokorrelation von Variablen zwischen Messzeitpunkten als eigenständige Variable modelliert werden, um die Residuen um diesen Varianzanteil zu bereinigen. Mit latenter Modellierung bestehen entsprechende Möglichkeiten, im Bereich manifester Verfahren stellt dies jedoch eine methodische Herausforderung dar. In der Forschung zu inklusionsbejahenden Einstellungen wird diese autokorrelative Struktur z. T. nicht berücksichtigt und stattdessen mit schrittweiser Korrelation oder deskriptiven Beschreibungen gearbeitet, was das Risiko fehlerhafter Ergebnisse und Interpretationen enorm steigert (exemplarisch: Batsiou, Bebetos, Panteli, & Antoniou., 2008; im Detail: siehe Kapitel 9.1).

Häufig fällt die Entscheidung darauf, die Datenstruktur so zu vereinfachen, dass nur ein einziger Zusammenhang berechnet werden muss. Dafür existieren verschiedene methodische Vorgehensweisen (Dunlap, Jones, & Bittler, 1983). Erstens könnte die Datenstruktur komplett ignoriert werden. Die Korrelation (r_{ignore}) würde dann jeden Datenpunkt als ein vollständiges Element der Population mit unabhängigem Varianzanteil behandeln. Dadurch erhöht sich jedoch einerseits die Stichprobengröße und andererseits sinkt die Varianz in den Messwerten, was zu einem erhöhten Risiko

führt, Effekte zu identifizieren, die nicht vorhanden sind. Zweitens könnte in einem schrittweisen Vorgehen eine Korrelation aus Mittelwerten der messwiederholten Datenpunkte jeder Person berechnet werden (r_{within}), was aber sowohl die Stichprobengröße um den Faktor der Anzahl von Messwiederholungen reduziert als auch die Varianz in den Daten. Drittens könnten zuerst Korrelationen innerhalb jedes Messzeitpunktes berechnet und dann gemittelt werden ($r_{between}$), was jedoch ebenfalls zur Unterschätzung der Varianz führt (Sainani, 2010). Die letzten beiden Verfahren der sukzessiven Aggregation senken die Teststärke, d. h. sie erhöhen das Risiko, einen vorhandenen Effekt nicht zu identifizieren.

Messwiederholte Daten können auf manifester Ebene prinzipiell mit messwiederholter Varianzanalyse (rmANOVA) modelliert werden. Das Modell zielt jedoch darauf, den Effekt von Stufen einer kategorialen Variable auf eine abhängige Variable um den Varianzanteils des autokorrelativen Zusammenhangs auf intra-individueller Ebene zu bereinigen. Erweitert man das Modell um eine kontinuierliche Variable als Kovariate, zielt das Modell darauf, die Gesamtmittelwerte der Stufen der kategorialen Variable um die Kovarianz zwischen abhängiger Variable und Kovariate zu bereinigen und weiterhin die Ausprägung der abhängigen Variable auf Unterschiede zwischen den Stufen der kategorialen Variable zu testen (Davison & Sharma, 1994). Interessiert jedoch nicht der Unterschied zwischen diesen Stufen, sondern der Zusammenhang zwischen Kovariate und abhängiger Variable, kann das Modell so modifiziert werden, dass eben diese Korrelation unter Berücksichtigung der messwiederholten Datenstruktur berechnet wird.

Dazu stellten Bland und Altman (1995) die messwiederholte Korrelation vor. Diese basiert auf einer einfaktoriellen ANCOVA und schätzt eine Korrelation zwischen zwei Variablen unter Berücksichtigung systematischer Personenunterschiede. In dieser ursprünglichen Fassung zielten Bland und Altman (1995) jedoch auf die praktische Anwendung, so dass keine Gleichungen ausformuliert wurden. Bakdash und Marusich (2017) griffen dieses statistische Modell für eine Softwareentwicklung auf (rmcorr) und explizierten vor diesem Hintergrund auch die zugehörigen Gleichungen. Im Gegensatz zur klassischen Anwendung der ANCOVA für die Schätzung von inter-individuellen Unterschieden wird in der messwiederholten Korrelation (rmcorr) für jede Person eine dichotome Indikatorvariable gebildet, sodass die Anzahl der Personen der Anzahl der Faktorstufen entspricht und die Kovariate der jeweiligen Ausprägung des Faktors zum jeweiligen Messzeitpunkt. Diese Modellspezifikation führt dazu, dass die intra-individuelle Korrelation die messwiederholte Struktur berücksichtigt und um die inter-individuellen Unterschiede bereinigt wird.

Bei genauerer Betrachtung der Gleichungen in Bagdash und Marusich (2017) im Rahmen dieser Dissertation wurden jedoch Fehler in der mathematischen Umformung von der allgemeinen ANCOVA Gleichung hin zur Gleichung der messwiederholten Korrelation gefunden und korrigiert (siehe Fußnote in Bagdash & Marusich, 2019). Im Folgenden wird daher zunächst die Korrektur dargestellt, um daran anschließend weitere Analysemöglichkeiten des Modells aufzuzeigen.

Den Ausgangspunkt der Zusammenhangsschätzung bildet das allgemeine ANCOVA Modell

$$Y_{ij} = \mu + \tau_j + \beta(X_{ij} - \bar{X}_j) + \varepsilon_{ij} \quad (13)$$

mit der abhängigen Variable Y_{ij} , dem globalen Mittelwert μ , dem Effekt der jeweiligen Faktorstufe τ_j sowie dem Term der Kovariate X_{ij} und dem Term des Messfehlers ε_{ij} . In Bakdash und Marusich (2017) wurde bei der Reformulierung des Modelles in konventioneller Darstellung als GLM der globale Mittelwert μ nicht als Intercept der abhängigen Variable interpretiert, d. h. als Mittelwert aller messwiederholten Werte der abhängigen Variable, sondern als Mittelwert aller messwiederholten Werte der Kovariate. Dies führt zu einer systematischen Unterschätzung der Korrelation bzw. des Anstiegs der geschätzten Regressionsgerade. So wurde die Gleichung im Rahmen eines generalisierten linearen Modells (GLM) als Wortgleichung formuliert durch

$$Measure1_{ij} = \overline{Measure2}_i + Participant_i + \beta(Measure2_{ij} - \overline{Measure2}_i) + \varepsilon_{ij} \quad (14)$$

und lautet korrigiert

$$Measure1_{ij} = \overline{Measure1}_i + Participant_i + \beta(Measure2_{ij} - \overline{Measure2}_i) + \varepsilon_{ij} \quad (15)$$

Das Modell hat eine offenkundige Ähnlichkeit mit einem hierarchischen linearen Modell (HLM) mit Random-Intercept und Fixed-Slope Komponenten, da der Effekt der Faktorstufe τ_j einem Random-Intercept für jede Gruppe von Messwerten bzw. Personen entspricht. Der zentrale Unterschied beider Methoden besteht in der Schätzung von Varianzkomponenten. Multilevel-Modelle treffen die Annahme, dass sich die Varianzen innerhalb und zwischen den Faktorstufen gegenseitig beeinflussen, was die Anwendung von Partial Pooling und damit die gleichzeitige Schätzung dieser Varianzkomponenten ermöglicht. Partial Pooling ist eine Schätzstrategie, in der für die Bestimmung von Effekten stichprobenbezogen kleinerer Faktoren bzw. Gruppen die Daten aus den umfangreicheren Faktoren bzw. Gruppen genutzt werden (Gelman, 2006). Dagegen erfolgt die Schätzung in einer ANCOVA bzw. einem einfachen GLM schrittweise, was die Genauigkeit der Schätzung für Faktoren bzw. Gruppen mit wenigen Messwerten limitiert aber im Gegenzug nur eine verhältnismäßig kleine Stichprobe voraussetzt.

Für die Berechnung einer messwiederholten Korrelation werden die Summe der Abweichungsquadrate der Kovariate und des Messfehlers genutzt (Bland & Altman, 1995). Der Effekt der Zwischenssubjektebene wird dazu als eigenständige Varianzkomponente vom Innersubjekteffekt getrennt und in der letztendlichen Berechnung der messwiederholten Korrelation nicht weiterverwendet. Demgegenüber können aus der für autokorrelierte Daten verwendeten ANCOVA Gleichung jedoch auch weitere Modellierungsmöglichkeiten abgeleitet werden. Da auch ein Regressionskoeffizient geschätzt wird, kann dieser ebenfalls interpretiert werden. In der ANCOVA Gleichung wird dieser um den Trend aller Gruppen bzw. Personen korrigiert, da dieser Trend einen eigenständigen Faktor bildet. Dies ist der globale Mittelwert μ . Die Korrektur hat zur Folge, dass Ausreißer in einzelnen Messzeitpunkten weniger Bias in die Schätzung einführen als bei einer nur auf den einzelnen Gruppen bzw. Personen basierenden Schätzung. Darüber hinaus wird die Genauigkeit der Schätzung bei Stichproben mit wenig Messwiederholungen deutlich erhöht. Der

Regressionskoeffizient kann dementsprechend als hierarchisch korrigierte Korrelation interpretiert werden.

Der korrigierte Regressionskoeffizient gibt neben der Stärke des Zusammenhangs durch die Messwiederholung auch dessen Veränderung über die Zeit an. Das Vorzeichen des Koeffizienten zeigt dementsprechend an, ob ein Zusammenhang wächst oder schrumpft (siehe Abbildung 9). Eine direkt anschließende Konsequenz ist die Möglichkeit des Vergleichs der so geschätzten Koeffizienten zwischen verschiedenen Personengruppen. Dazu kann bspw. in einer Interventions- und einer Kontrollgruppe in getrennten Modellen eine Schätzung erfolgen, die sich dann mit t -Tests für den β -Koeffizienten (Cohen, Cohen, West, & Aiken, 2003) auf Unterschiede prüfen lässt. Dieses statistische Vorgehen lässt also Aussagen über Unterschiede in Zusammenhangsveränderungen zu. Für Aussagen dieser Art werden auf latenter Ebene Wachstumskurvenmodelle geschätzt, die eine etwa zehnfach größere Stichprobe benötigen, aber dementsprechend auch nicht-lineare Veränderungen abbilden können (Hamilton, Gagné, & Hancock, 2003; Seaman, Algina, & Oljenik, 1985). Inwiefern mit diesem manifesten Modellierungsverfahren von linearen Wachstumskurven auch nicht-lineare Funktionen akkurat geschätzt werden können, ist bislang eine offene Frage, an die Simulationsstudien anschließen können.

Zusammengefasst eröffnet diese manifeste Modellierung von Zusammenhangsveränderungen der Forschung die Möglichkeit, komplexe Fragestellungen auch mit kleinen Stichproben zu überprüfen. Das GLM und insbesondere die Varianzanalyse sind sehr flexible Verfahren zur manifesten Modellierung von Daten, zu dem nach wie vor immer wieder neue Anwendungen entdeckt werden (Schneider, Avivi-Reich, & Mozuraitis, 2015). Im Kontext der Interventionsforschung können mit der vorgestellten Datenmodellierung Fragen zur Wirksamkeit von Maßnahmen nicht auf inter-individueller, sondern auch auf intra-individueller Ebene beantwortet werden. Da Kompetenzentwicklung auf die funktionelle Verknüpfung kognitiver Module und Strukturen zielt, sollte deren Zusammenhang im Verlauf einer Intervention statistisch gesehen steigen. In der vorliegenden Forschungsarbeit beschäftigt sich Kapitel 9 mit einer solchen Frage und prüft mit dem entwickelten statistischen Modell die Veränderung des Zusammenhangs von Diagnosewissen und inklusionsbejahenden Einstellungen in der für die Forschungsarbeit entwickelten Intervention.

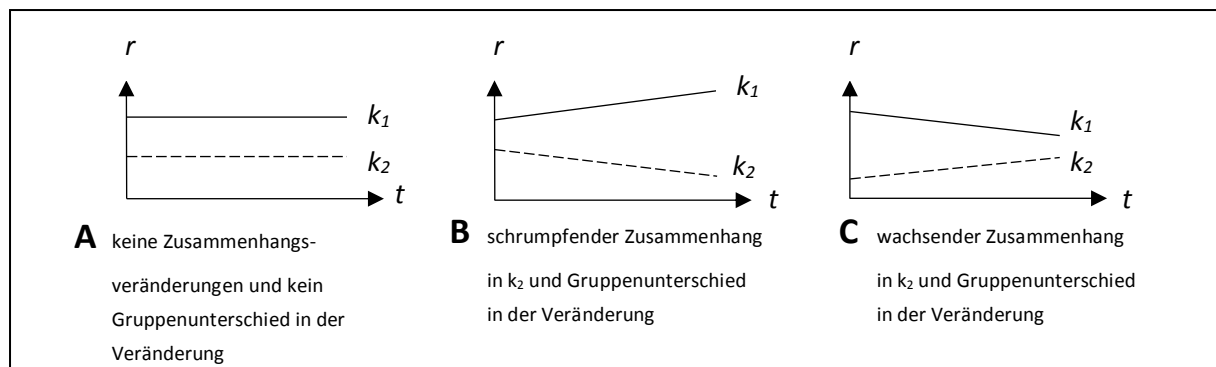


Abbildung 9. Lineare Zusammenhangsveränderungen. Veränderung des Zusammenhangs r in den Gruppen k_1 und k_2 über die Zeit t . Eigene Darstellung.

6 Ziele, Fragestellungen und Hypothesen

Die vorliegende Forschungsarbeit hat das Ziel, einen Beitrag zum Kenntnisstand über förderliche Bedingungen der Entwicklung diagnostischer Kompetenz angehender Lehrkräfte zu leisten. Der dargestellte theoretische Hintergrund und die empirischen Befunde in diesem Kontext weisen sowohl die Verknüpfung fallbezogenen und theoretischen Wissens als auch das kooperative diagnostische Handeln als zentrale Entwicklungsaspekte aus. Demgegenüber fehlen Erkenntnisse darüber, inwiefern eine kooperative Lernumgebung die Entwicklung von Diagnosewissen anstoßen kann (siehe Kapitel 4). Der spezifische Beitrag der vorliegenden Dissertation zu diesem Forschungsstand liegt in der empirischen Untersuchung einer kooperativ und problemorientiert angelegten Intervention zur Förderung diagnostischer Kompetenz. Diese wurde im Rahmen des Dissertationsvorhabens entwickelt und ist auf selbstgesteuertes Lernen mit Textfällen ausgerichtet. Die ersten drei Studien fokussieren jeweils die interventionsbedingte Veränderung des Diagnosewissens im Kontrast zu einer Kontrollgruppe, differenzieren diese Veränderung jedoch in unterschiedliche Richtung und mit unterschiedlichen Analyseverfahren aus. Die übergreifende Forschungsfrage dazu lautet:

Inwiefern trägt eine kooperativ und problemorientiert angelegte Intervention zur Entwicklung der diagnostischen Kompetenz angehender Lehrkräfte bei?

Die erste Studie in Kapitel 7 geht von dem identifizierten Erkenntnismangel zu Interventionen aus, denen kooperative Anforderungssituationen aus dem Bereich der Diagnostik zugrunde liegen. Die Studie arbeitet daher besonders den kooperativen Charakter der Intervention heraus und vergleicht Effekte auf die Kompetenzentwicklung von Lehramtsstudierenden in einem quasi-experimentellen Design. Ziel der Studie ist eine Prüfung von Effekten der entwickelten Intervention auf das Diagnosewissen von Lehramtsstudierenden, das Fähigkeitsselbstkonzept im Bereich Diagnostik und die Motivation zur Beschäftigung mit Diagnostik. Diese Auswahl erfolgte systematisch vor dem Hintergrund der zentralen Bedeutung des Wissens und der Überzeugungen in allgemeinen und diagnostikspezifischen Kompetenzmodellen (siehe Kapitel 2.2 und 3.1), sowie von Weinert (1999) eingeforderten Kontrolle motivationaler Variablen in Kompetenzmessungen. Folgende Hypothesen mit einem varianzanalytischen Vorgehen geprüft:

1. Problemorientiertes Lernen mit Textfällen hat einen positiven Effekt auf die Entwicklung des Diagnosewissens. D. h. der Anstieg des Diagnosewissens ist in der Interventionsgruppe größer als in der Kontrollgruppe.
2. Problemorientiertes Lernen mit Textfällen hat einen positiven Effekt auf die Entwicklung des Fähigkeitsselbstkonzepts im Bereich der Diagnostik. D. h. der Anstieg des Fähigkeitsselbstkonzepts ist in der Interventionsgruppe größer als in der Kontrollgruppe.
3. Problemorientiertes Lernen mit Textfällen hat einen positiven Effekt auf die Entwicklung der Motivation zur Beschäftigung mit Diagnostik. D. h. der Anstieg der Motivation ist in der Interventionsgruppe größer als in der Kontrollgruppe.

Die zweite Studie in Kapitel 8 reagiert auf die Replikationskrise in der Psychologie (Nosek et al., 2015) und untersucht die Reproduzierbarkeit der Interventionseffekte über verschiedene Kohorten hinweg mit einer exakten Replikation. In einem ersten Schritt werden daher dieselben Hypothesen mit demselben Studiendesign und denselben Analysemethoden untersucht. In einem zweiten Schritt wird

mit meta-analytischen Methoden einerseits die Reproduzierbarkeit bzw. Heterogenität der Studien untersucht. Andererseits lässt sich dadurch trotz der jeweils kleinen Stichproben eine Aussage über um studienbedingte Messfehler bereinigte Effekte treffen. Die zweite Studie fokussiert darüber hinaus stärker den Aspekt der systematischen Konstruktion von diagnostischen Textfällen. Analog zur ersten Studie werden folgende Hypothesen geprüft:

1. Die entwickelte Intervention hat auch in einer Replikationsstudie einen positiven Effekt auf das Diagnosewissen.
2. Die Intervention hat auch in einer Replikationsstudie einen positiven Effekt auf das Fähigkeitsselbstkonzept im Bereich Diagnostik.
3. Die Intervention hat auch in einer Replikationsstudie einen positiven Effekt auf die Motivation zur Beschäftigung mit Diagnostik.

Die dritte Studie in Kapitel 9 greift die theoretische Auseinandersetzung mit dem Kompetenzkonstrukt sowie den weiterhin bestehenden Bedarf an empirischen Untersuchungen zur Struktur diagnostischer Kompetenz auf. Entsprechend der Kompetenzdefinition in Kapitel 2.4 zeichnet sich Kompetenzentwicklung auch durch einen wachsenden Zusammenhang von kontextspezifisch relevanten kognitiven Modulen aus. Die Studie untersucht daher die Veränderung des Zusammenhangs von Diagnosewissen mit inklusionsbejahenden Einstellungen und dem diagnostischen Fähigkeitsselbstkonzept im Laufe der Intervention und im Vergleich zu einer Kontrollgruppe. Dazu kommt ein eigens entwickeltes methodisches Verfahren zum Einsatz, das trotz der geringen Stichprobengröße ähnliche Aussagen erlaubt, wie sie mit Wachstumskurvenmodellen möglich sind. Dadurch leistet die Studie neben dem inhaltlichen Erkenntnisgewinn einen methodischen Beitrag zur Interventionsforschung mit kleinen Stichproben. Folgende Hypothesen werden untersucht:

1. Es besteht ein positiver Zusammenhang zwischen dem Diagnosewissen, der Einstellung zu Inklusion und dem diagnostischen Selbstkonzept in beiden Gruppen.
2. Die Zusammenhänge sind in der Interventionsgruppe größer als in der Kontrollgruppe.
3. Die Zusammenhänge wachsen in der Interventionsgruppe stärker als in der Kontrollgruppe.

Die theoretische Auseinandersetzung mit Kompetenz in Kapitel 2 stellte heraus, dass Kompetenz in der Bildungsforschung überwiegend als Konstrukt verstanden wird, dass alle kognitiven Module und Strukturen umfasst, die den nicht auf kognitive bzw. sozial-kognitive Grundfähigkeiten zurückzuführenden Varianzanteil in einer gemessenen Performanz erklären. Für Kompetenzentwicklung stellt sich daher nicht nur die Frage, welche kognitiven Dispositionen förderlich sind, sondern auch, welche Performanz als Indikator für diagnostische Kompetenz verwendet werden kann. Da ein Großteil verfügbarer theoretischer Modelle diagnostische Kompetenz auf Urteilsgenauigkeit reduzieren, ohne den Effekt kognitiver Grundfähigkeiten auf diese Performanz zu prüfen (einzige Ausnahme: Kaiser et al., 2012), besteht dringender Bedarf an entsprechenden Untersuchungen. Eine vierte Studie untersucht daher experimentell in einer fallbasierten Testumgebung, inwiefern Zusammenhänge der diagnostischen Urteilsgenauigkeit mit Intelligenz und Empathie bestehen. Damit wird der Aspekt der kognitiven Dispositionen diagnostischer Kompetenz grundlagenorientiert vertieft und ein Beitrag zum Kenntnisstand über Komponenten diagnostischer Kompetenz geleistet. Weiterhin zeigt die in Kapitel 4.1 dargestellte Expertiseforschung zu mentalen

Modellen, dass diese in ihrer aggregierten Form im Sinne fallbezogenen Wissens zu kompetentem Handeln befähigen. Daher untersucht die Studie ebenfalls den Effekt unterschiedlich stark elaborierter mentaler Modelle diagnostischer Textfälle auf die Urteilsgenauigkeit. Die grundlegende Fragestellung lautet:

Inwiefern haben kognitive Grundfähigkeiten und der Elaborationsgrad mentaler Modelle einen Effekt auf die diagnostische Urteilsgenauigkeit angehender Lehrkräfte?

Da sich mentale Modelle je nach Situation unterscheiden, sollte in einer Untersuchung dieser Fragestellung neben der Messung von Dispositionen im Sinne von Personenmerkmalen auch eine Kontrolle der Situationsmerkmale erfolgen. In einer auf Textfällen basierenden Testumgebung sind sowohl der Textstil als auch der Textinhalt wichtige Merkmale der Anforderungssituation. Daher werden diese als Designfaktoren der Experimentalstudie in einem 2x2 Design manipuliert. Folgende Hypothesen werden regressionsanalytisch untersucht:

1. Fluide Intelligenz hat einen positiven Effekt auf die Urteilsgenauigkeit von Lehramtsstudierenden.
2. Kognitive Trait-Empathie hat einen positiven Effekt auf die Urteilsgenauigkeit von Lehramtsstudierenden.
3. Affektive State-Empathie hat einen positiven Effekt auf die Urteilsgenauigkeit von Lehramtsstudierenden.
4. Ein maximal elaboriertes mentales Modell (das Situationsmodell) hat einen positiven Effekt auf die Urteilsgenauigkeit.
5. Ein psycho-narrativer Textstil (d. h. eine ‚fallartig‘ narrative Kontextualisierung) der vorgestellten Diagnosefälle hat einen positiven Effekt auf die Urteilsgenauigkeit.

Vor der Darstellung der durchgeführten Studien erfolgt zur besseren Nachvollziehbarkeit der Analysen eine methodische Diskussion ausgewählter Verfahren in Kapitel 5. Da die statistischen Analyseverfahren für die Hypothesenprüfung der Studien zwei und drei (Kapitel 8 und 9) über die in sozialwissenschaftlichen Standardlehrbüchern vermittelten Methoden hinausgehen (z. B. Döhring & Bortz, 2016; Eid, Gollwitzer, & Schmitt, 2011; Sedlmeier & Renkewitz, 2013), fokussiert sich der anschließende Methodenteil auf deren Diskussion.

Empirischer Teil

7 Studie 1: Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen¹

7.1 Einleitung und theoretischer Hintergrund

Diagnostische Kompetenz ist eine zentrale Facette des professionellen Handelns von Lehrkräften und eine zentrale Voraussetzung eines lernwirksamen und inklusiven Unterrichts (Baumert & Kunter, 2006; Greiten, Geber, Gruhn, & Königer, 2018). Um unter den Bedingungen inklusiven Unterrichts mit heterogenen Lerngruppen einen adaptiven Unterricht durchführen zu können, sind alle Lehrkräfte zur Entwicklung ihrer diagnostischen Wissensbestände und Fähigkeiten aufgerufen (KMK, 2004; KMK, 2011). Der systematische Einbezug von Lernvoraussetzungen in das unterrichtliche Handeln ist notwendig, um Unterricht zu planen, Schülerinnen und Schüler in ihrem Lernprozess individuell zu fördern, aber auch, um langfristig unterstützende Selektionsentscheidungen in schulischen Laufbahnen treffen zu können (z. B. Übergangsempfehlungen). Dennoch variieren die inklusionsbezogenen Kompetenzen von Lehrkräften enorm (OECD, 2013). Daher wird in der Lehrkräftebildung verstärkt nach evidenzbasierten Lerngelegenheiten zur gezielten Förderung diagnostischer Kompetenz gesucht (Ricken, 2017; Schrader, 2017).

Zwei besondere Herausforderungen sind dabei die Vermittlung praxisbezogenen Wissens und die Anschlussfähigkeit der Kompetenzentwicklung im Studium an weitere Lernprozesse im Beruf (Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper, & Lautenbach, 2016). Besonders in der ersten Phase der Lehrkräftebildung, in der die Hochschulen nicht systematisch an pädagogische Vorerfahrungen von Studierenden anknüpfen können, haben sich fallbasierte Lerngelegenheiten als wirksam für die Kompetenzentwicklung erwiesen (Stürmer, Königs, & Seidel, 2013; Syring et al., 2015).

Bereits entwickelte Lerngelegenheiten orientieren sich überwiegend an kognitiven Ansätzen der Kompetenzentwicklung (Stahnke et al., 2016). Im Fokus dieser Ansätze steht die individuelle Förderung kontextspezifischer und stabiler Dispositionen (z. B. Wissen, Überzeugungen, Fähigkeiten), die zu professionellem Handeln befähigen sollen (Baumert und Kunter, 2006; Shavelson, 2013). Weit weniger beforscht sind sozial-interaktive Ansätze, die auf den Aufbau kollektiver Lern- und Handlungsformen zielen, die situative Rahmenbedingungen wie ‚professional communities‘ stärker einbeziehen und damit an die Schul- und Unterrichtspraxis anschlussfähig sind (Ghousseini & Herbst, 2016; Hoth et al., 2016; Santagata & Yeh, 2016). Inwiefern Lerngelegenheiten mit sozial-interaktivem Ansatz zur

¹ Diese empirische Studie wurde in ähnlicher Form 2020 veröffentlicht und kann wie folgt zitiert werden: Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A. (2020). Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen. In I. Gogolin, B. Hannover, & A. Scheunpflug (Hrsg.), *Evidenzbasierung in der Lehrkräftebildung. Edition der Zeitschrift für Erziehungswissenschaft Band 4* (S. 95-122). Wiesbaden: Springer VS. Postprint Version. https://doi.org/10.1007/978-3-658-22460-8_5
Reprinted by permission from Springer Nature: Springer, Zeitschrift für Erziehungswissenschaften, Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen, Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A., © 2020.

Entwicklung diagnostischer Kompetenz beitragen, wurde bislang nicht untersucht. Insbesondere fehlen hier längsschnittlich angelegte empirische Studien (Buholzer & Zulliger, 2013; Schrader, 2017). Die vorliegende Interventionsstudie greift diese Forschungslücke auf und untersucht die Effekte einer solchen Lerngelegenheit in der ersten Phase der Lehrkräftebildung.

Individuelle diagnostische Kompetenz von Lehrkräften

Diagnostische Kompetenz wurde in früheren Studien häufig durch die Urteilsgenauigkeit von Lehrkräften abgebildet, d. h. durch die Übereinstimmung der Lehrkräfteeinschätzung von Lern- und Leistungsmerkmalen mit den objektiven Testergebnissen dieser Schülerinnen und Schüler (Schrader und Helmke, 1987). Anschließende Forschung konnte für die Urteilsgenauigkeit aber weder einen positiven Effekt auf schulische Leistungen nachweisen (Behrmann & Souvignier, 2013), noch einen für Kompetenz erwartbaren Zusammenhang mit der Dauer der Lehrerfahrung (Hoffmann & Böhme, 2014). Weiterhin konnte eine Studie von Kaiser, Helm, Retelsdorf, Südkamp und Kaiser (2012) mit angehenden Lehrkräften zeigen, dass die Urteilsgenauigkeit zu schulischer Leistung im simulierten Klassenraum am ehesten durch allgemeine kognitive Fähigkeiten vorhergesagt wird. Inhaltlich besteht insofern Kritik am Modell der Urteilsgenauigkeit, dass eine genaue Einschätzung der Lernenden nur eingeschränkt die Auswahl und Zielsetzung pädagogischer Anschlusshandlungen (z. B. die Selektion oder Förderung) informieren kann (van Ophuysen & Behrmann, 2015).

Neben der Urteilsgenauigkeit existieren mittlerweile alternative Modelle diagnostischer Kompetenz, die ihren Fokus auf die Informationsgenerierung und -verarbeitung legen (Krolak-Schwerdt, Böhmer, & Gräsel, 2012). Besonders ressourcenorientiert sind Modelle die dem ‚response-to-intervention‘ (RTI) Prinzip folgen (Fuchs, 2003). RTI bedeutet, dass die Informationen über Lernvoraussetzungen von Schülerinnen und Schülern durch den Effekt von Präventions- und Fördermaßnahmen generiert werden. Eine Meta-Analyse von Burns, Appleton und Stehouwer (2005) zeigt, dass die Anwendung von RTI im Unterricht einen großen positiven Effekt auf die schulische Leistung hat ($d = 1.02$). Entsprechend bezieht sich die vorliegende Studie auf das Modell diagnostischer Kompetenz von Klug (2011), das RTI als integralen Bestandteil pädagogischer Diagnostik betrachtet.

Klug, Bruder und Schmitz (2016) beschreiben diagnostische Kompetenz als überfachliche Kompetenz der Beurteilung und Förderung des Lernverhaltens von Schülerinnen und Schülern. Das Modell unterscheidet drei aufeinander aufbauende Phasen individuellen diagnostischen Handelns. Die präaktionale Phase beinhaltet die kognitiven Prozesse der Zielsetzung förderrelevanten Handelns und der Aktivierung methodischen und theoretischen Wissens. Die daran anschließende aktionale Phase beschreibt das hypothesenbasierte Sammeln und Interpretieren von Informationen mit dem Ziel eines vorübergehenden diagnostischen Urteils. In der postaktionalen Phase folgen die Kommunikation mit den Beteiligten (z. B. Eltern, Schülerinnen und Schüler) und die Förderplanung. Das Modell fokussiert also primär individuelle Reflexions- und Handlungsprozesse, benennt aber neben individuellen Dispositionen auch professionelle Interaktionen als Ressource für diese Vorgänge. Es folgt grundlegend dem Rubikon-Modell der Handlungsphasen (Heckhausen & Gollwitzer, 1987) und eignet sich auch zur Anwendung auf individuelle Lernvoraussetzungen und -störungen im Allgemeinen. Die

dreidimensionale Struktur des Modells wurde mittels konfirmatorischer Faktoranalyse eines schriftlichen Szenario-Tests bestätigt (Klug, Bruder, Keller, & Schmitz, 2012). Lehrkräfte sowie Lehrkräfte im Vorbereitungsdienst erreichten in diesem Test höhere Werte als Studierende. Die stärksten Prädiktoren der Leistung im Szenario-Test waren das Wissen, die Motivation und die Einstellung zu Diagnostik sowie das professionelle Selbstkonzept (Klug et al., 2012; Klug et al., 2016).

Diese Ergebnisse entsprechen dem COACTIV-Modell professioneller Lehrkompetenzen, das sowohl kognitive als auch affektiv-volitionale individuelle Dispositionen als Voraussetzung professionellen Handelns definiert (Kunter, Klusmann, & Baumert, 2009). Kompetenzen sind in diesem Modell situationsspezifische Wahrnehmungs-, Interpretations- und Entscheidungsfähigkeiten, deren notwendige Voraussetzung stabile individuelle Dispositionen sind und die in konkrete Performanz, d. h. beobachtbarem Handeln münden. Im Fokus steht somit die intraindividuelle Ebene, auf der eine lineare Beziehungen zwischen diesen drei Konstrukten angenommen wird (Blömeke, Gustafsson, & Shavelson, 2015). Die kognitiven Dispositionen des COACTIV-Modells umfassen fünf verschiedene Facetten des Professionswissens, wobei Wissen über fachübergreifende Prinzipien des Diagnostizierens, Prüfens und Bewertens in den Facetten des fachdidaktischen Wissens und des pädagogisch-psychologischen Wissens verortet werden. Die affektiv-volitionalen Dispositionen umfassen Selbstregulation, Motivation und Überzeugungen („beliefs“). Motivation wird u. a. im Rahmen der Selbstbestimmungstheorie von Deci und Ryan (2000) als Grad des positiven emotionalen Erlebens während einer Tätigkeit definiert. Im Bereich der Überzeugungen unterscheidet das COACTIV-Modell zwischen epistemologischen Überzeugungen, subjektiven Lerntheorien, gegenstandsbezogenen subjektiven Lehrtheorien und selbstbezogenen Fähigkeitskognitionen. Letztere stellen nach Dickhäuser (2002) einen Teil des Selbstkonzeptes dar. Pädagogisches Diagnosewissen, Motivation zu pädagogischer Diagnostik und das diagnostische Selbstkonzept stellen damit zentrale Dispositionen diagnostischer Kompetenz dar.

Interaktion im diagnostischen Handeln von Lehrkräften

Wenngleich individuelle Dispositionen eine notwendige Voraussetzung für kompetentes Handeln sind, ist die Annahme, Performanz sei zum größten Teil auf individuelle Kompetenz zurückzuführen, gerade im Fall pädagogischer Diagnostik nur in ausgewählten Handlungsbereichen zutreffend. Dies ist für den gut untersuchten Bereich der individuellen Handlungen der Fall. Dazu gehören bspw. die Einschätzung von Schülerinnen und Schülern, die Durchführung und Dokumentation von Test- und Beobachtungsverfahren im Unterricht sowie die binnendifferenzierende Planung des eigenen Unterrichts. Für die Performanz in anderen diagnostischen Handlungsbereichen ist neben individueller Kompetenz auch die Zusammenarbeit von Personen zentral für kompetentes Handeln. Kollegiale Fallbesprechungen, die multiprofessionelle Zusammenarbeit in der Förderplanung und -durchführung sowie die Kommunikation mit Eltern und anderen Beteiligten sind einige Beispiele dafür. Dementsprechend benennen einige Modelle diagnostischer Kompetenz explizit interaktive Handlungsbereiche, wenngleich diese nicht immer systematisch ausdifferenziert sind (Klug et al., 2016; Ingenkamp & Lissmann, 2008). Behrmann und van Ophuysen (2017) legen ein Modell für die Prozess-

und Ergebnisqualität pädagogischer Diagnostik vor, das interaktive diagnostische Handlungen explizit einfordert (z. B. Multiperspektivität). Diese Anforderung an kompetentes Diagnostizieren ist darin begründet, dass Informationen über das Lernverhalten und die Lernvoraussetzungen von Schülerinnen und Schülern auch außerunterrichtlich und von verschiedenen Personen um eine Schülerin oder einen Schüler herum erfasst und genutzt werden.

Dass die Beziehung zwischen Kompetenz und Performanz als linearer und intraindividueller Prozess konzipiert wird, kritisieren auch andere Studien (Scheiner, 2016; Stahnke, Schueler, & Roesken-Winter, 2016). Santagata & Yeh (2016) kommen auf Grundlage von Unterrichtsvideos und Interviews zu dem Ergebnis, dass die Definition dessen, was ‚guter‘ bzw. ‚kompetenter‘ Unterricht ist und ihn ermöglicht, weder zeitlich stabil noch situationsunabhängig ist. Vielmehr sei Kompetenz in der Praxis der Lehrkräfte ein beständig neu verhandeltes Konstrukt, das insbesondere von den im jeweiligen Arbeitskontext relevanten professionellen Zusammenschlüssen geprägt werde. Dieser Befund relativiert die Annahme einer unidirektionalen Beziehung zwischen Kompetenz und Performanz zugunsten einer interaktionistischen Perspektive. In dieser bleibt zunächst offen, inwiefern in eine Interaktion involvierte Personen die gezeigte Performanz auf Kompetenzen einzelner Person zurückführen oder nicht. Kompetenz ist in dieser Perspektive eher durch Zuschreibungs- und Aneignungsprozesse bestimmt als durch stabile individuelle Dispositionen. Wenngleich die Vereinbarkeit einer interaktionistischen Perspektive mit der Definition von Standards der Lehrkräftebildung einige Fragen aufwirft, rückt sie dennoch professionelle Interaktionen und Zusammenschlüsse als wichtiges Lern- und Entwicklungsfeld der Lehrkräftebildung in den Blick.

Eine Meta-Studie zu Effekten und Voraussetzungen verschiedener Formen professioneller Zusammenarbeit an Schulen von Vangrieken, Dochy, Raes und Kyndt (2015) nutzt für eine grundlegende Definition professioneller Zusammenschlüsse das sozialpsychologische Konstrukt der Entitativität. Entitativität beschreibt das Ausmaß, in dem eine Ansammlung von Personen als miteinander verbundene, kohärente Einheit wahrgenommen wird (Campbell, 1958; Lickel et al., 2000). Grundlegende Merkmale sind: (1) Geteilte Ziele und Verantwortungsbereiche, (2) Aufgabenorientierte Kohäsion, (3) Identifikation mit dem Zusammenschluss, (4) Wechselseitige Abhängigkeit in der Aufgabenerfüllung und (5) Wechselseitige Abhängigkeit in der Zielerreichung. Professionelle Zusammenschlüsse haben sowohl einen positiven Effekt auf die beruflichen Lernvorgänge erfahrener Lehrkräfte (Vescio, Ross, & Adams, 2008; Toom, 2017; Vangrieken, Meredith, Parker, & Kyndt, 2017) als auch auf die Leistungen von Schülerinnen und Schülern (Reeves, Pun, & Chung, 2017) und auf ein bedürfnisorientiertes Schulklima (Westheimer, 2008). Negative Effekte sind ein erhöhter Workload für die Lehrkräfte und eine wahrgenommene Autonomieeinschränkung (Vangrieken, Dochy, Raes, & Kyndt, 2013). Studien zu Effekten professioneller Zusammenschlüsse auf Qualität, Frequenz und Bedeutung pädagogischer Diagnostik liegen dagegen kaum vor (Brinkmann & Twiford, 2012).

Ein weiterer wichtiger Befund für die Integration interaktiver Bestandteile in Lerngelegenheiten zur Entwicklung diagnostischer Kompetenz stammt aus Studien zum Umgang von Lehrkräften mit professionellen Zusammenschlüssen während des Berufseinstiegs. Sie weisen darauf hin, dass die hohen Anforderungen an die kritische und kommunikative Auseinandersetzung mit etablierten gruppenspezifischen Normen und Routinen in den Schulen beim Berufseinstieg häufig eine retrospektive Nivellierung universitär erworbener Wissensbestände bewirkt (Ghousseini & Herbst,

2016; Hoth et al., 2016). Auch Main (2007) identifiziert die Schulung professionellen Zusammenarbeitens als Voraussetzung gelingender Zusammenschlüsse pädagogischen Fachpersonals. Da die bisherigen Modelle diagnostischer Kompetenz überwiegend individuelles Handeln fokussieren, könnte der vorübergehende Bedeutungsverlust erworbener Wissensbestände auch ein Hinweis einen Bedarf an stärkerer Implementierung interaktiver diagnostischer Handlungsbereiche in der Lehrkräftebildung sein. Bislang liegen dazu keine systematisch aufbereiteten Daten vor (z. B. Analyse von Curriculae).

Wenngleich die Entwicklung individueller Dispositionen eine zentrale Facette der Entwicklung diagnostischer Kompetenz bleibt, sollte die interaktive Dimension diagnostischer Kompetenz stärker in der Gestaltung von Interventionen bzw. Lerngelegenheiten aufgegriffen und messbar gemacht werden.

Förderung diagnostischer Kompetenz mit problemorientiertem Lernen an Textfällen

Um Lehrkräften in der ersten Phase ihrer Ausbildung praxisbezogene Wissensbestände und Fähigkeiten zu vermitteln, werden zunehmend Methoden der Fallarbeit verwendet. Als Fälle dienen sowohl Videos als auch Texte oder Erzählungen über retrospektiv erinnerte professionsrelevante Situationen. Didaktisch aufbereitete Fallarbeit eröffnet den Lernenden eine authentische, aber komplexitätsreduzierte Auseinandersetzung mit Unterrichtssituationen und zielt auf die Förderung von Wahrnehmungs-, Interpretations- und Entscheidungsfähigkeiten ab. Die Entlastung vom Handlungsdruck der Realität soll Kapazitäten für die systematische Reflexion und Handlungsplanung freisetzen sowie die Entwicklung problembezogener Interessen und Lernziele stimulieren (Reh & Schnelle, 2010).

Bisherige Studien zur Wirksamkeit von Fallarbeit in der Entwicklung diagnostischer Kompetenz konzipieren fallbasiertes Lernen in Anknüpfung an kognitive Ansätze der Kompetenzentwicklung als individuellen Prozess, etwa mit persönlichen Praxisreflexionen in Tagebüchern (Klug, 2011) und Arbeitsaufgaben (Trittel, Gerich, & Schmitz, 2014) sowie mit computerbasierten Trainings im virtuellen Klassenraum (Praetorius et al., 2016) oder mit angeleitetem Tutoring mit videobasierten Fällen (Seidel, Blomberg, & Stürmer, 2010). Videobasierte Fälle zeigten u. a. positive Effekte auf die Wahrnehmung diagnostisch relevanter Unterrichtssituationen (Krammer, 2014; Santagata & Guarino, 2011). Kognitive Elaborationen wie etwa die differenzierte Beschreibung und Erklärung von Videoszenen verändern sich durch das Lernen mit videobasierten Fällen dagegen abhängig von den studierten und in den Videos gezeigten Unterrichtsfächern sowie abhängig von der vorangegangenen Erfahrung mit Videos aus dem eigenen Unterricht (Seidel & Schäfer, 2015; Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011). Insbesondere in der ersten Phase der Lehrkräftebildung, in der Erfahrungen mit eigenem Unterricht und Videoanalyse nicht vorausgesetzt werden können, ist textbasierte Fallarbeit eine sinnvolle didaktische Alternative für die Kompetenzentwicklung. Textbasierte Fallarbeit und deren Effekt auf die Kompetenzentwicklung angehender Lehrkräfte sind im Gegensatz zu videobasierter Fallarbeit weit weniger gut untersucht (Syring et al., 2015).

Didaktische Konzepte und Studien zu kollektiver Fallarbeit mit textbasierten Fällen liegen bislang vor allem in der Medizin vor. Hier hat sich insbesondere problemorientiertes Lernen (POL) als wirksame didaktische Methode der textbasierten Fallarbeit bewährt. POL ist eine konstruktivistische, lernerinnen- und lernerzentrierte Unterrichtsmethode, die authentische und alltagsnahe Problemsituationen für selbstgesteuertes Lernen in Kleingruppen aufbereitet (Barrows, 1985). Die wichtigsten Elemente von POL sind die gemeinsame Analyse eines Falls, die Diskussion bereits vorhandener Wissensbestände und die eigenständige Entscheidung für weiterführende Lernziele. Bekannt sind positive Effekte problemorientierten Lernens mit Textfällen auf den Aufbau epistemologischer Überzeugungen zur aktiven Wissenskonstruktion (Kuhn, 2007) und den langfristigen Erhalt erworbener Wissensbestände (Wirkala & Kuhn, 2011). Meta-Analysen zum Wissenserwerb zeigen uneindeutige Ergebnisse zum Erwerb deklarativen Wissens, jedoch eindeutig positive Effekte auf den Erwerb prozeduralen und anwendungsorientierten Wissens sowie auf die Argumentationsfähigkeiten (Dochy, Segers, van den Bossche, & Gijbels, 2003; Strobel & van Barneveld 2009; Bergstrom, Pughs, Phillips, & Machlev, 2016). Implementationen von POL führen überwiegend zu einem höheren Anstieg der Lernmotivation als lehrkräftezentrierte Lerngelegenheiten, u.a. durch eine höhere Nutzenerwartung (Sungur & Tekkaya, 2006; Hwang & Kim, 2006; Sangestani & Khatiban, 2013). Wichtige Moderatoren für den Motivationseffekt sind die Qualität der Fälle (Noordzij & Te Lindert, 2010), die Unterstützung bzw. Moderation des Lernprozesses (Wijnia, Loyens, & Derous, 2011) und die curriculare Kombination mit anderen Lernmethoden (Wijnen et al., 2017). Für den Bereich der Lehrkräftebildung mangelt es im Vergleich zur Medizin an empirischen Untersuchungen über POL (Walker & Leary, 2009).

Um weitere Erkenntnisse zu interaktiven und textbasierten Lernformaten sowie zur Entwicklung diagnostischer Kompetenz zu erlangen, wurde für die vorliegende Studie eine Intervention zum Erwerb diagnostischer Kompetenz auf der Grundlage problemorientierten Lernens mit Textfällen entwickelt. Da bisher keine verlässlichen Instrumente zur Kompetenzmessung in stärker interaktiven Bereichen pädagogischer Diagnostik existieren, greift die vorliegende Studie ebenfalls auf Methoden individualisierter Kompetenzmessung zurück. Diese Einschränkung in der Messung von Interventionseffekten gewährt jedoch einerseits Vergleichbarkeit mit o.g. Studien und andererseits kann dadurch die Frage beantwortet werden, ob sich eine stärker interaktiv ausgerichtete Kompetenzentwicklung positiv auf die individuelle Kompetenz auswirkt.

7.2 Fragestellung und Hypothesen

Bisherige Interventionsstudien zur Förderung diagnostischer Kompetenz planen den Kompetenzerwerb trotz großer interaktiver Anteile beim Diagnostizieren als individuellen Prozess. Es fehlt an Erkenntnissen über kollektive Formen der Kompetenzentwicklung, wie auch an empirischer Evidenz über die Eignung textbasierter Fallarbeit für den Erwerb diagnostischer Kompetenz. Ziel der vorliegenden Studie ist es, diese Forschungslücke mit Daten zu schließen, die den zentralen Dispositionen des COACTIV-Modells für Lehrkräfteprofessionalität entsprechen. Untersucht werden das Diagnosewissen, die Motivation für pädagogische Diagnostik und das diagnostische Selbstkonzept. Während das COACTIV-Modell in Anlehnung an das Modell der Urteilsgenauigkeit sowohl

fachdidaktisches als auch der pädagogisch-psychologisches Wissen vorsieht, beschränkt sich das vorliegende Modell diagnostischer Kompetenz auf pädagogisch-psychologisches Wissen zu Diagnostik und überfachlichen Lernvoraussetzungen (z. B. Motivation, Prüfungsangst). Entsprechend überprüft die vorliegende Studie folgende Hypothesen: Problemorientiertes Lernen mit Textfällen (POL-TF) hat einen positiven Effekt auf das Wissen über pädagogische Diagnostik (Hypothese 1), auf die Motivation zur Beschäftigung mit pädagogischer Diagnostik (Hypothese 2) und auf das Selbstkonzept zu pädagogischer Diagnostik (Hypothese 3).

7.3 Methoden und Design

Studiendesign und Intervention

Die vorliegende Studie ist eine quasi-experimentelle Interventionsstudie mit zwei Messzeitpunkten. Zum ersten und zweiten Messzeitpunkt wurden jeweils ein Wissenstest zur Erfassung des pädagogischen Diagnosewissens und ein Fragebogen zur Erfassung der Motivation und des diagnostischen Selbstkonzepts eingesetzt. Vor Beginn der Intervention fanden in beiden Versuchsgruppen fünf identische Lerneinheiten zu den Grundlagen pädagogischer Diagnostik statt.

Interventionsgruppe. Nach diesen fünf Lerneinheiten wurde den Studierenden zunächst POL-TF als didaktische Methode vorgestellt. Nach einem theoretischen Input über Forschungsstand und Konzept von POL-TF stellte der Dozent ad-hoc eine Kleingruppe zusammen und führte mit dieser im Fishbowl-Design exemplarisch die vorgesehenen POL-TF Schritte durch. Beobachtungen und Erfahrungen der Studierenden wurden in der zweiten Lerneinheit reflektiert und zu erwartetem Unterstützungsbedarf in der Fallarbeit verdichtet. Anschließend bildeten die Studierenden Kleingruppen von 5–7 Personen und bearbeiteten die theoretisch konstruierten Textfälle in zwei Phasen, die den Schritten pädagogischer Diagnostik im Kompetenzmodell von Klug et al. (2016) entsprechen. Die erste Phase bestand aus der Themen- und Informationssammlung, der Auswahl und der Aktivierung relevanter Wissensbausteine und der gemeinsamen Festlegung von Lernzielen, z. B. die Entstehung von Prüfungsangst erklären können. Im Abstand von einer Woche bearbeiteten alle Gruppenmitglieder jeweils alle Lernziele, diskutierten und visualisierten anschließend das erworbene Wissen, planten das weitere Vorgehen im Fallbeispiel, formulierten eine diagnostische Arbeitshypothese und reflektierten abschließend ihren Arbeitsprozess. Eine Person aus der Gruppe moderierte diesen Prozess. Die Gruppen blieben in ihrer Zusammensetzung für die Dauer der Intervention gleich. Dies entspricht den Kriterien professioneller Zusammenschlüsse von Vangrieken et al. (2015) und ermöglichte den Kleingruppen die notwendige Identifikations- und Kohäsionsbildung, wie auch relevante Ziele und Werte über die Formulierung von Lernzielen zu finden. Aus praktischer Sicht erleichtern konstante Gruppen die Planungs-, Dokumentations- und Austauschprozesse für die Studierenden. Nach Bearbeitung des letzten Textfalls fand eine gemeinsame Lerneinheit zur Reflexion und Ergebnissicherung statt.

Die Implementation von POL-TF erfolgte schrittweise nach dem 4C-ID Modell (Four Component Instructional Design) von van Merriënboer und Kirschner (2007), das auf den Aufbau professioneller

Routinen und Schemata zielt. Die Studierenden erhielten zu Beginn zusätzlich zum Fall Strukturierungshilfen (z. B. ein Arbeitsblatt zu POL-TF Schritten), anschließend wiederkehrende Übungsaufgaben (z. B. Hypothesen- und Feedbackformulierung) und unterstützende Informationen in Form von Hinweisen zu Literaturdatenbanken und Suchwörtern sowie Literatur über Lernvoraussetzungen und -störungen. Die Literatur ermöglichte einen Einstieg in mögliche Themenbereiche der Textfälle und wurde durch die Studierenden selbstständig erweitert.

Die sechs verwendeten und durch Expertinnen und Experten validierten Textfälle skizzieren auffällige Unterrichtssituationen mit Bezug zu individuellen stabilen Lernvoraussetzungen und -störungen. Informationen werden sowohl im Fließtext als auch in Zusatzmaterialien präsentiert, z. B. in Notenlisten, ausgefüllten Beobachtungsbögen und Tests. Ausgewählt wurden Lernstörungen mit den höchsten Prävalenzraten unter Schülerinnen und Schülern, z. B. Prüfungsangst, Aufmerksamkeitsstörung oder Lese-Rechtschreibschwäche (Linderkamp & Grünke, 2007). Die Schwierigkeit der Textfälle wurde über die Qualität und Quantität der Informationen (z. B. Kongruenz, Mehrdeutigkeit, Distraktoren) und den Grad der Lösungsoffenheit der Fälle bestimmt und im Verlauf der Intervention gesteigert.

Kontrollgruppe. Die Studierenden der Kontrollgruppe besuchten ein thematisch identisches Seminar mit theoretischen Inputs und instruierter Kleingruppen- und Einzelarbeit, z. B. Feedback für einen fiktiven Schüler formulieren. Um die Vergleichbarkeit der Interventions- und Kontrollgruppe zu gewährleisten, wurden beide Seminare durch denselben Dozenten angeleitet. Zusätzlich wurde den Studierenden zu Beginn und nach der Hälfte des Seminars dieselben Lernziele mitgeteilt.

Stichprobe

Die Stichprobe bestand aus Lehramtsstudierenden der beruflichen Fachrichtungen und der Arbeitslehre, die im Wintersemester 2016/2017 an zwei parallelen Seminaren ‚Methoden pädagogisch-psychologischer Diagnostik‘ teilnahmen. Zum ersten Messzeitpunkt haben 57 Studierende ($N_{\text{Interventionsgruppe}} = 19$, $N_{\text{Kontrollgruppe}} = 38$) und zum zweiten Messzeitpunkt 43 Studierende ($N_{\text{Interventionsgruppe}} = 18$, $N_{\text{Kontrollgruppe}} = 25$) an der Befragung teilgenommen. In der Interventionsgruppe waren die Studierenden 23–44 Jahre ($M = 29.5$ Jahre, $SD = 5.1$ Jahre, 55 % weiblich) und in der Kontrollgruppe 21–39 Jahre alt ($M = 26.4$ Jahre, $SD = 4.6$ Jahre, 78 % weiblich). Interventions- und Kontrollgruppe unterschieden sich signifikant in Bezug auf das Alter ($U = 131.500$, $z = -2.264$, $p = .024$), nicht aber in Bezug auf das Geschlecht ($\chi^2 = 3.390$, $p = .066$). Beide Gruppen wurden im ersten Fachsemester des Masterstudiums begleitet. Die Teilnehmenden in der Interventionsgruppe studierten im Erst- und Zweitfach verschiedene berufliche Fachrichtungen wie Ernährung/Lebensmittelwissenschaft oder Mathematik. Die Teilnehmenden der Kontrollgruppe studierten das Erstfach Arbeitslehre (allgemeinbildendes Schulfach Wirtschaft – Arbeit – Technik) mit verschiedenen Zweitfächern wie Mathematik oder Deutsch als Fremdsprache. In Bezug auf die bildungswissenschaftlichen Studieninhalte gab es keine Unterschiede zwischen den Gruppen.

Messinstrumente

Das Wissen über pädagogische Diagnostik wurde mit dem PDW, dem Test zum pädagogischen Diagnosewissen, von Wedel und Müller (2016) erfasst. Der Test beinhaltet auf einer ersten Achse entsprechend des Modells diagnostischer Kompetenz von Klug (2011) fünf Inhaltsbereiche: Urteilsfehler, Feedback, Ablauf pädagogischer Diagnostik, individuelle Lernvoraussetzungen und Testgütekriterien. In Anlehnung an die Taxonomie kognitiver Lernziele von Bloom (1976) unterscheidet der Test auf einer zweiten Achse für jeden dieser Inhaltsbereiche die kognitiven Fähigkeiten Erkennen, Verstehen und Anwenden. Das Erkennen und Verstehen wird mit insgesamt 20 Multiple-Choice-Items (je vier Items pro Inhaltsbereich) und das Anwenden mit insgesamt 10 offenen Items (je 2 Items pro Inhaltsbereich) erfasst, z. B. „Sie korrigieren einen schriftlichen Test Ihrer Klasse. Die ersten zehn Tests fallen durchgehend schlecht aus. Worauf achten Sie bei der weiteren Korrektur besonders?“ oder „Ein Schüler zweifelt sehr an sich und hat Angst vor Prüfungen. In der letzten Prüfung weinte er nach kurzer Zeit. Sie mussten die Arbeit auch mit einer schlechten Note bewerten. Welches Feedback geben Sie ihm? Bitte formulieren Sie einen Satz.“ Das erste Beispielitem erfasst das Verstehen im Inhaltsbereich Urteilsfehler (z. B. Kontrasteffekt) und das zweite Beispielitem das Anwenden im Inhaltsbereich Feedback. Die Antworten wurden anhand eines Kodierleitfadens entweder als $0 = falsch$ oder als $1 = richtig$ kodiert. Die Inter-Rater-Reliabilität für die offenen Fragen lag bei $\kappa = .92$. Abweichungen wurden in einer Teamsitzung besprochen und rekodiert. Die Schwierigkeiten bzw. die Lösungswahrscheinlichkeiten der Items variierten zwischen 0.02 und 0.91 (siehe Tabelle 1). Es liegen zwei als Paralleltests konzipierte Versionen des Tests vor.

Die *Motivation* wurde in Anlehnung an die Selbsteinschätzungsskala von Fernet et al. (2008) gemessen. Die Skala erfasst mit 3 Items auf einer 4-stufigen Antwortskala von $1 = trifft\ gar\ nicht\ zu$ bis $4 = trifft\ genau\ zu$, wie intrinsisch motiviert die Studierenden sind, sich mit pädagogischer Diagnostik zu beschäftigen, z. B. „Ich beschäftige mich mit pädagogischer Diagnostik, weil mir das Spaß macht.“ Die Skala ist zufriedenstellend reliabel, $\alpha_{t1} = .86$, $\alpha_{t2} = .78$.

Das *Selbstkonzept* wurde in Anlehnung an Helmke (2009) und Dickhäuser (2002) mit einer selbstentwickelten Skala gemessen. Die Skala erfasst mit 15 Items auf einer 4-stufigen Antwortskala von $1 = trifft\ gar\ nicht\ zu$ bis $4 = trifft\ genau\ zu$, wie die Studierenden ihre Fähigkeiten in pädagogischer Diagnostik einschätzen, z. B. „Ich kann Instrumente zur Erfassung von Lernvoraussetzungen einsetzen (z. B. Beobachtungsbögen, Tests).“ Die Skala ist zufriedenstellend reliabel, $\alpha_{t1} = .91$, $\alpha_{t2} = .89$.

Tabelle 1

Itemschwierigkeiten des PDW

	Messzeitpunkt 1	Messzeitpunkt 2
<i>Inhaltsbereiche</i>		
Urteilsfehler	0.13–0.78	0.09–0.64
Feedback	0.19–0.61	0.22–0.66
Ablauf pädagogischer Diagnostik	0.04–0.91	0.02–0.68
Lernvoraussetzungen	0.02–0.56	0.07–0.64
Testgütekriterien	0.02–0.56	0.04–0.73
<i>Kognitive Fähigkeiten</i>		
Erkennen	0.04–0.56	0.02–0.73
Verstehen	0.02–0.91	0.07–0.66
Anwenden	0.02–0.61	0.04–0.52

Anmerkungen. Die fünf Inhaltsbereiche beinhalten jeweils 6 Items und die drei kognitiven Fähigkeitsstufen jeweils 10 Items. Die angegebenen Itemschwierigkeiten sind die jeweils niedrigsten und höchsten Itemschwierigkeiten des Inhaltsbereichs bzw. der kognitiven Fähigkeitsstufe. Beispielsweise hat das schwierigste der sechs Items des Inhaltsbereichs Urteilsfehler eine Itemschwierigkeit von 0.13 und das leichteste Item eine Itemschwierigkeit von 0.78.

7.4 Ergebnisse

Die deskriptiven Ergebnisse für Wissen, Motivation und Selbstkonzept sowie die Mann-Whitney-*U*-Tests der Unterschiede zwischen Interventions- und Kontrollgruppe sind in Tabelle 2 dargestellt.

Die Hypothesen wurden mit einfaktoriellen ANOVAs mit Messwiederholung mit dem Zwischensubjektfaktor Interventions- oder Kontrollgruppe und mit den Kovariaten Alter und Geschlecht geprüft (siehe Tabelle 3). Shapiro-Wilk-Tests für kleine Stichproben zeigen, dass mit Ausnahme des anwendungsbezogenen Wissens die abhängigen Variablen bzw. die Residuen der abhängigen Variablen in der Interventions- und der Kontrollgruppe normalverteilt sind. Allerdings ist die ANOVA relativ robust gegenüber Verletzungen der Normalverteilungsannahme (Pagano, 2010; Wilcox, 2012).

Tabelle 2

Mittelwerte, Standardabweichungen und Mann-Whitney-U-Tests für Wissen, Motivation und Fähigkeitsselbsteinschätzungen in der Interventions- und Kontrollgruppe

	Messzeitpunkt	Interventionsgruppe	Kontrollgruppe	U-Test	
		<i>M (SD)</i>	<i>M (SD)</i>	<i>Z</i>	<i>p</i>
Wissen (Gesamt)	1	10.69 (2.92)	7.94 (3.71)	-2.394	.017
	2	13.17 (4.19)	8.31 (3.58)	-3.458	<.001
Erkennen	1	4.44 (1.71)	2.77 (1.55)	-2.865	.004
	2	5.17 (1.79)	3.92 (1.85)	-2.078	.038
Verstehen	1	3.44 (1.50)	3.59 (1.52)	-0.270	.787
	2	4.50 (1.43)	2.77 (1.42)	-3.467	<.001
Anwenden	1	2.81 (1.72)	1.87 (1.59)	-1.931	.054
	2	3.50 (1.92)	1.62 (1.27)	-3.038	<.001
Motivation	1	2.71 (0.40)	2.62 (0.85)	-.056	.955
	2	3.15 (0.42)	2.97 (0.58)	-1.267	.205
Selbstkonzept	1	2.40 (0.42)	2.70 (0.44)	-1.848	.065
	2	3.03 (0.34)	2.91 (0.38)	-1.333	.183

Anmerkungen. Wissen (Gesamt) 0–30 Punkte, Erkennen, Verstehen und Anwenden 0–10 Punkte, Motivation 1–4, Selbstkonzept 1–4; $29 \leq N \leq 30$.

Tabelle 3

ANOVA mit Messwiederholung für Wissen, Motivation und Selbstkonzept

	Messzeitpunkt			Messzeitpunkt x Gruppe		
	<i>F</i>	<i>p</i>	η^2_p	<i>F</i>	<i>p</i>	η^2_p
Wissen (Gesamt)	4.554	.042	.149	2.831	.104	.098
Erkennen	3.124	.089	.111	1.086	.307	.042
Verstehen	2.501	.126	.088	5.204	.031	.167
Anwenden	0.966	.335	.037	4.227	.050	.145
Motivation	0.084	.774	.003	2.297	.142	.081
Fähigkeiten	26.651	<.001	.506	6.894	.014	.210

Anmerkungen. Kovariaten Alter und Geschlecht; $29 \leq N \leq 30$.

Wissen zu pädagogischer Diagnostik

Zum ersten Messzeitpunkt unterschieden sich die Interventions- und Kontrollgruppe in Bezug auf ihr Wissen über pädagogische Diagnostik. Dieser Unterschied zeigte sich beim Erkennen, nicht aber beim Verstehen und Anwenden (siehe Tabelle 2). Zum zweiten Messzeitpunkt erreichte die Interventionsgruppe im Erkennen, Verstehen und Anwenden signifikant höhere Werte als die Kontrollgruppe (siehe Tabelle 2).

Die ANOVA mit Messwiederholung zeigt einen starken Haupteffekt für das Wissen über pädagogische Diagnostik, das über das Semester sowohl in der Interventionsgruppe als auch in der Kontrollgruppe zunimmt. Auf Ebene der kognitiven Fähigkeiten, d. h. Erkennen, Verstehen und Anwenden, zeigt sich kein signifikanter Haupteffekt des Messzeitpunktes (siehe Tabelle 3).

Die ANOVA zeigt zudem einen nicht signifikanten, schwachen Interventionseffekt für das Erkennen und signifikante, starke Interventionseffekte für das Verstehen und Anwenden (siehe Tabelle 3). Die Interventionseffekte für das Erkennen, Verstehen und Anwenden sind in Abbildung 10 veranschaulicht. Die Interventionsgruppe ist immer fett, die Kontrollgruppe nicht fett dargestellt. Während das Erkennen vom ersten zum zweiten Messzeitpunkt in beiden Gruppen zunimmt (siehe durchgezogene Linien), nimmt das Verstehen und Anwenden nur in der Interventionsgruppe zu (siehe fette gestrichelte und fette gepunktete Linien). In der Kontrollgruppe zeigt sich beim Verstehen und Anwenden eine Abnahme (siehe nicht fette gestrichelte und nicht fette gepunktete Linie).

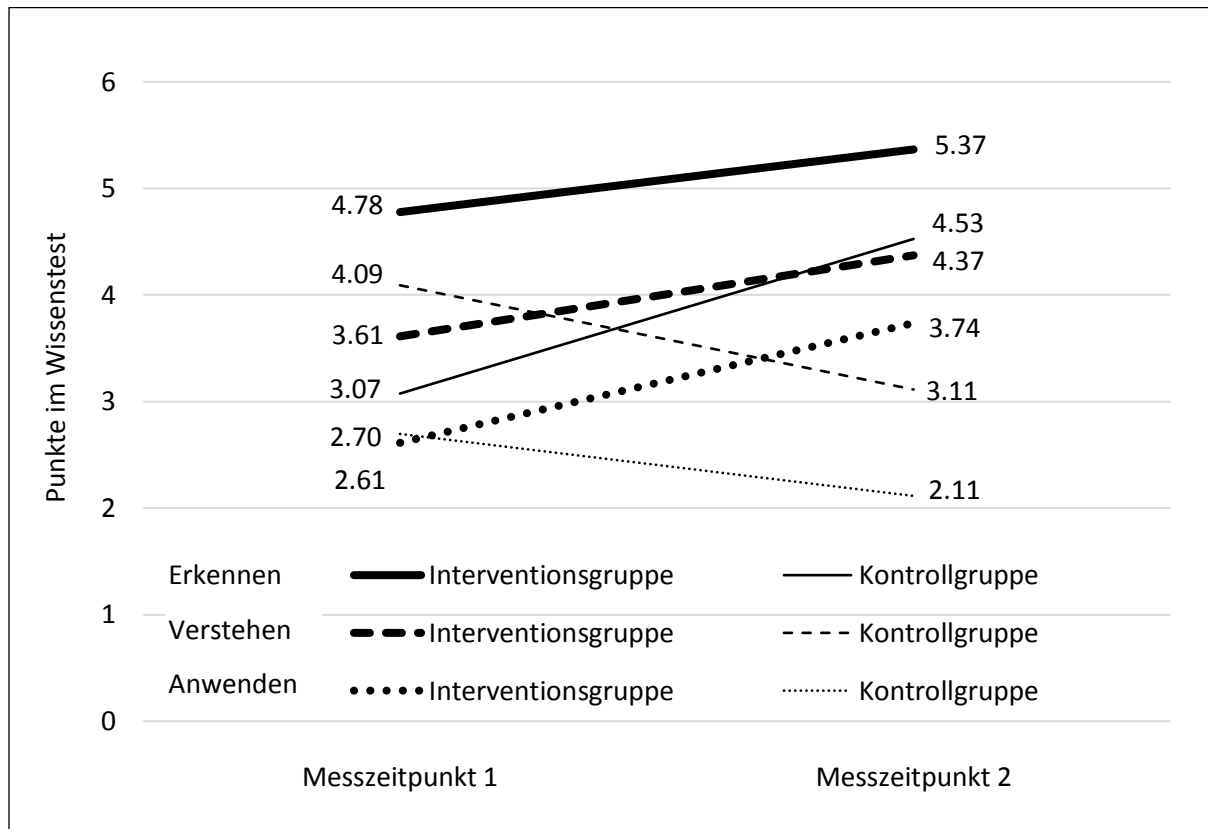


Abbildung 10. Effekt von Messzeitpunkt x Gruppe auf das Erkennen, Verstehen und Anwenden, Range 0–10.

Motivation zur Beschäftigung mit pädagogischer Diagnostik

Zum ersten Messzeitpunkt unterschieden sich Interventions- und Kontrollgruppe in Bezug auf die Motivation zur Beschäftigung mit pädagogischer Diagnostik nicht signifikant voneinander (siehe Tabelle 2). Auch zum zweiten Messzeitpunkt zeigte sich kein signifikanter Unterschied, da beide Gruppen zum Ende des Semesters motivierter waren als zu Beginn (siehe Tabelle 2).

Die ANOVA mit Messwiederholung zeigt keinen signifikanten Effekt des Messzeitpunkts und einen nicht signifikanten, mittleren Interventionseffekt (siehe Abbildung 11 und Tabelle 3).

Selbstkonzept zu pädagogischer Diagnostik

Zum ersten Messzeitpunkt unterschieden sich die Interventions- und Kontrollgruppe in Bezug auf das Selbstkonzept zu pädagogischer Diagnostik nicht signifikant voneinander (siehe Tabelle 2). Auch zum zweiten Messzeitpunkt zeigte sich kein Unterschied (siehe Tabelle 2).

Die ANOVA mit Messwiederholung zeigt einen signifikanten, sehr starken Effekt des Messzeitpunktes und einen signifikanten, sehr starken Interventionseffekt (siehe Tabelle 3). Das Selbstkonzept nahm vom ersten zum zweiten Messzeitpunkt in beiden Gruppen, besonders stark aber in der Interventionsgruppe zu (siehe Abbildung 12).

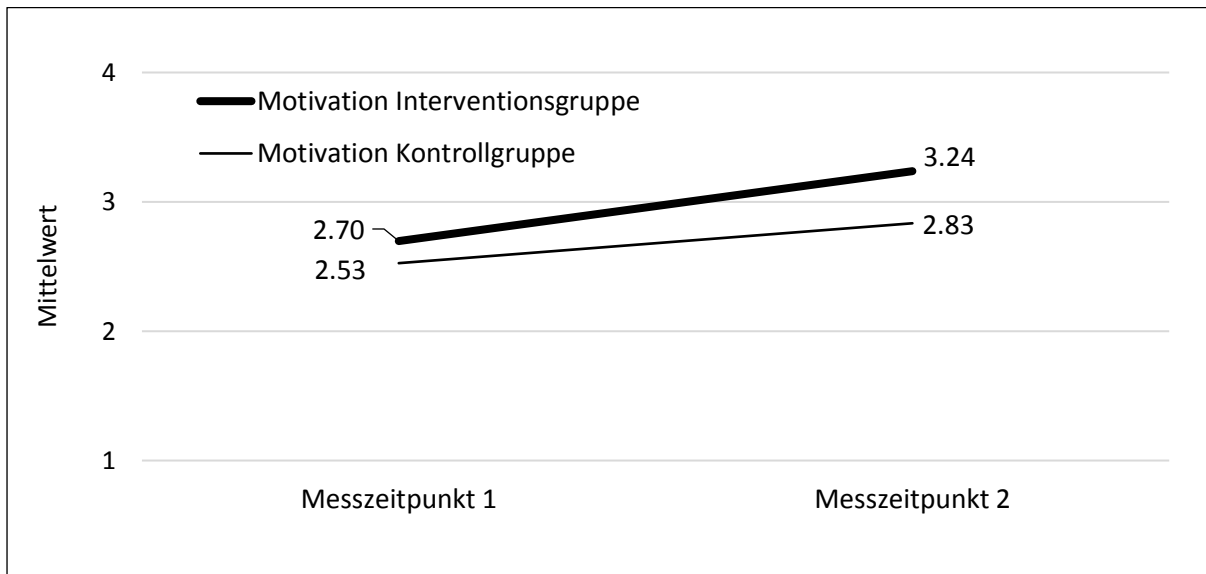


Abbildung 11. Effekt von Messzeitpunkt x Gruppe auf die Motivation.

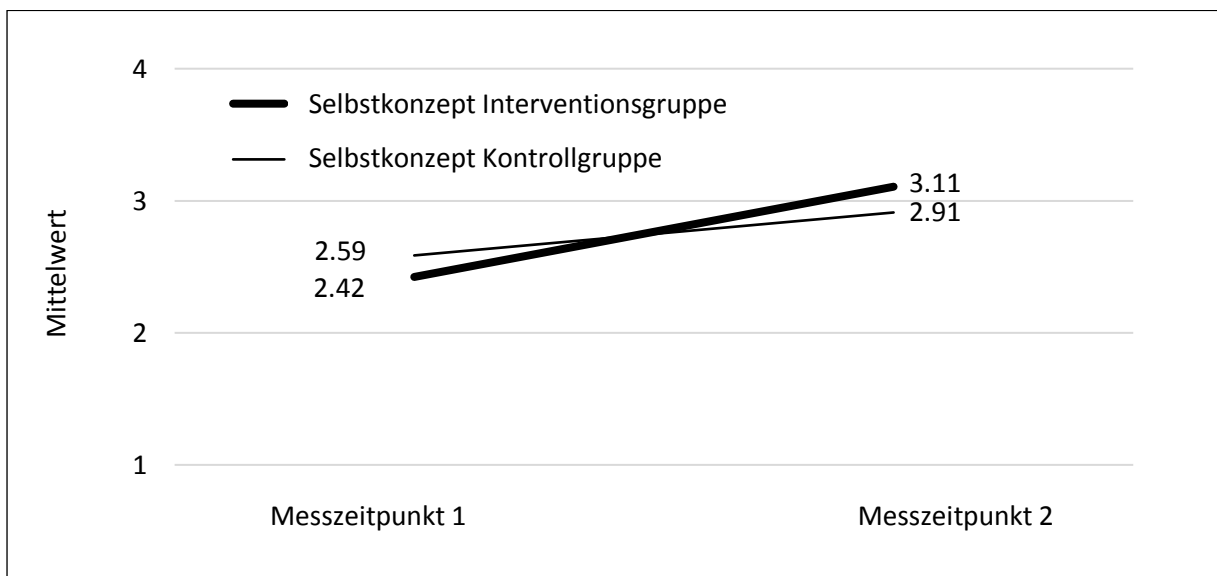


Abbildung 12. Effekt von Messzeitpunkt x Gruppe auf das Selbstkonzept.

7.5 Diskussion

Die vorliegende Studie untersuchte die Effekte problemorientierten Lernens mit textbasierten Fällen (POL-TF) auf die Entwicklung diagnostischer Kompetenz von Lehramtsstudierenden. Die Kompetenzmessung erfolgte in Anlehnung an das COACTIV-Modell professioneller Lehrkompetenz. Angenommen wurde, dass POL-TF einen positiven Effekt auf das Wissen, die Motivation und das Selbstkonzept in Bezug auf pädagogische Diagnostik hat.

Wissen, Motivation und Selbstkonzept

POL-TF fördert pädagogisches Diagnosewissen. Während in beiden Untersuchungsgruppen ein Wissenszuwachs zu verzeichnen war, stiegen in der Interventionsgruppe das Verstehen und Anwenden signifikant stärker als in der Kontrollgruppe. Die Arbeit mit Textfällen stößt offenbar tiefere Verarbeitungsprozesse an und fokussiert besonders Fähigkeiten der kognitiven Elaboration (wie Verstehen und Anwenden). Sie stellt damit eine gute Ergänzung zur Förderung der professionellen Unterrichtswahrnehmung mit Videofällen dar (Seidel & Schäfer, 2015). Dass die Fähigkeiten zur kognitiven Elaboration in der Kontrollgruppe eher absinken, könnte durch die unterschiedlichen Angebote zur Wissensorganisation in den Gruppen zu erklären sein. Die wiederkehrenden Schritte der Fallarbeit in der Interventionsgruppe führen möglicherweise zur Bildung inhaltlich flexibler kognitiver Routinen und Schemata, z. B. wie auf diagnostisch relevante Merkmale geschlossen wird. In der Kontrollgruppe könnte es zu einer ‚Verinselung‘ von Wissensbausteinen in Verbindung mit jeweils verschiedenen kognitiven Routinen und Schemata gekommen sein. Dies würde einen zusätzlichen Aufwand in der Informationsspeicherung und -abrufung bedeuten und könnte zu einer größeren Fehlerrate in der Anwendung führen. Entsprechend der ‚Cognitive Load Theory‘ (Sweller, 2003) wirkt sich eine klare kognitive Struktur positiv auf die Informationsverarbeitung und -speicherung aus. Die selbstständige und gemeinsame Festlegung von Lernzielen im Rahmen von POL-TF könnte dementsprechend die Anschlussfähigkeit der Lerninhalte an das Vorwissen und die flexible kognitive Strukturierung der neuen Wissensbausteine gefördert haben. Auch dass sich die Schritte der Intervention und der Ablauf pädagogischer Diagnostik stark ähneln, spricht für diese Interpretation. Die erste Hypothese zum positiven Effekt von POL-TF auf das pädagogische Diagnosewissen wird demnach also in dieser Studie – nicht zum Erkennen, wohl aber zum Verstehen und Anwenden – bestätigt.

POL-TF und instruktionsorientiertes Seminar fördern Motivation. Überraschenderweise unterschied sich die Veränderung der Motivation nicht signifikant zwischen der Interventions- und der Kontrollgruppe. Tatsächlich nahm die Motivation für pädagogische Diagnostik in beiden Gruppen zu. Dieser durchaus erfreuliche Befund wurde aber nicht zusätzlich ergänzt durch einen besonderen Interventionseffekt. Eine mögliche Erklärung dafür ist die fehlende Differenzierung des Motivationsmaßes für die Fallarbeit und die besonderen Effekte selbstgesteuerter Lernprozesse. Da beide Gruppen durch denselben Dozenten angeleitet wurden, kann nicht ausgeschlossen werden, dass der Interventionseffekt durch den Effekt des Dozenten auf die Motivation der Teilnehmerinnen und Teilnehmer überschattet wurde. In zukünftigen größer angelegten Studien sollte dieser Effekt kontrolliert werden, z. B. durch systematische Variation der Dozierenden und die Erhebung motivationsförderlicher Eigenschaften der bzw. des Dozierenden und der Seminargestaltung. Darüber hinaus müsste für die über die Seminarzeit hinausgehende Lernzeit kontrolliert werden. Dazu gehören die Bearbeitung der Lernziele in der Interventionsgruppe und die Bearbeitung der Zusatzaufgaben in der Kontrollgruppe. Die zweite Hypothese zum Anstieg der Motivation kann demzufolge nur für die allgemeine Beschäftigung mit pädagogischer Diagnostik, nicht aber spezifisch für POL-TF bestätigt werden.

POL-TF fördert diagnostisches Selbstkonzept. Neben den kognitiven Wissensdimensionen zeigte sich auch ein Interventionseffekt für das Selbstkonzept. Die Werte nahmen sowohl in der Kontroll- als auch in der Interventionsgruppe zu, allerdings besonders stark in der Interventionsgruppe. Die textbasierte Arbeit mit Fällen trägt also auch zur differenzierten und positiven Einschätzung der eigenen Fähigkeiten im Bereich pädagogischer Diagnostik bei. Da soziale Vergleichsprozesse und Rückmeldungen eine grundlegende Informationsquelle für das Selbstkonzept sind (Dickhäuser, 2002), erklären der kontinuierliche gemeinsame Austausch und die Festlegung gemeinsamer Lernziele im Sinne einer Gruppennorm und eines gemeinsamen Erfolgserlebnisses möglicherweise diesen Anstieg. Die dritte Hypothese zum Anstieg des Selbstkonzeptes durch POL-TF kann demnach beibehalten werden. Dieses Ergebnis ist besonders erfreulich, da das Selbstkonzept ein vergleichsweise stabiles Konstrukt ist, dessen ‚baseline‘ auch beim Übergang in die Arbeitswelt erhalten bleibt (Cairns, McWhirter, Duffy, & Barry, 1990).

Interventionsgestaltung, Kompetenzentwicklung und Textfälle

Durch den Einsatz problemorientierten Lernens kann diagnostische Kompetenz in der ersten Phase des Lehramtsstudiums gefördert werden. Dass eine systematisch interaktiv organisierte Lerngelegenheit den individuellen Kompetenzerwerb befördert, spricht für einen stärkeren Einbezug von Kontextvariablen in Modelle der Kompetenzentwicklung. Denkbar sind etwa sozialpsychologische Konstrukte wie Gruppennormen, Rollenverteilungen und -konflikte, Einstellungen der Lernenden zueinander oder die Identifikation mit der sozialen Gruppe. Analysen zur Kompetenzentwicklung sollten entsprechend auch interpersonale Konstrukte wie Empathie und soziale Unterstützung stärker einbeziehen. In einer Interventionsstudie mit Unterrichtsvideos von Seidel et al. (2011) haben sich Lehrkräfte erst dann auf eine Unterrichtsanalyse eingelassen, nachdem sie ein gemeinsames Lernziel festgelegt hatten. Diese gemeinsamen Sozialisationsprozesse sind Teil der Entwicklung professioneller Kompetenzen und sollten in der Kompetenzmessung und -entwicklung berücksichtigt werden. In der vorliegenden Intervention haben sich Textfälle als geeignetes, anpassbares und vergleichsweise ökonomisches Fallmedium erwiesen. Insbesondere die diagnostischen Zusatzinformationen in Form von Test- und Beobachtungsbögen können einen Eindruck über die pädagogische Diagnostik im Schulkontext vermitteln, mit dem sich Lehrkräfte bereits in einem frühen Stadium ihrer Ausbildung ohne Risiken für sich und andere mit pädagogischer Diagnostik auseinandersetzen können. Sowohl die Lösungsoffenheit als auch die Mehrdeutigkeit der Informationen führten in unserer Erfahrung zu angeregten Diskussionen in den POL-Gruppen.

Grenzen der Studie

Grenzen der Studie betreffen zum einen die begrenzte Stichprobengröße und zum anderen das Fehlen einer weiteren Interventionsgruppe, die problemorientiertes Lernen mit videobasierten Fällen durchführt. Daran anschließend bleibt auch offen, welche Effekte ein kombiniertes Design aus video- und textbasierten Fällen auf die Wahrnehmungsfähigkeiten einerseits und die kognitiven

Elaborationen andererseits hat sowie welche individuellen Dispositionen und Gruppenmerkmale die Kompetenzentwicklung moderieren.

Zusätzlich war eine Randomisierung der Teilnehmenden zu Interventions- und Kontrollgruppe aus studienorganisatorischen Gründen nicht möglich. Die Seminare wurden für spezifische Studierendengruppen angeboten, die im Rahmen des Curriculums nicht frei wählbar waren. Empirisch zeigte sich jedoch eine gute Vergleichbarkeit der Gruppen und trotz der überschaubaren Stichprobengröße konnten starke Effekte nachgewiesen werden.

Die entwickelte Intervention konzentrierte sich insbesondere auf die Diagnostik individueller stabiler Lernvoraussetzungen bei Schülerinnen und Schülern. Wünschenswert wäre eine Ausweitung auf weitere Inhalte pädagogischer Diagnostik, bspw. die Feststellung und Beantragung von Förderschwerpunkten. Diese haben mit 6–7% aller Schülerinnen und Schüler in Deutschland eine hohe Prävalenz im Schulsystem (Malecki, 2014). Hier wäre auch ein interdisziplinäres Vorgehen mit den Studiengängen der Sonderpädagogik und Psychologie denkbar. Da sich ein Großteil bisheriger Erkenntnisse über diagnostische Kompetenz in der Lehrkräftebildung auf das Modell der Urteilsgenauigkeit im Rahmen fachdidaktischer Forschung bezieht, sind Vergleiche mit dem überfachlichen Modell der vorliegenden Studie nur eingeschränkt möglich.

In Anlehnung an die oben beschriebene Nivellierung universitär erworbener Wissensbestände ist weiterhin offen, inwiefern der durch POL-TF erworbene Zuwachs an Wissen, Motivation und Selbstkonzept resistent gegenüber dem sogenannten ‚Praxischock‘ oder dominanten schulischen Normen und Routinen ist. Dies müsste eine Langzeitstudie klären, auch um förderliche und hinderliche Faktoren für den Transfer erworbener Kompetenzen genauer benennen zu können. Im Rahmen der hier vorgestellten Studie sind daher weitere Messzeitpunkte geplant.

Ausblick

Die Förderung diagnostischer Kompetenz bei Lehramtsstudierenden steht vor der besonderen Herausforderung, ein komplexes Handeln in der Praxis vorbereiten zu müssen, das eine große Ambiguitätstoleranz und Orientierungssicherheit in professionellen Zusammenschlüssen im Schulkontext verlangt. In vielen diagnostisch relevanten Fällen sind konkurrierende Hypothesen denkbar und unterschiedliche Informationssammlungen notwendig, bevor ein (vorläufiges) diagnostisches Urteil gefällt werden kann. Gerade problemorientiertes Lernen in Kleingruppen bereitet unter komplexitätsreduzierten Bedingungen gut auf diese Anforderungen vor, da interaktiv Deutungen ausprobiert und argumentativ untermauert werden, alternative Sichtweisen diskutiert und gemeinsame Entscheidungen gefällt werden. Neben dem selbstgesteuerten Lernprozess ist besonders die soziale Dimension des problemorientierten Lernens hilfreich, um auf pädagogische Diagnostik in der Praxis vorzubereiten.

Mehrere theoretische Kompetenzmodelle greifen die interaktiven Bestandteile pädagogischen Diagnostizierens bereits auf, bislang jedoch lediglich als Rahmenbedingung für kompetentes diagnostisches Handeln, nicht als eigene Dimension des diagnostischen Handelns. Dass bislang entwickelte Lerngelegenheiten zu diagnostischer Kompetenz in der Lehrkräftebildung keine

systematische Implementation interaktiver Lern- und Kommunikationsprozesse vornehmen, erklärt sich möglicherweise aus der Uneinigkeit über die Qualität ‚guter‘ pädagogischer Diagnostik und zu messender Indikator- und Outcomevariablen (Altmann & Nückles, 2017). Ein Konstrukt, das die interaktive Dimension diagnostischer Kompetenz aufnimmt, ist das ‚diagnostic agreement‘, d. h. die Übereinstimmung von Expertinnen und Experten in ihrem diagnostischen Urteil. Empirische Studien aus klinischen Kontexten zeigen, dass ‚diagnostic agreement‘ einen positiven Effekt auf die Motivation von Patientinnen und Patienten sowie auf den Erfolg im diagnostisch-therapeutischen Prozess hat (Jensen-Doss & Weiss, 2008; Pogge et al., 2001). Es besteht jedoch Forschungsbedarf zu Voraussetzungen und Effekten des ‚diagnostic agreement‘ im Schulkontext, sowie unterschiedlichen Formen professioneller Zusammenarbeit im Kontext pädagogischer Diagnostik. Auch im Bereich der Konzeptualisierung diagnostischer Kompetenz und der Entwicklung valider Verfahren zur Kompetenzmessung braucht es weitere Forschung. Insgesamt besteht weiterhin Forschungsbedarf zur theoretischen Konzeptualisierung pädagogischer Diagnostik und der Entwicklung valider Verfahren zur Kompetenzmessung.

Weitere Interventionsstudien in der Lehrkräftebildungsforschung bieten speziell in der ersten Phase des Lehramtsstudiums das Potenzial, den Hochschulen evidenzbasierte Möglichkeiten zur Vorbereitung angehender Lehrkräfte auf die komplexen Praxisanforderungen in der Schule zu erschließen.

8 Study 2: Training Teachers' Diagnostic Competence with Problem-Based Learning: A Pilot and Replication Study²

8.1 Introduction and Current State of Research

The increasing diversity in schools poses a challenge for developing inclusive teaching competences in higher education, especially diagnostic competence (Guerra, Kupczynski, Mundy, & Garza, 2017; Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper, & Lautenbach, 2016). In the TALIS Teaching and Learning International Surveys (Organization for Economic Cooperation and Development [OECD], 2009, 2014), teachers ranked their need for professional development in teaching students with special needs above all other fields. At the same time, the impact rating of available professional training in this field was among the lowest (OECD, 2014). Thus, integrating suitable and evidence-based learning environments such as case-based learning in teacher education curricula becomes increasingly important (Darling-Hammond, 2017; Ricken, 2017).

Case-based interventions mostly follow cognitive learning approaches, focusing on students' individual learning, competences, and dispositions (Shavelson, 2013; Stahnke, Schueler, & Roesken-Winter, 2016). Knowledge on the effectiveness of interventions that follow a more constructivist approach and encourage learning in collaborative environments to foster pre-service teachers' diagnostic competence, is still lacking (Buholzer & Zulliger, 2013; Ghousseini & Herbst, 2016; Hoth et al., 2016; Santagata & Yeh, 2016; Schrader, 2017).

Since learning in collaborative environments strongly depends on the participants' specific characteristics and activities such as pre-knowledge and role-taking, replicating findings is an important step in evaluating the conducted research. While replication studies are accompanied by many difficulties (e.g., variability in study conditions, participants, and results), the benefits are highly valuable (e.g., validity and estimates of true effect sizes) (Open Science Collaboration, 2015). Especially in teacher education, it is important to implement trainings only after replication to reduce the risk of distributing potentially negative effects into the educational system (Makel & Plucker, 2014). To bring the discussion on teachers' diagnostic competence and collaboration together both theoretically and empirically, we designed a learning environment for teacher education, and conducted a pilot study and a subsequent exact replication study.

Collaboration in Teachers' Diagnostic Competence and its Training

Teachers' diagnostic competence (i.e., diagnostic activities and skills) are complex, manifold, and often highly dependent on specific school policies and practices (Santagata & Yeh, 2016). While in the scientific discourse no single definition of diagnostic competence has been agreed upon (Herppich et al., 2018; Praetorius et al., 2012), most definitions focus on the teacher-student relationship and on

² Diese empirische Studie ist 2019 erschienen und kann folgendermaßen zitiert werden: Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A. (2019). Training teachers' diagnostic competence with problem-based learning: A pilot and replication study. In: Teaching & Teacher Education, 86, 1-14. <https://www.doi.org/10.1016/j.tate.2019.102909>.

classroom interactions, as well as on teachers using, preparing, and evaluating instructional and diagnostic material. Definitions address the assessment and support of learning behavior (Klug, Bruder, Kelava, Spiel, & Schmitz, 2013), identification and modification of pre-conditions for learning as well as school-track selection (Ingenkamp & Lissmann, 2008), and assessment of content knowledge (Leuders, Dörfler, Leuders, & Philipp, 2018). Curiously, definitions of diagnostic competence address neither professional interactions nor collaborative activities sufficiently.

Vangrieken, Dochy, Raes, and Kyndt (2015) introduced a flexible and yet precise definition of professional collaboration in a systematic review about effects of different forms of teacher collaboration and communities in schools. Their study employed the psychological construct of entativity. Entativity describes the extent to which a group of people is perceived as a connected and coherent unity (Campbell, 1958; Lickel et al., 2000). Defining characteristics of entativity are: (1) shared goals and responsibilities, (2) task-specific cohesion, (3) identification with the group, (4) interdependency in fulfilling tasks, (5) and interdependency in reaching goals. Professional collaborations with these characteristics have a positive effect on the professional development of experienced teachers (Toom, Pietarinen, Soini, & Pyhältö, 2017; Vangrieken, Meredith, Packer, & Kyndt, 2017; Vescio, Ross, & Adams, 2008), on students' academic achievement (Reeves, Pun, & Chung, 2017; Warwas & Helm, 2018; Zito, 2011), on changes in instructional practice (Woodland, Lee, & Randall, 2013; Zito, 2011), and on a supportive school climate (Westheimer, 2008). Negative effects include a higher workload for teachers and a perceived reduction in autonomy (Vangrieken et al., 2017). The convincing concept offered by Vangrieken et al. (2015) provides the conceptual framework for professional interaction in the present studies and is leveraged using problem-based learning where collaborative activities play a central role.

Empirical studies point out several situations in which teachers make use of their professional diagnostic skills and knowledge in collaborative environments, such as parental counselling (Klug, Bruder, Keller, & Schmitz, 2012), general and special education teacher collaboration (Jurkowski & Müller, 2018), and case conferences in multi-professional teams (Hesjedal, Hetland, & Iversen, 2015). Accordingly, multi-professional teams are discussed as a quality criterion for teachers' diagnostics (Behrmann & Glogger-Frey, 2017). Studies on decision-making support this proposition. Not only did Gigerenzer, Hoffrage, and Kleinbölting (1991) show that an outsider's perspective reduces cognitive bias and overconfidence in one's own knowledge, but systematic reviews confirmed the overall reduction of diagnostic errors through professionals' case-specific collaboration (Graber et al., 2012; McDonald et al., 2013).

Working together with other professionals is a core element of the teaching profession. Berliner (1985) sees teachers as "[...] decision maker[s] who handle a complex set of interacting variables in a dynamic social environment" (p. 6). However, this social environment is not restricted to the classroom, but also entails professional interactions on an institutional level. While collaboration requires context specific skills to be successful (Arthaud, Aram, Breck, Doelling, & Bushrow, 2007; Brownell, Adams, Sindelar, Waldron, & Varover, 2006), collaborative diagnostic activities serve as an important interface between the educational and the medical system (e.g., allocation of resources for special needs students; Byrne, 2017). Nonetheless, Moldavsky, Pass, and Sayal (2014) found that

teachers prefer working with other teachers to get a second opinion, rather than with psychologists when working out strategies for diagnosing students with deficits in learning prerequisites.

There are continuous calls for including professional teacher-teacher collaboration in curricula of teacher education (Brownell, Ross, Colón, & McCallum, 2005; Harvey, Yssel, Bauserman, & Merbler, 2010; Shippen, Crites, Houchins, Ramsey, & Simon, 2005; Vangrieken et al., 2017). And yet, learning environments fostering pre-service teachers' diagnostic competence often focus on developing dispositions and situation-specific skills in the teacher-student relationship, e.g., by reflecting personal teaching experiences in diaries (Klug, 2011) or by doing frequent tasks (Trittel, Gerich, & Schmitz, 2014), just as in computer-based trainings in virtual classrooms (Praetorius et al., 2016) or tutoring with video-based cases (Seidel, Blomberg, & Stürmer, 2010).

The present studies addressed the outlined gap in the scientific discourse as well as in teacher education by introducing a learning environment for pre-service teachers grounded in problem-based learning (PBL) with text-based cases.

Training Diagnostic Competence with Problem-Based Learning and Text-Based Cases

The designed learning environment builds on models for diagnostic competence with a focus on goal-specific information processing (Dünnebier, Gräsel, & Krolak-Schwerdt, 2009; Herppich et al., 2018; Klug et al., 2012; Klug, Bruder, & Schmitz, 2016; van Ophuysen, 2006). Cross-sectional studies about differences in diagnostic information processing between expert and novice teachers provided the general learning goals for the designed learning environment. Experts identify relevant and irrelevant diagnostic information more successfully (Klug et al. 2016), make use of specific information rather than social categories when asked to make predictions (Krolak-Schwerdt, Böhmer, & Gräsel, 2012), and build more accurate inferences about student development (Santagata & Yeh, 2016). Pre-service teachers cannot be expected to perform on the same level as experts, but the learning environment should aim to promote the cognitive performance in information processing. The specific learning goals for pre-service teachers were derived from the diagnostic information processing model of Klug et al. (2016) that distinguishes three consecutive phases of diagnostic activities. The pre-actional phase includes the cognitive processes of goalsetting (supporting an individual student) and activating existing diagnostic as well as student-specific knowledge. The actional phase includes generating, gathering, and interpreting information based on hypotheses to build a plausible working hypothesis. The post-actional phase includes the resulting pedagogical action (e.g., redesigning instructional material). The proposed three phases were substantiated in a confirmatory factor analysis with 12 open-ended items in a scenario test (Klug et al., 2012). In line with the research on diagnostic information processing, teachers as well as in-service teachers achieved higher scores in the scenario test than pre-service teachers. However, Klug et al. (2016) primarily focused on individual cognitive and behavioral processes, but not on professional interactions. Professional interactions and collaboration are merely considered a supplementary resource for individual diagnostic action.

Case-based learning for diagnostic activities, especially PBL, has a long tradition in medical education (Johnson & Finucane, 2000), and is effective for developing diagnostic competence in this

context (Hmelo-Silver, 2004). PBL is a constructivist instructional design method that focusses on students' interaction and self-regulated learning in small groups by systematically working on authentic cases (Barrows, 1985).

Firstly, PBL has several characteristics that strongly resemble the diagnostic process modelled by Klug et al. (2016), and thus provides a promising instructional framework for training diagnostic competence. These are the step-by-step action, identification, and selection of relevant information, hypothesis and evidence-based reasoning, subsequent planning and reflecting on the process, and reiterating these steps. Interactive elements highlighted in teacher collaboration research are also explicitly integrated into PBL. These include the collaborative information gathering and problem definition, as well as updating results, peer-feedback, and making group supported decisions.

Secondly, PBL strongly encourages collaborative learning. In fact, analyzing the time spent on tasks in videos of PBL sessions in medical education, Ramaekers (2011) found that participants spent 96 % of the time on interactively discussing, questioning, and explaining parts of the case. Of that time, 60 % was spent on gathering and organizing information, 7 % on developing interpretations, 25 % on making decisions, judgements and justifications, and 8 % on coordinating the activities.

Thirdly, meta-analyses about knowledge acquisition with PBL showed positive effects on developing procedural and application-oriented knowledge, as well as positive effects on cognitive elaboration skills and learning motivation, but also mixed results for declarative knowledge gains (Bergstrom, Pugh, Phillips, & Machley, 2016; Dochy, Segers, van den Bossche, & Gijbels, 2003; Hwang & Kim, 2006; Strobel & van Barneveld, 2009; Sangestani & Khatiban, 2013; Sungur & Tekkaya, 2006). An important moderator for the motivational effects is the quality of the cases (Noordzij & Te Lindert, 2010). Since case design is a crucial and practical element of the learning environment, we developed and tested a framework for designing text-based diagnostic cases. We thereby follow Syring et al.'s (2015) suggestion to study the effects of learning with text-based cases on competence development, before discarding them straight away in favor of videos.

Designing Text-Based Diagnostic Cases

L. Shulman (1992) defines a case as being a tangible example of a larger theoretical concept, broader category, or class of events. This definition accommodates positive and negative examples as well as content related problems. In teacher education, working with video-based cases yielded positive effects on the perception of diagnostic cues in classroom situations (Krammer, 2014; Santagata & Guarino, 2011). Cognitive elaboration (i.e., theoretically enriching the description and explanation of perceptions), however, was only pronounced when students had previous experience with reflecting on videos from their own teaching or when the videos portrayed lessons from their main subject (Schäfer & Seidel, 2015; Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011). Thus, learning with text-based cases seems to be a reasonable alternative for pre-service teachers who do not always have experience in teaching and working with video analysis. How can cases be designed to specifically support cognitive elaboration about diagnostics?

Diagnostic activities are commonly employed when teachers observe a sudden drop in the academic achievement or psychological wellbeing of their students. Since the students themselves cannot always identify the causes of their problems and the resources to cope with them, teachers must entertain different possibilities in the diagnostic process until they can act with reasonable certainty. This situation bears a striking resemblance to an ill-structured problem (Jonasson, 2000; Jonasson & Hung, 2008). Ill-structured problems lack a clear problem definition and usually contain informational complexities like missing or distracting information. Thus, different perspectives on any given situation are possible and procedures for identifying and selecting problems are needed. This process particularly depends on individual values and beliefs (Grohs, Kirk, Soledad, & Knight, 2018), as well as on communicative interaction with others (Hmelo-Silver, 2004; Selmer, Jonasson, & Luring, 2013). Typical problem-solving activities in ill-structured problem situations involve assessing one's own resources, identifying additional needs and limitations, as well as exploring underlying assumptions influencing interpretation and reasoning (Jonasson, 2000). Ill-structured problems contain little constraints and thus closely resemble the complexity of reality and professional action (McNeill, Douglas, Koro-Ljungberg, Therriault, & Krause, 2016).

Designing ill-structured problems for PBL requires a framework that allows for precise and reliable levels of difficulty in learning with cases. Klug et al. (2016) suggested the content of the case to interact with the perception of difficulty. Accordingly, they varied the information about the school climate (e.g., attitude towards diagnostics), the diagnostic situation (e.g., routine assessment, in danger of failing class), and the student (e.g., attitude towards cooperation). In contrast, Hung (2006) and Jonasson and Hung (2008) suggested the structure and presentation of information to be decisive for the perception of difficulty.

Both models were successfully applied in case-based learning (Mohd-Ali et al., 2016; Trittel, Gerich, & Schmitz, 2014). In contrast to Klug et al. (2016), however, who regarded difficulty as a case characteristic, we conceptualized difficulty as an outcome, based on the difference of individual problem-solving skills and the complexity of information in the case. In our design framework, case complexity has two dimensions: (1) content structure and (2) information structure.

Table 4 shows the design matrix that we developed for the present studies. Naturally, combining characteristics with low and high complexity allows for tailoring cases to learning goals and pre-conditions.

For the two present studies, we planned the case complexity to be alternating but steadily increasing. Visualizing the complexity rating in a graph would result in a zigzag shaped line with an upwards trend. Using a short questionnaire in a small subsample of 4 to 11 pre-service teachers we found this shape to be true on a descriptive level for the first four of five text-based cases ($M_{\text{Case 1}} = 3.10$, $SD_{\text{Case 1}} = 0.88$; $M_{\text{Case 2}} = 2.73$, $SD_{\text{Case 2}} = 1.10$; $M_{\text{Case 3}} = 4.30$, $SD_{\text{Case 3}} = 1.06$; $M_{\text{Case 4}} = 3.82$, $SD_{\text{Case 4}} = 1.25$; $M_{\text{Case 5}} = 3.50$, $SD_{\text{Case 5}} = 1.29$). Inter-rater-reliability for the assigned ranks was high (Kendall's $W = .91$, $p = .042$).

Table 4

Design Elements for Constructing Text-Based Cases in the Domain of Teachers' Diagnostics

Case Characteristics	Low Complexity	High Complexity
<i>1. Content structure</i>		
1.1 Diagnostic situation	Routine assessment without time constraints (e.g., academic achievement)	Immediate danger of psychological or physical damage (e.g., bullying and self-harm)
1.2 Pre-knowledge demands	No domain specific pre-knowledge needed	Specific pre-knowledge needed (e.g., specifics about diagnostic instruments)
1.3 Topic of diagnosis	Culturally well-known and high prevalence (e.g., aggressive behavior)	Culturally less known and low prevalence (e.g., underachievement)
1.4 Competing diagnoses	One possible diagnosis	Multiple possible diagnoses
1.5 People involved in the described scene	Oneself as teacher and one student	Oneself and multiple other students, parents and colleagues
1.6 Attitudes towards diagnostics in the described scene	Positive and supportive	Negative and discouraging

Table 4 (Continued)

Design Elements for Constructing Text-Based Cases in the Domain of Teachers' Diagnostics

Case Characteristics	Low Complexity	High Complexity
<i>2. Information structure</i>		
2.1 Typicality	Identical to typical characteristics	Individualized and situated characteristics
2.2 Missing and distracting information	Only key information present	Key information missing and luring into judgment errors
2.3 Coherence	Logical framework for information pieces (i.e., making sense in the narrative world)	Varying frameworks for information pieces (i.e., not making sense)
2.4 Cohesion (i.e., linguistic elements supporting text-comprehension)	Using references between the information pieces	No references linking the information pieces
2.5 Consistency	Information points in the same direction	Conflicting information
2.6 Type of information	Uniform	Heterogenous (e.g., observation notes and test results)

8.2 Research Questions and Hypotheses

Previous interventions for promoting diagnostic competence primarily focused on individual learning processes, although professional interaction is often central to the diagnostic process. Since knowledge about interactive and collaborative forms of competence development, as well as empirical evidence on the effects of case-based learning is still scarce, the present pilot and replication studies aimed to examine the effects of problem-based learning on diagnostic competence. Does problem-based learning with text-based cases support the development of diagnostic competence?

We hypothesized that problem-based learning with text-based cases has a positive effect on dispositions of diagnostic competence, namely 1a) the cognitive skills in the domain of diagnostic knowledge, 2a) the motivation for pedagogical diagnostics, and 3a) the diagnostic self-concept. Furthermore, we proposed that effects of the learning environment can be replicated. Accordingly, we hypothesized successful replication for 1b) cognitive skills, 2b) motivation, and 3b) the diagnostic self-concept.

8.3 Methods and Design

Procedure

The present studies are quasi-experimental intervention studies with two measurement points each, that we conducted in the winter terms of 2016/2017 and 2017/2018. At the first measurement point, participants of the control and the intervention groups completed a knowledge test on pedagogical diagnostics and self-report questionnaires on the motivation for pedagogical diagnostics and the diagnostic self-concept. Afterwards, participants of the control and the intervention groups attended five seminar units on the basics of pedagogical diagnostics. Participants of the intervention groups then took part in problem-based learning with text-based cases (PBL-TF).

Preparing and Implementing PBL-TF

Naturally, the initial effort of constructing the material for the learning environment is substantial. The cases must be written with the specific target group of pre-service teachers in mind, as well as the curriculum and the scientific facts concerning the addressed learning pre-requisites and disorders. The cases employed in this study are available as Open Educational Resource on the first authors' website (Appendix 1). Additional work lies in preparing the students to be able to benefit from the learning environment. In our experience, not all students are equipped with the skills for successful self-regulated learning. Not only does self-regulated learning require a basic vocabulary of the content to engage in meaningful discussions as well as to search and understand thematically relevant literature. Groups also need support in keeping close to the steps of PBL.

We addressed these needs by working with the four-component instructional design model (4C-ID) by van Merriënboer and Kirschner (2007). Students received structuring aids in addition to the text-based cases (e.g., a worksheet on the steps of problem-based learning), worked on recurrent exercises (e.g., formulating hypotheses or feedback), and received supportive information such as literature on learning pre-conditions and learning disorders. The literature gave a first insight into possible topics of the respective text-based cases (e.g., bullying, test anxiety, ADHD). Surprisingly, a substantial number of students seemed to struggle with phrasing learning goals. This skill, however, improved over the course of learning with PBL-TF.

Working with the student groups often initially took the form of rephrasing and directing their thoughts, only to change after the first two or three cases towards observing their learning activities and addressing observations of productive behavior before beginning with the next case. McCaughan (2013) who compared guiding problem-based learning with Carl Rogers' approach to communication in client centered therapy documented similar experiences.

Intervention condition. After the five seminar units on the basics of pedagogical diagnostics, the lecturer introduced the theoretical background and the state of research on problem-based learning with text-based cases. Subsequently, the lecturer formed small groups of 5–7 students and exemplarily went through the steps of problem-based learning in a fishbowl design for the other groups to observe. The students' observations and experiences were reflected in the following seminar unit, deriving the support needs for the remainder of the seminar. For the remainder of the seminar, students worked on the text-based cases.

The text-based cases were theoretically constructed and validated by experts. The cases illustrate striking teaching situations relating to the most frequent, stable individual learning disorders such as test anxiety, attention deficit disorder or reading-spelling weakness (Linderkamp & Grünke, 2007), and contain supplementary material such as lists of school grades, completed observation sheets, or tests. How much of this information was included in the text, and how congruent, ambiguous, or distracting that information was, determined how complex the text-based cases were. According to the 4C-ID model by van Merriënboer and Kirschner (2007), which aims at promoting professional routines and schemata, the complexity of the text-based cases was intermittently increased during the course of the training. In line with the competence model by Klug et al. (2016), in a first phase of problem-based learning, students collected and discussed possible topics of the text-based cases, selected relevant information in group decisions and defined 2–4 shared learning goals such as being able to explain causes of test anxiety. Over the course of the following week, each student worked on every learning goal individually, followed by a second group session on the case. In this second session, the groups discussed their findings and thoughts, including next possible steps, phrasing a diagnostic working hypothesis, and finally reflecting their collaborative process. One person from the group moderated this process. In line with the criteria for professional communities by Vangrieken et al. (2015), the groups remained the same during the intervention. This supposedly allowed the groups to build the necessary identification and cohesion and enabled them to find shared goals and values for formulating their learning goals. Furthermore, from a practical point of view, keeping the groups stable supported the students' planning, documentation, and coordination efforts. After the last text-based case, all groups reflected the problem-based learning process and exchanged their learning outcomes in a group discussion.

Control condition. The students in the control condition attended a seminar with theoretical inputs and with instructed small group and individual work (e.g., providing feedback for a fictional student). To ensure that the intervention and the control condition were comparable, the same lecturer led the seminars, the seminars had the same curricular learning goals and the same literature was provided.

Sample

The sample consisted of pre-service teachers for general and vocational schools who participated in two parallel seminars on pedagogical diagnostics in winter terms of 2016/2017 and 2017/2018, respectively.

In the first study, 57 students participated in the first measurement point at the beginning of winter term 2016/2017 ($N_{\text{Intervention group}} = 19$, $N_{\text{Control group}} = 38$), and 43 students in the second measurement point at the end of the term ($N_{\text{Intervention group}} = 18$, $N_{\text{Control group}} = 25$). 39 students participated in both measurement points. The students in the intervention group were 55 % female and aged 23–44 years ($M = 29.5$, $SD = 5.1$). Their main subject was a vocational subject such as nutrition or food science. The students in the control group were 78 % female and slightly younger than in the intervention group (21–39 years, $M = 26.4$, $SD = 4.6$, $U = 131.500$, $z = -2.264$, $p = .024$). Their main subject was *Work Studies* (i.e., basic activities in workshops, laboratories, kitchens etc.). The students of the intervention and the control group did not significantly differ with respect to sex ($\chi^2 = 3.390$, $p = .066$), and were all in the first semester of master's program.

In the second study, 65 students participated in the first ($N_{\text{Intervention group}} = 24$, $N_{\text{Control group}} = 41$), and 58 students in the second measurement point ($N_{\text{Intervention group}} = 28$, $N_{\text{Control group}} = 30$), with 45 students participating in both measurements. In the intervention group, students (78 % female) were aged 22–31 years ($M = 26.8$, $SD = 2.6$). Their main subject was a vocational study such as nutrition or food science, and their minor subject was mathematics, languages etc. In the control group, students (67 % female) were 21–34 years old ($M = 26.5$, $SD = 3.6$), and they studied *Work Studies*. The intervention and the control group did not significantly differ with respect to gender, age, and study semester.

Measurement Instruments

The *knowledge on pedagogical diagnostics* was assessed using the PDW by Authors (2016). According to the model of diagnostic competence by Klug et al. (2012, 2016), the PDW contains items for five different content areas: judgment errors, feedback, procedure of diagnostics, individual learning pre-conditions, and test quality criteria. For each of these contents, the PDW contains items on the cognitive level of recognizing, understanding, and applying (Bloom, 1976).

Recognizing and understanding were each assessed with 20 multiple-choice items (4 per content area) and applying with 10 open items (2 per contents area), e.g., *You correct a written test of your class. The first ten tests are consistently bad. What do you pay attention to in the further correction?* or *A student doubts himself and is afraid of exams. In the last test he cried after a short time. You also had to give him a bad grade. What feedback do you give? Please phrase a sentence.* The first item assessed understanding of judgment errors such as the contrast-effect and the second item assessed applying of knowledge on feedback.

Responses to the open items were coded either as 0 = *false* or 1 = *correct* using a coding guide. The mean inter-rater-reliability of the open items in the first study (two raters) was almost perfect for the first measurement point ($\kappa_{\text{Cohen_t1}} = .92$). For the second measurement point the same two raters compared a small sample of open items and showed perfect agreement. For study 2, inter-rater-reliability of three raters was substantial $\kappa_{\text{Free_t1}} = .79$, $\kappa_{\text{Free_t2}} = .75$. Differing inter-rater-assessments were discussed and recoded in a team meeting. The item difficulties for study 1 varied between 0.02 and 0.91 at the first, and between 0.03 and 0.85 at the second measurement point. For study 2, the item difficulties varied between 0.00 and 0.92 at the first, and between 0.02 and 0.81 at the second measurement point. There are two versions of the test, one for the first and one for the second measurement point. To assess convergent validity, participants in study 1 also worked on a case-based scenario test developed by Klug et al. (2016). Correlations to the PDW levels of recognizing, understanding, and applying ranged between .25 and .42 for the first, and between .81 and .86 for the second measurement point. Given these correlations, the PDW shows convergent validity.

The *intrinsic motivation to engage in pedagogical diagnostics* was assessed with a self-assessment scale by Fernet, Senécal, Guay, Marsh, and Dowson (2008). The scale consists of three items (e.g., *I engage in pedagogical diagnostics because I enjoy it.*) that could be rated on a 4-point Likert scale from 1 = *does not apply at all* to 4 = *fully applies* ($\alpha_{\text{Study 1 t1/t2}} = .86/.78$, $\alpha_{\text{Study 2 t1/t2}} = .71/.87$).

The *diagnostic self-concept* was assessed using 15 items by Authors (2017), that were developed based on Helmke (2009) and Dickhäuser, Schöne, Spinath, and Stiensmeier-Pelster (2002). Students rated each item from 1 = *does not apply at all* to 4 = *fully applies* ($\alpha_{\text{Study 1 t1/t2}} = .91/.89$, $\alpha_{\text{Study 2 t1/t2}} = .93/.91$), e.g., *I am able to use instruments to measure learning pre-conditions (e.g., observation sheets, tests).*

8.4 Results

Descriptive Results

Descriptive results for knowledge, motivation, and self-concept, and *Mann-Whitney-U*-tests for the differences between the intervention and the control group are displayed in Table 5.

Knowledge on pedagogical diagnostics. In study 1, the intervention and the control group significantly differed in terms of their knowledge on pedagogical diagnostics. At the first measurement point this was true for the overall knowledge and for recognizing, but not for understanding and applying. At the second measurement point, the intervention group achieved significantly higher scores in recognition, understanding, and applying than the control group (see Table 5). In study 2, the intervention and the control group significantly differed in terms of their overall knowledge at the first measurement point, and in terms of their overall knowledge, recognizing, and understanding at the second measurement point (see Table 5).

Table 5

Means, Standard Deviations, and Mann-Whitney-U-Tests for Knowledge, Motivation, and Self-Concept

		Study 1			Study 2		
		IG	CG	<i>p</i>	IG	CG	<i>p</i>
T1	Knowledge (Overall)	10.69 (2.92)	7.94 (3.71)	.017	10.83 (2.46)	9.20 (3.11)	.050
	Recognizing	4.44 (1.71)	2.77 (1.55)	.004	4.13 (1.63)	3.55 (1.45)	.150
	Understanding	3.44 (1.50)	3.59 (1.52)	.787	4.04 (1.52)	3.75 (1.53)	.619
	Applying	2.81 (1.72)	1.87 (1.59)	.054	2.65 (1.30)	2.00 (1.38)	.076
	Scenario Test	7.00 (2.56)	6.11 (2.29)	.231	-	-	-
	Motivation	2.71 (0.40)	2.62 (0.85)	.955	2.59 (0.46)	2.71 (0.67)	.542
	Self-concept	2.40 (0.42)	2.70 (0.44)	.065	2.49 (0.62)	2.65 (0.57)	.391
T2	Knowledge (Overall)	13.17 (4.19)	8.31 (3.58)	.001	13.61 (3.48)	8.23 (4.20)	< .001
	Recognizing	5.17 (1.79)	3.92 (1.85)	.038	5.11 (2.11)	2.96 (1.43)	< .001
	Understanding	4.50 (1.43)	2.77 (1.42)	.001	5.44 (1.15)	3.44 (1.79)	< .001
	Applying	3.50 (1.92)	1.62 (1.27)	.002	3.24 (1.99)	2.35 (0.77)	.192
	Scenario Test	9.22 (2.90)	6.15 (2.98)	.001	-	-	-
	Motivation	3.15 (0.42)	2.97 (0.58)	.205	2.59 (0.53)	2.44 (0.77)	.708
	Self-concept	3.03 (0.34)	2.91 (0.38)	.183	2.92 (0.43)	2.83 (0.42)	.437

Notes. T1 = Time 1, T2 = Time 2, IG = Intervention Group, CG = Control Group; Knowledge 0–30 points, recognizing, understanding, and applying 0–10 points, Scenario Test 0–11, motivation 1–4, self-concept 1–4.

Motivation to engage in pedagogical diagnostics. In study 1 and 2, at the first and the second measurement point, the intervention and the control group did not significantly differ in terms of their motivation to engage in pedagogical diagnostics.

Diagnostic self-concept. In study 1 and 2, at the first and the second measurement point, the intervention and the control group did not significantly differ in terms of their pedagogical self-concept.

Repeated Measures ANCOVAs

We tested hypotheses 1a, 2a, and 3a using *repeated measures ANCOVAs* with the covariates age and sex. *Shapiro-Wilk* tests for small sample sizes showed that, except for applying, the dependent variables and the residuals of the dependent variables were normally distributed in the intervention and in the control groups, respectively. Since *ANCOVAs* are robust to violations of the assumption of normal distribution, we conducted *ANCOVAs* for applying as well (Norman, 2010; Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). The results for knowledge, motivation, and self-concept are presented in Table 6.

Knowledge on pedagogical diagnostics. In study 1 and study 2, the repeated measures *ANCOVAs* showed a medium to large main effect of the measurement point on the knowledge on pedagogical diagnostics, which increased over the course of the term both in the intervention and in the control group. For recognizing, understanding, and applying, there was no significant main effect of the measurement point (see Table 6). In study 1, the *ANCOVA* also showed a marginally significant, small intervention effect on recognizing and significant, large intervention effects on understanding and applying (see Table 6). While recognizing increased from the first to the second measurement point in both groups, understanding and applying increased only in the intervention group and decreased in the control group (see Figure 13). In study 2, in contrast, the *ANCOVA* showed a marginally significant, medium intervention effect on recognizing and non-significant, small intervention effects on understanding and applying (see Table 6). That is, recognizing and understanding increased in the intervention group and decreased in the control group from the first to the second measurement point. Applying increased both in the intervention and in the control group (see Figure 13).

Table 6

Repeated Measures ANCOVAS for Knowledge, Motivation, and Self-Concept

	Measurement point				Measurement point x Group			
	Study 1		Study 2		Study 1		Study 2	
	η^2_p	p	η^2_p	p	η^2_p	p	η^2_p	p
Knowledge (Overall)	.149	.042	.000	.909	.098	.104	.138	.026
Recognizing	.111	.089	.006	.655	.042	.307	.086	.088
Understanding	.088	.126	.011	.544	.167	.031	.052	.186
Applying	.037	.335	.001	.861	.145	.050	.019	.431
Scenario Test	.204	.046	-	-	.117	.140	-	-
Motivation	.003	.774	.000	.920	.081	.142	.006	.666
Self-concept	.506	< .001	.005	.696	.210	.014	.080	.118

Notes. Covariates age and sex. In study 2, age and sex are not significant, but bias the results for Measurement point x Group. ANOVAs without age and sex show that recognizing is significant, $\eta^2_p = .104$, $p = .048$, and self-concept is marginally significant, $\eta^2_p = .100$, $p = .068$.

Motivation to engage in pedagogical diagnostics. In study 1 and 2, the repeated measures ANCOVAs showed no significant main and no significant intervention effects for the motivation to engage in pedagogical diagnostics (see Table 6 and

Figure 14). In study 1, the motivation increased over the course of the term both in the intervention and in the control group. In study 2, the motivation did not change in the intervention group, whereas it slightly decreased in the control group.

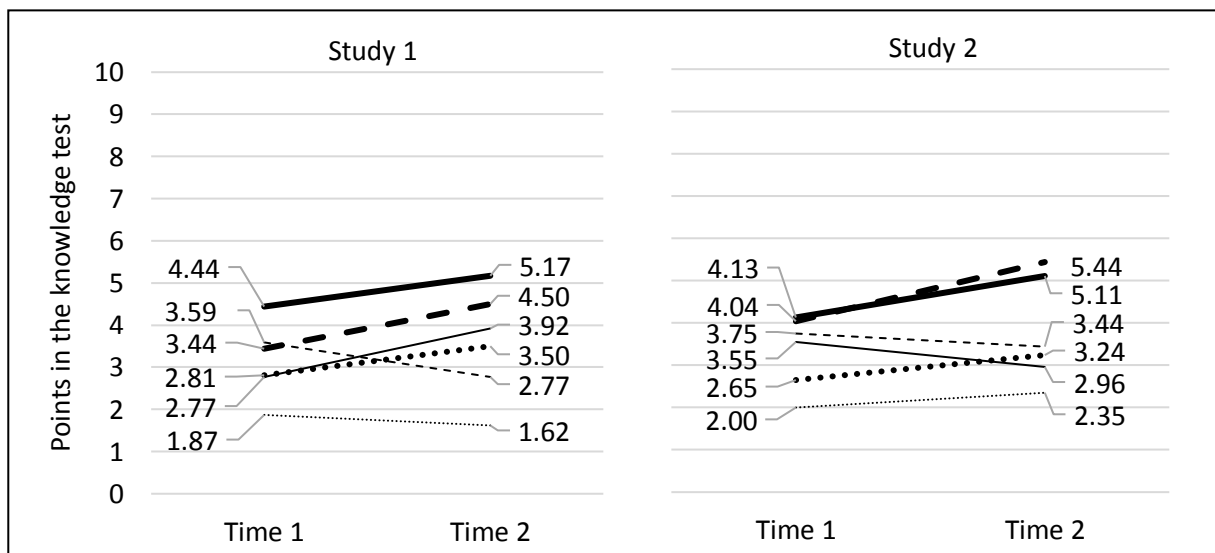


Figure 13. Effect of Measurement Point x Group on Recognizing (solid line), Understanding (dashed line), and Applying (dotted line); line for the intervention group is bold, and line for the control group is not; range 0–10.

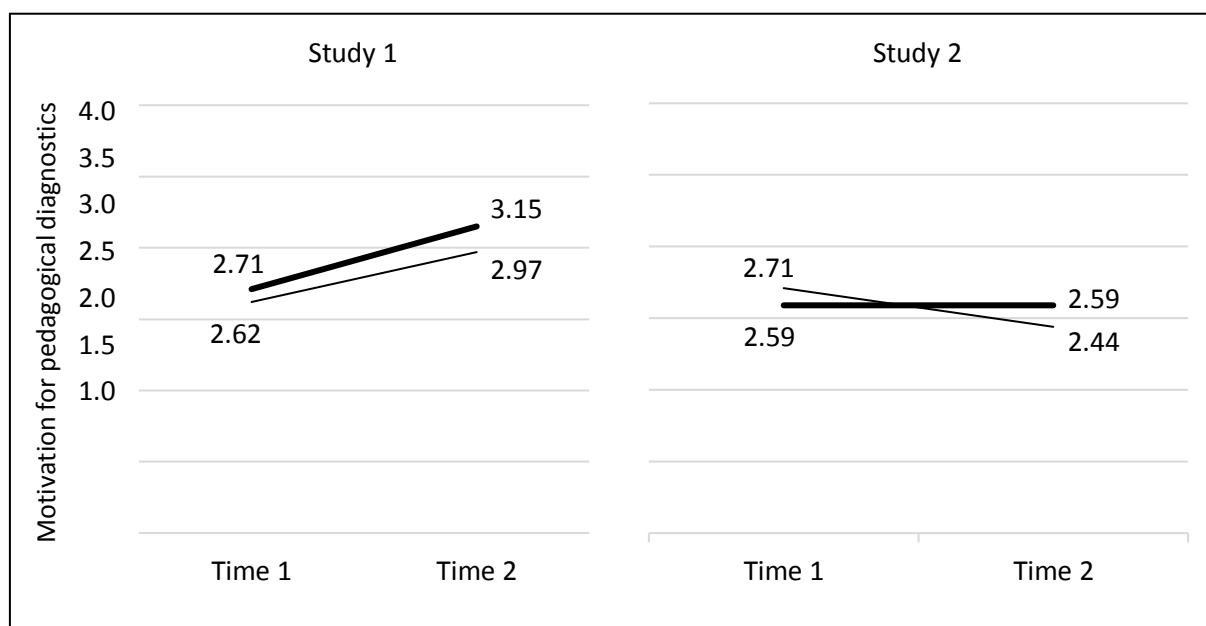


Figure 14. Effect of Measurement Point x Group on Motivation to engage in pedagogical diagnostics; line for the intervention group is bold, and line for the control group is not; range 1–4.

Diagnostic self-concept. In study 1, the *repeated measures ANCOVAs* showed a significant, large main effect of the measurement point and a significant, large intervention effect on the diagnostic self-concept (see Table 6). The diagnostic self-concept increased in both groups, but especially in the intervention group (see Figure 15). In study 2, on the other hand, there was no effect of the measurement point and no significant intervention effect, because the diagnostic self-concept increased both in the intervention and in the control group (see Figure 15).

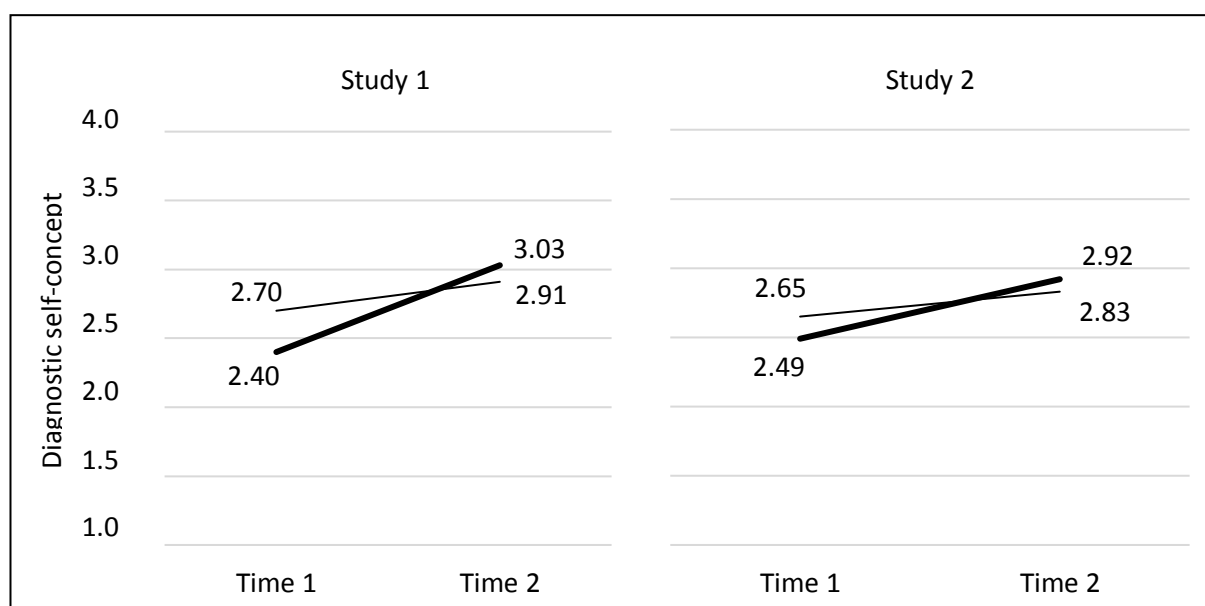


Figure 15. Effect of Measurement Point x Group on diagnostic self-concept; line for the intervention group is bold, and line for the control group is not; range 1–4.

Meta-Analytic Models for Summarizing the Replication

To evaluate the replicability and combined effect estimates (hypotheses 1b, 2b, 3b), we applied a meta-analytic approach. Since simply averaging effect sizes or pooling data can result in incorrect estimations of standard errors and false test statistics (Moeyaert, Ferron, Beretvas, & van den Noortgate, 2014), meta-analytic regression models facilitate correct estimation of true intervention effects. Thus, we tested the hypotheses with multivariate fixed- and random-effects regression models of the standardized mean changes in the outcome variables from the pre- and post-measurement (Becker, 1988; Berkey, Hoaglin, Antczak-Bouckoms, Mosteller, & Colditz, 1998; Morris, 2008). Despite the comparatively small sample size ($k = 2$), a power analysis yielded a satisfactory power of .69 for outcomes with high heterogeneity between the studies and of .93 and above for outcomes with moderate to low heterogeneity (Valentine, Pigott, & Rothstein, 2010). Heterogeneity (I^2) is a measurement that is central for interpreting meta-analysis results. It evaluates the comparability of studies included in a random-effects model, with higher heterogeneity referring to less comparable studies and consequently less precise estimation of combined effects (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003). Suggested cut-off values are 25 % for low, 50 % for intermediate, and 75 % for high heterogeneity (Higgins et al., 2003). For example, $I^2 = 85$ % in the random-effects model about recognizing knowledge (see Figure 16) means: from the total variance of the studies, 85 % is due to systematic between-study differences and the estimation cannot be expected to be precise. We conducted the analyses with the *metafor* package in R and estimated the effects with REML (Viechtbauer, 2010).

Results of the meta-analytic regression models are displayed in Table 7 and Figure 16. Table 7 displays the regression coefficients β as well as comparative fit indices for the fixed- and random-effects models. A regression coefficient of $\beta = 1$ indicates that the intervention group changed one standard deviation more than the control group. Comparative fit indices allow a descriptive comparison between the fixed- and the random-effects model. Both models show similar results, with differences in estimates and p -values mostly due to the limited sample size, and much higher degrees of freedom in the random-effects model (see Table 7).

Precise interpretation of the results can be supported by visualizing the data in a lattice-like forest plot. Figure 16 shows the effect sizes of the individual studies, as well as the combined effect sizes according to the meta-analytic models. Heterogeneity (I^2) is added for each subgroup. Suggested cut-off values (stemming from meta-analyses in the medical context) are 25 % for low, 50 % for intermediate, and 75 % for high inconsistency (Higgins et al., 2003). For example, $I^2 = 75$ % means that 75 % of the total variance in the intervention effects is due to between-study variation and not measurement error. In this case, aside from random measurement error, systematic sources for the difference between studies likely exist. Since fixed-effect models assume one true effect size for all studies, between-study variance and consequently heterogeneity is set to zero.

Table 7

Multivariate meta-analytic regression models with comparative fit indices

		Fixed-Effect Model			Random-Effects Model		
		β	SE	p	β	SE	p
Outcomes	Knowledge (Overall)	1.176	0.376	0.002	1.059	0.462	0.022
	Recognizing	0.193	0.359	0.590	0.316	0.700	0.652
	Understanding	1.131	0.366	0.002	1.082	0.327	< 0.001
	Applying	0.467	0.328	0.155	0.366	0.330	0.267
	Motivation	0.629	0.299	0.035	0.525	0.285	0.066
	Self-concept	0.734	0.313	0.019	0.665	0.387	0.085
Comparative Fit Indices	AIC	21.858			54.528		
	AICc	49.858			1566.528		
	BIC	25.252			48.906		
	Log-likelihood	-3.929			-0.264		
	Deviance	11.480			0.528		
	Df	7			27		

Knowledge on pedagogical diagnostics. Both the fixed-effect and the random-effects model showed a significant combined intervention effect on overall knowledge (see Figure 16). Heterogeneity regarding this construct was moderate to low, so that the combined effect size is likely to be a good estimate of a common mean. For recognizing, despite the significant and strong intervention effect in study 2, there was no significant combined effect (see Figure 16). There was high heterogeneity, consequently the combined effect size was much lower than effects of the individual studies and a source of systematic between-study variance was present. For understanding, both models showed a significant combined intervention effect with very low heterogeneity (see Figure 16). For applying, in contrast to the strong and significant intervention effect in study 1, there was moderate heterogeneity and the combined effect of the intervention was not significant (see Figure 16).

Motivation to engage in pedagogical diagnostics. Both the fixed-effect and the random-effects model showed a similar combined effect size estimation, with the fixed-effect model showing a significant effect and the random-effects model showing a marginally significant effect. There was moderate to large heterogeneity, translating to some systematic difference between the studies. In contrast to the negative effect in study 1, the combined effect was positive (see Figure 16).

Diagnostic self-concept. Both models, again, showed a similar combined effect size estimate. There was moderate heterogeneity and the random-effects model showed a marginally significant combined effect. The fixed-effect model estimated the combined intervention effect to be significant (see Figure 16).

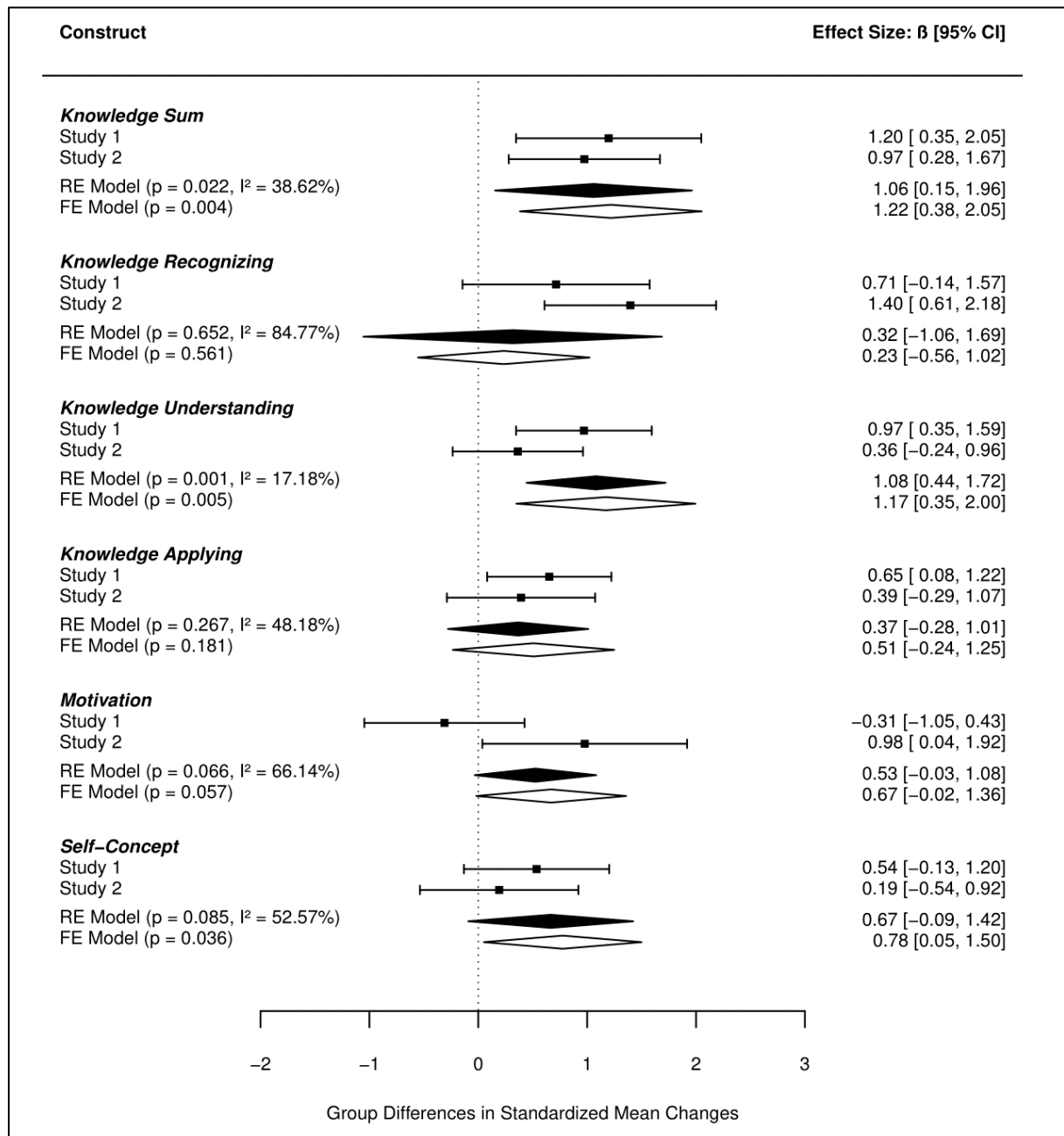


Figure 16. Forest Plot for Standardized Mean Change Differences in Individual Studies and Estimated Multivariate Fixed-Effect and Random-Effects Models; FE = fixed-effect, RE = random-effects, p -values account for the significance of the regression coefficient, not for the model fit.

8.5 Discussion

The present studies examined the effects of problem-based learning with text-based cases on pre-service teachers' diagnostic competence. Problem-based learning was chosen as instructional method, because the stepwise learning activities that the students engage in closely resemble the diagnostic process and specifically incorporate collaboration in these steps. We assessed diagnostic competence based on the COACTIV model of teachers' professional competence. We hypothesized that problem-based learning with text-based cases has a positive effect on 1a) cognitive skills in the domain of knowledge about diagnostics, 2a) the motivation for diagnostics, and 3a) the diagnostic self-concept. These hypotheses were extended to replicability of the effects on 1b) cognitive skills, 2b) motivation, and 3b) self-concept.

Effects of Problem-Based Learning with Text-Based Cases on Knowledge, Motivation, and Self-Concept

Cognitive skills in the domain of diagnostic knowledge. In study 1, the overall knowledge of pedagogical diagnostics increased in both the control and the intervention group. However, understanding and applying decreased in the control group and increased in the intervention group. The decrease in the control group can be explained by having varied the instruction method and exercises with the content. In line with cognitive load theory (Sweller, 2003), students would have to accommodate new content and new instructional methods in parallel, possibly resulting in shared processing and reduced content memorization.

The increase in the intervention group is in line with research on comparing problem-based learning with direct instruction (Bergstrom et al., 2016). This research suggests that PBL-TF successfully triggers deep cognitive processing involving cognitive elaboration, which in turn enhances knowledge transfer and connectedness. The self-directed and joint definition of learning goals in problem-based learning likely fostered the integration of new learning content with pre-knowledge. In line with research on expertise (Carbonell, Stalmeijer, Könings, Segers, & van Merriënboër, 2014; Kim & Klassen, 2018; Lachner, Jarodzka, & Nückles, 2016), the clear cognitive structure of the recurring steps of problem-based learning possibly encouraged forming content flexible cognitive routines and schemata (e.g., how to select and interpret diagnostically relevant information). Specifically, this possibly was facilitated through the similarity of learning behavior in the intervention and application of diagnostic knowledge in the applied test. Thus, problem-based learning with text-based cases is a good complement to the training of perceptual skills with video cases (Schäfer & Seidel, 2015). Results of the scenario test by Klug et al. (2016) showed very high correlations with cognitive skills in the knowledge domain. While Klug et al. (2016) constructed the test as a performance measure, our results rather indicated that the test captures dispositions in the knowledge domain, but not diagnostic performance. It remains an open question, how diagnostic performance can economically be tested in an experimental setting.

In study 2, the overall knowledge on pedagogical diagnostics also increased in both the control and the intervention group. Unlike study 1, understanding and applying did not differ after the intervention, but recognizing did. Recognizing decreased in the control group and increased in the intervention group. Thus, problem-based learning might not per se stimulate interpretation and transfer skills in teachers' diagnostics, as study 1 would suggest. In fact, the trained skillset might depend on the specific learning goals that students define. It is possible that in study 1 students defined learning goals that focused on understanding and applying, and students in study 2 defined learning goals that focused on recognition.

Although there was variation in the specific cognitive skill that pre-service teachers gain with PBL-TF, we observed positive effects in both studies. Hypothesis 1a can therefore be retained.

Hypothesis 1b can be judged by the meta-analytical models. Heterogeneity concerning recognition was much higher than concerning understanding, applying, and overall knowledge. This leads to a more conservative estimation of the intervention effect on recognizing, even though the individual studies suggest otherwise. The mixed results about recognizing (i.e., declarative knowledge), are in line with research on problem-based learning (Dochy et al., 2003; Strobel & van Barneveld, 2009). This can be explained by the very central element of the students' learning goals. While all students gained experience in understanding and applying knowledge, a focus on declarative knowledge can only be set by specifically addressing it in the learning goals, e.g., what questions instead of how or why questions about learning pre-requisites or disorders. The positive intervention effect on understanding and overall knowledge was much more pronounced in the multivariate regression models, possibly due to correlations between the measured outcomes, that *ANCOVAs* do not sufficiently account for. This result additionally supports the interpretation that problem-based learning leads to more integrated and transferable knowledge.

Hypothesis 1b can therefore only be retained for the cognitive skills of understanding and applying in the domain of diagnostic knowledge. To gain more insight in this regard, future studies should assess and control learning goals.

Motivation to engage in pedagogical diagnostics. Surprisingly, in study 1 the motivation to engage in pedagogical diagnostics increased in both the intervention and the control group despite there was no specific intervention effect in the *ANCOVAs*. In study 2, motivation did not change in the intervention group and slightly decreased in the control group. These results are very much in contrast to the typically positive effect of PBL on motivation (Hwang & Kim, 2006; Sangestani & Khatiban, 2013; Sungur & Tekkaya, 2006). A first possible explanation for this is that the motivation scale did not differentiate for casework and for the specific effects of self-regulated learning. Furthermore, the presented study only measured intrinsic motivation, but neither identified nor introjected motivation. Both are constructs in the self-determination theory of motivation (Ryan & Deci, 2000) that could not be assessed reliably in the presented studies. This is likely due to the requirements of the utilized measurement instrument, as some items are difficult to answer without prior experience in diagnosing students. If possible, additional research on the topic should keep these motivational constructs in mind or work with goal-specific motivation regarding learning and achievement (Elliot, 1999). For instance, in comparing text- and video-cases, Syring et al. (2015) found that pre-service teachers were both more learning-motivated and more performance-anxious when working with videos.

The text-based cases that were a source for motivational differences in the study of Noordzij and Te Lindert (2010), are very unlikely to play the same role in the present studies. Since the groups in both studies worked with the same cases and had comparable pre-knowledge on teacher's diagnostics and motivation, the cases could only then induce the observed differences if perceptual skills or preferences differed between the studies. Reading comprehension and related constructs were not measured, but are unlikely to have a strong impact, since there was no time limit for reading the cases and since they were discussed in the groups.

A third possible explanation refers to methodological considerations. Both groups were taught by the same lecturer. That is, the intervention effect may be overshadowed by the lecturer's effect on the student's motivation. In future studies, this effect should be controlled, e.g., by systematically varying the lecturers and by separately assessing the motivational qualities of the lecturer and the learning environment.

A fourth explanation can be thought of with results of meta-studies about teacher collaboration (Vangrieken et al., 2015; Vangrieken et al., 2017). These reported a perceived reduction of autonomy while pursuing collaborative activities. Since autonomy is a determinant of intrinsic motivation (Ryan & Deci, 2000), our studies might show that this mechanism can already be prevalent in pre-service teacher education. Based on the results of both studies, hypothesis 2a must be rejected. But what about the replication?

The meta-analytic results, although only marginally significant, suggested a positive intervention effect. The moderate to high heterogeneity pointed to systematic differences between the studies that cannot be accounted for by measurement error alone. This difference could also be due to the time spent learning outside the training sessions. Anecdotally we can report that seemingly more motivated pre-service teachers also spent more time on researching for their learning goals (e.g., by additionally working on a learning diary at home). For future studies with problem-based learning, we strongly advise to control for the time spent on studying and possibly setting time limits. Especially in settings of self-regulated learning, the time spent on learning can be an indicator for motivation. Still, based on the study results, hypothesis 2b must also be rejected.

Diagnostic self-concept. In study 1, though diagnostic self-concept increased in both the intervention and the control group, increases in the intervention group were significantly higher. In study 2, the results were not significant but were very similar on a descriptive level. Thus, problem-based learning with text-based cases contributed to a more differentiated and more positive assessment of one's own skills in teachers' diagnostics. This intervention effect might be due to continuously sharing ideas and setting common learning goals in the problem-based learning groups, as social comparison processes and feedback are fundamental sources for the self-concept (Dickhäuser et al., 2002). Hypothesis 3a can therefore be retained. Small differences in the studies might be explained by differences in the composition of problem-based learning groups. The meta-analytic results showed a marginally significant positive effect for the more complex random-effects model and a significant positive effect for the fixed-effect model. Thus, hypothesis 3b can also be retained, which is particularly promising since the self-concept is a relatively stable construct (Cairns et al., 1990).

In summary, problem-based learning with text-based cases was judged a learning environment for pre-service teachers, in which they especially develop skills relating to diagnostics in the cognitive

domain. While effects on strictly declarative knowledge varied with the participant groups, our studies showed stable positive effects on the more elaborate skills of reasoning and applying diagnostic knowledge. Furthermore, the training had a stable positive effect on the diagnostic self-concept. Centering the learning environment on collaborative activities came, however, with the cost of unchanging or decreasing motivation.

Implications

Collaborative activities encouraged by the instructional method of problem-based learning can promote diagnostic competence in the first phase of teacher education. That is, pre-service teachers who have experience in collaborative learning activities about diagnostics, perform better than students without this experience. This is a strong argument for including professional interaction and collaboration in models of diagnostic competence. This furthermore supports a more situated view on professional competence, as suggested by Santagata and Yeh (2016). Research is only beginning to understand the interplay of teachers' dispositions and their working environment.

To further examine diagnostic competence, future studies on diagnostic activities and trainings should include constructs that are common in social psychology and socialization research, such as group decisions, norm production, and success or failure attribution. Studies on teachers' professional development would also gain validity by including interpersonal constructs such as empathy and social support. In line with this implication, in an intervention study with classroom videos by Seidel et al. (2011) teachers did not engage in teaching analysis until they set a common learning goal. Colleagues and other professionals are part of teachers' social environment (L. Shulman, 1992), therefore ongoing socialization processes in the professional community should be considered in both the assessment and the development of diagnostic competence.

Well-established educational norms and routines in schools are perceived as overpowering by pre-service teachers, who then resort to masking their competences developed in higher education rather than communicating their own perspectives and risking a conflict (Ghousseini & Herbst, 2016; Hoth et al., 2016). In line with this result, Main (2007) identified early exposition to environments of professional collaboration as pre-requisite for building effective collaborations in schools. L. Shulman (1992) addressed this problem in his outline of case-based learning in teacher education as well. Based on cognitive psychology, L. Shulman (1992) argued that teaching demands cognitive flexibility and situated cognition to balance the tension between current and desired practice, and that "this tension becomes particularly acute when students are sent out as observers, interns, student teachers, or apprentices to work under real conditions" (p. 8). Since previous models of teachers' diagnostic competence mostly focus on the teacher-student relationship, the limited transfer of competences into the school environment might also point to the necessity of training professional interactions in teacher education. Insights about transitioning from teacher education to teaching could be gained in longitudinal studies, which are still rare.

Text-based cases are an appropriate, adaptable, and comparatively economical case medium. In comparison to video-based cases, the demands on perception are much lower, highlighting demands

on interpretation and decision-making skills. Working with text-based cases including additional diagnostic material – as provided in the present interventions – promotes skills that are necessary in the future working environment (e.g., making choices based on standardized diagnostic instruments). Problem-based learning gives pre-service teachers the opportunity to experience these situations at an early stage of their higher education training without risks for themselves and others. One of the few studies comparing text-based and video-based cases (Syring et al., 2015) did unfortunately not control for the comparability of groups prior to the interventions. If our studies would be replicated by other researchers, and we strongly encourage this, another intervention group working with video-based cases would provide more insights about benefits of each learning material.

However, one must keep in mind that implementation of problem-based learning with text-based cases faces some institutional constraints (e.g., incongruous forms of student examination, the size of the learning group, and specific curricular goals in relation to other parts of the curriculum). Moreover, it is still unclear whether the increase in knowledge, motivation, and self-concept gained through problem-based learning can be carried over to working in schools, regardless of dominant school norms and routines. A longitudinal study could identify beneficial and obstructive factors for transferring the acquired competences into practice.

From a methodological perspective, comparing individual study results with meta-analytic models was surprisingly insightful. Where results from the individual studies diverged (e.g., motivation and recognizing), the regression models indicated whether differences were more likely due to measurement error or systematic differences. Additionally, since the meta-analytic models can take advantage of the combined sample size, estimations are more precise than from single *ANCOVAs* and have greater statistical power. From a practical standpoint, a reliable answer about replicability can be very beneficial for arguing institutional changes in teacher education, as results are less likely to be biased by uncontrolled factors that are introduced by different groups of students. Consequently, the effort of replicating a study concerning teacher education contributes to establishing objective findings and estimation of true effects in a field, where subjectivity and perceptual differences are very pronounced and necessary to create a rich learning environment for students.

Limitations and Future Research

The presented studies have limitations concerning the sample. First, a bigger sample size and additional intervention groups working on classroom videos or video- and text-based cases would allow for additional interpretations about perceptual abilities and cognitive elaborations. This would allow interpretations about which individual dispositions and which group characteristics moderate the competence development. Second, randomizing the participants to the intervention and control group was not possible for organizational reasons (the seminars were offered for specific groups of students who were not free to choose within the curriculum). However, the groups were empirically well comparable and despite the small sample size we were able to show strong effects.

Another limitation concerns the intervention. The developed training focused on a selection of cognitive, behavioral, and affective-motivational learning pre-conditions among students, not, for

example, on further topics such as content knowledge, physical disabilities, or parent-teacher communication. Moreover, the results of the present intervention study cannot directly be compared with previous findings on diagnostic competence, since previous findings mostly refer to the model of judgment accuracy.

Preparing pre-service teachers for conducting professional diagnostic activities in classrooms and professional interactions places strong demands on teacher training programs in higher education. Trainings need to address the necessary skills and dispositions in both contexts. In many diagnostic cases, competing hypotheses exist, resulting in the need for diverse information gathering, as well as goal specific interactions with other professionals, before a preliminary diagnosis can be reached. Problem-based learning is a viable method to prepare for these demands, since it not only entails applying knowledge to a case, but also weighing different interpretations against each other, making collaborative decisions and engaging in self-directed learning.

While several models of diagnostic competence consider professional interaction a beneficial factor, this assumption is not always based on context specific research and does not address the dispositions and situation specific skills that are necessary for successful collaborative efforts. This is possibly due to the lack of research on outcomes, quality, and objective measurement of teachers' diagnostics aside from judgment accuracy (Altmann & Nückles, 2017). A construct that at least builds on professional interactions is the diagnostic agreement. Studies in medicine showed that diagnostic agreement not only has a positive effect on patients' compliance and motivation, but also on treatment success (Jensen-Doss & Weisz, 2008; Pogge et al., 2001). As a first step in schools, different settings of collaborative and interactive settings could be identified (e.g., case conferences, documentations), as well as their effects on individual diagnostic competence and students' academic achievement, learning behavior, and achievement emotions. Because a certain degree of autonomy in the choice of when and how to collaborate can be very beneficial for the success of collaborative efforts, mainly due to fear of autonomy loss being a predictor of unsuccessful collaboration (Johnson, 2003; Somech, 2010).

Since research on developing professional competences and designing evidence-based interventions involves manifold decisions and systematic differences between studies and local implementations, replicating results is a justifiable effort. While the additional time spent on collecting data may lead to more complex data and longer publication intervals, the benefits of more precise interpretations, estimates of true effect sizes, and avoiding premature publication and implementation far outweigh the drawbacks (Open Science Collaboration, 2015).

9 Studie 3: Evidenzbasierung und Zusammenhangsanalysen in der Lehrkräftebildung – Diagnosewissen, Selbstkonzept und Einstellungen zu Inklusion in manifesten Wachstumsmodellen³

9.1 Einleitung und theoretischer Hintergrund

Evidenz kann die Lehrkräftebildung über Rahmenbedingungen, Prozesse und Ergebnisse von Bildung informieren, die zum gegenwärtigen Zeitpunkt mit wissenschaftlichen Methoden messbar sind. Mit dem Ziel eines weitgehend inklusiven Schulsystems steigt auch der Bedarf an evidenzbasierten Lerngelegenheiten zur Förderung inklusiver Lehrkompetenzen (Darling-Hammond, 2017). Da klassische Kompetenzmodelle Einstellungen als zentrale Komponente professioneller Lehrkompetenz betrachten (Baumert & Kunter, 2006), stellen inklusionsbejahende Einstellungen und Möglichkeiten ihrer Förderung einen Schwerpunkt der Forschung zur Lehrkräftebildung dar (Hellmich & Görel, 2014).

Positive Einstellungen zu inklusivem Unterrichten gehören zu den wichtigsten Einflussfaktoren einer lernförderlichen Unterrichtspraxis in heterogenen Lerngruppen (Avramidis & Norwich, 2002; Savage & Erten, 2015). Trotz überwiegend positiver Einstellungen berichten viele Lehrkräfte und Lehramtsstudierende jedoch Ängste und Befürchtungen in Bezug auf die Unterrichtsdurchführung (Schwab & Seifert, 2015; Syring, Tillmann, Weiß, & Kiel, 2018) und Lehrkräfte geben einen Mangel an inklusionsbezogenen Aus- und Weiterbildungsinhalten an (OECD, 2014).

Die Forschung zur Lehrkräftebildung knüpft an diese Ergebnisse durch die Untersuchung von Wissensbeständen und Maßnahmen zur Wissensvermittlung an, die eine positive Einstellungsbildung und -änderung bewirken können (Schürer, 2018). Dies ist auch der Schwerpunkt der vorliegenden Studie, die den Zusammenhang von kognitiven Fähigkeiten im Bereich des Diagnosewissens mit Einstellungen zu inklusivem Unterrichten in einer quasi-experimentellen Studie mit einem selbstgesteuerten versus instruktionsbasierten hochschuldidaktischem Lernsetting untersucht. Dabei wird ein methodischer Weg der Analyse dargestellt, der einen unverzerrten Blick auf diese Zusammenhänge von Diagnosewissen, Einstellungen zu inklusivem Unterrichten und diagnostischem Selbstkonzept sowie deren Veränderung ermöglichen. Kern des vorliegenden Beitrags ist die Frage: Wie entwickeln sich diese Zusammenhänge, wenn Diagnosewissen mit selbstgesteuertem Lernen oder mit instruktionsbasiertem Lernen vermittelt wird?

³ Diese empirische Studie erscheint 2020 und kann wie folgt zitiert werden: Wedel, A., Pfetsch, J., & Ittel, A. (2020). Evidenzbasierung und Zusammenhangsanalysen in der Lehrkräftebildung – Diagnosewissen, Selbstkonzept und Einstellungen zu Inklusion in manifesten Wachstumsmodellen. In: *Journal für Psychologie*, 27(2), 356-381. <https://doi.org/10.30820/0942-2285-2019-2-356>

Theoretischer Rahmen und Forschungsstand: Diagnosewissen und Selbstkonzept als Basis der Einstellung zu inklusivem Unterrichten

Einstellungen sind evaluative Reaktionen auf Einstellungsobjekte und beinhalten eine kognitive, affektive und behaviorale Komponente (Bohner & Wänke, 2002). Einstellungsbildung und -änderung werden zumeist mit dualen Prozessmodellen und insbesondere mit dem Elaboration Likelihood Model (ELM) erklärt (Petty & Cacioppo, 1986). Das ELM erklärt Einstellungsänderungen durch das Ausmaß an kognitiver Elaboration, mit der eine Botschaft in Bezug auf das Einstellungsobjekt verarbeitet wird, und die wahrgenommene Stärke, die diese Botschaft hat.

Mehrere Studien weisen darauf hin, dass sowohl wissenschaftliche Evidenz als starkes Argument wahrgenommen wird (Winter & Krämer, 2016; Winter, Krämer, Rösner, & Neubaum, 2015) als auch selbstgenerierte Informationen (Darke et al., 1998). So weist Wenden (1998) darauf hin, dass selbstgesteuertes Lernen zur Analyse und Revision bisheriger Einstellungen und Überzeugungen führen kann. Diesen Effekt auf Einstellungsänderungen unterstützen auch aktuellere Studien zu problemorientiertem Lernen. Diese Form selbstgesteuerten Lernens führt im Vergleich zu lehrkräftezentrierten Lernformen eher zu Einstellungsänderungen in Bezug auf die Lerninhalte, was sowohl Meta-Analysen zeigen (Demirel & Dağyar, 2016) als auch Studien spezifisch in Bezug auf Lehramtsstudierende (z. B. Kuvac & Koc, 2018).

Im ELM kommt auch dem Selbstkonzept eine Rolle in der Einstellungsänderung zu. Die ‚functional matching‘ Hypothese besagt, dass die Verarbeitung von zum Selbstkonzept kongruenten Informationen eine positive Bewertung der Informationen und eine stärkere Einstellungsänderung bewirkt (Lavine & Snyder 1996; Petty & Wegener 1998), insbesondere wenn das Selbstkonzept in der Beurteilungssituation salient ist (Fleming & Petty, 2000).

In einem engen konzeptionellen Zusammenhang zu inklusionsorientiertem Unterrichten steht nicht nur das professionelle Selbstkonzept, sondern auch das Wissen aus dem Bereich pädagogischer Diagnostik. So wird Diagnosewissen nicht nur konzeptionell als Bestandteil des Professionswissens angesehen (Voss et al., 2015), sondern in neueren Modellen adaptiver Lehrkompetenz auch als Voraussetzung inklusiven Unterrichts beschrieben (Baumert & Kunter, 2011; Rey, Lohse-Bossenz, Wacker, & Heyl, 2018) und ist darüber hinaus ein zentraler Bestandteil von Testinstrumenten zu inklusionsrelevanten Wissensbeständen (König, Gerhard, Kaspar, & Melzer, 2019; Marx et al., 2017).

Die empirischen Befundlage des Zusammenhangs von Diagnosewissen und Einstellungen zu inklusivem Unterrichten wird in systematischen Reviews als sehr einheitlich dargestellt. Sowohl das Meta-Review von Miegheem und anderen (2018), als auch vorherige Reviews zu inklusionsbezogenen Einstellungen von Lehrkräften (Avramidis & Norwich, 2002; de Boer, Pijl, & Minnaert, 2011) berichten einen positiven mittleren Zusammenhang. Grundlage dieser Bewertung sind jedoch Studien mit bedeutsamen methodischen Einschränkungen. Dies betrifft zum einen deskriptive Beobachtungen eines parallelen Anstiegs bzw. Niveaus von Wissen und Einstellungen (Kurniawati, de Boer, Minnaert, & Mangunsong, 2017; Liu et al., 2016; Mulholland, 2016; Youssef & Youssef, 2015) und zum anderen Interventionsstudien, die lediglich querschnittliche Zusammenhänge mit Pearson Korrelationen berechnen, nicht aber Zusammenhangsveränderungen schätzen oder für eine autokorrelative Struktur

der Konstrukte kontrollieren (Avramidis, Bayliss, & Burden, 2002; Batsiou et al., 2008; Ghanizadeh, Bahredar & Moeini, 2006). So bleibt unklar, ob in einer großen Zahl verfügbarer Studien das Wissen eine Grundlage der Einstellung bildete oder die Einstellung zum Wissenserwerb motivierte. Beide Ansätze der Datenanalyse resultieren größtenteils aus einem Mangel adäquater Stichproben bzw. geeigneter Analyseverfahren für die vorliegenden Stichproben.

Eine weitere Einschränkung ist die Erfassung der Wissensbestände durch Selbsteinschätzungen (Avramidis et al., 2002; Batsiou et al., 2008; Mulholland, Cumming & Jung, 2015; Mulholland, 2016; Sermier Dessemontet, Benoit, & Bless, 2011). Selbsteinschätzungen zu Wissen und Fähigkeiten sind Teil des professionellen Selbstkonzepts und stehen nur in einem mittleren Zusammenhang zum objektiven Professionswissen (Retelsdorf, Bauer, Gebauer, Kauper, & Möller, 2014). In einem Review beurteilen Ewing, Monsen und Kielblock (2018) den Zusammenhang zwischen den Einstellungen zu inklusivem Unterrichten und dem professionellen Selbstkonzept, das auch explizit als solches gemessen wurde, als positiv. Das Review differenziert dabei zwar nicht zwischen verschiedenen Facetten des professionellen Selbstkonzepts, jedoch sind die Ergebnisse sehr einheitlich.

Zum Zusammenhang zwischen dem professionellen Selbstkonzept und dem Diagnosewissen ist die Studienlage klarer. So kommen Interventionsstudien mit Lehramtsstudierenden ebenso zum Ergebnis positiver mittlerer bis großer Zusammenhänge (Klug, Bruder, Kelava, Spiel, & Schmitz, 2013; Wedel, Müller, Pfetsch, & Ittel, 2019), wie auch querschnittlich angelegte Studien (Klug, Bruder & Schmitz, 2016; Ohle, McElvany, Horz, & Ullrich, 2015).

Zusammengefasst besteht nach wie vor Forschungsbedarf zu Zusammenhängen zwischen Diagnosewissen, Selbstkonzept und Einstellungen zu inklusivem Unterrichten. Wenngleich die zahlreichen Interventionsstudien ein Beleg für die Veränderbarkeit dieser Konstrukte in formalen Bildungskontexten sind (z. B. Lehrveranstaltungen, Weiterbildungen), fehlt es an belastbaren statistischen Ergebnissen, die mit größerer Sicherheit als wissenschaftliche Evidenz interpretiert werden können. Besonderes Interesse besteht an Interventionsstudien, da diese über die querschnittliche Zusammenhänge hinaus auch die Veränderung dieser Zusammenhänge adressieren und damit Auskunft über die Effektivität von Wissenserwerb als Grundlage von Einstellungsänderungen geben.

Herstellung von inklusionsbezogener Evidenz in der Lehrkräftebildung

In der Wissenschaftstheorie beschreibt Evidenz die regelgeleitete Verbindung theoretischer Wissensbestände mit einem Ausschnitt der Realität (Mullen, 2016). Bromme, Prenzel und Jäger (2016) schlagen vor, dass Evidenz mit wissenschaftlichen Methoden gewonnene Daten oder Befunde sind, die zur Prüfung von Theorien verwendet werden. Diese Definition entspricht dem Begriff forschungsbezogener Evidenz im kritischen Rationalismus, dessen Falsifikationsprinzip zu einer bedeutenden Wissens- und Entscheidungsgrundlage in allen gesellschaftlichen Feldern geworden ist.

Die Erzeugung von Evidenz dieser Art ist in Bildungskontexten jedoch häufig problematisch. Spiel und Strohmeier (2012) stellen fest, dass nur selten randomisierte Experimental- und Kontrollgruppen oder ausreichend große Stichproben für geeignete statistische Modelle zur Verfügung stehen. Darüber

hinaus ist auf struktureller Ebene, d.h. im System der Lehrkräftebildung, häufig keine Population gegeben, auf die verallgemeinert werden könnte. Dies erschwert die Anwendung probabilistischer Messtheorien und Falsifikationen generell (Bromme, Jäger & Prenzel, 2014). Im Kontext inklusiven Unterrichtens spitzt sich diese Konstellation noch weiter zu. Auf bildungspolitischer Ebene existiert zwar die normative Anforderung inklusionsbezogene Kompetenzen zu vermitteln (KMK & HRK, 2015), doch existieren weder eine einheitliche Definition von Inklusion, noch direkt vergleichbare Ausbildungsstrukturen und -ziele (Greiten, Geber, Gruhn, & Königer, 2018).

Um in dieser methodologisch widersprüchlichen Situation evidenzbasierte Entscheidungen treffen zu können, existieren sowohl konzeptionelle als auch methodische Ansätze. Auf konzeptioneller Ebene wird Evidenz mit dem Begriff der evidenzbasierten Praxis vom Forschungs- auf das Praxissetting ausgeweitet. Dabei wird fallbezogenes Wissen und individuelle Expertise, die aus reflektierten subjektiven Erfahrungen und Routinen besteht, ebenfalls als Evidenz anerkannt (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Dies eröffnet jedoch noch weitestgehend unbeantwortete Fragen, bspw. der Gewichtung von Expertise und wissenschaftlicher Evidenz in Entscheidungsvorgängen. Auf methodischer Ebene können mit bayesianischer Inferenzstatistik die Ergebnisse von Studien aus unterschiedlichen Kontexten verglichen werden (Schum, 2012). Die dafür verfügbaren statistischen Methoden stellen jedoch so hohe Anforderungen an methodische Expertise, dass ihre Anwendung auch außerhalb der Lehrkräftebildung vergleichsweise selten erfolgt (Gopnik & Tenenbaum, 2007).

Probabilistische Zusammenhangsanalysen sind eine zentrale Grundlage der evidenzbasierten Maßnahmenentwicklung, nicht zuletzt, um Prädiktoren von Einstellungen zu inklusivem Unterrichten zu identifizieren und damit Konzepte professioneller Kompetenz von Lehrkräften zu informieren. Die vorliegende Studie stellt daher das Verfahren der messwiederholten Korrelation und die darauf aufbauende Modellklasse der manifesten Wachstumskurvenmodelle vor. Diese Verfahren können mit wenigen Einschränkungen auch mit kleineren Stichproben zuverlässige Parameterschätzungen in der Analyse von Zusammenhangsveränderungen vornehmen und damit die Bildung von Evidenz im Bereich inklusionsbezogener Kompetenzen unterstützen.

9.2 Fragestellung und Hypothesen

Wie könnte Evidenz über die Vermittlung von Diagnosewissen als Basis inklusionsbejahender Einstellungen beschaffen sein? Hierzu kann eine Aussage aus dem ELM und den Kompetenzmodellen der Lehrkräftebildung abgeleitet werden. Selbstgesteuertes Lernen mit wissenschaftlicher Evidenz zu pädagogischer Diagnostik sollte zu mehr selbstgenerierten Informationen und damit einer positiven Einstellungsänderung resultieren. Entsprechend des ELM sollte darüber hinaus ein Gruppensetting das diagnostische Selbstkonzept besonders salient machen und zu einer positiven Einstellungsänderung führen.

Dementsprechend leiten folgende Forschungsfragen die Untersuchung an: Wie stark ist der Zusammenhang zwischen dem Diagnosewissen, der Einstellung zu Inklusion und dem diagnostischen Selbstkonzept bei Lehramtsstudierenden? Wie verändert sich dieser Zusammenhang im Verlauf eines Semesters? Zur Untersuchung dieser Fragestellungen werden die folgenden Hypothesen auf

Grundlage einer Studie zum Training diagnostischer Kompetenz mit einer Form des selbstgesteuerten Lernens (problemorientiertes Lernen) an Textfällen (POL-TF) geprüft:

- Hypothese a. Es besteht ein positiver Zusammenhang zwischen dem Diagnosewissen, der Einstellung zu Inklusion und dem diagnostischen Selbstkonzept.
- Hypothese b. Die Zusammenhänge sind größer bei Personen, die selbstreguliert lernen, als bei Personen, die instruktionsbasiert lernen.
- Hypothese c. Die Zusammenhänge wachsen bei Personen, die selbstreguliert lernen, stärker, als bei Personen, die instruktionsbasiert lernen.

9.3 Methoden

Stichprobe und Studiendesign

Die vorliegende Studie hat ein quasi-experimentelles Design mit zwei Messzeitpunkten. An der Studie nahmen $N_{t1} = 23$, $N_{t2} = 18$ Lehramtsstudierende der beruflichen Fachrichtungen (Interventionsgruppe) und $N_{t1} = 40$, $N_{t2} = 26$ Lehramtsstudierende der Arbeitslehre (Kontrollgruppe) teil, die in jeweils einem von zwei parallel durchgeführten Seminaren zum Thema Diagnostik von Lernvoraussetzungen und -störungen im Wintersemester 2016/2017 teilnahmen. Die Teilnehmenden in der Interventionsgruppe waren 23–44 Jahre alt ($M = 29.5$ Jahre, $SD = 5.1$ Jahre, 78 % weiblich), die Teilnehmenden in der Kontrollgruppe 21–39 Jahre ($M = 26.4$ Jahre, $SD = 4.6$ Jahre, 67 % weiblich). Die Interventions- und Kontrollgruppe unterschieden sich weder signifikant in der Geschlechterverteilung ($\chi^2 = 2.017$, $p = .365$), noch im Alter ($U = 525.500$, $z = -0.599$, $p = .549$).

Interventionsgruppe. Nach fünf ersten Lerneinheiten zu Grundlagen pädagogischer Diagnostik wurde den Studierenden POL vorgestellt. POL ist eine Unterrichtsmethode, die realitätsnahe Problemsituationen für selbstgesteuertes Lernen in Kleingruppen aufbereitet (Barrows, 1985). Ein zentrales Merkmal von POL ist die kooperative Analyse eines Falls in einer festgelegten Schrittfolge. Dazu gehört das Lesen des Falls, die Diskussion fallbezogener Informationen und eigener themenbezogener Wissensbestände sowie die Entscheidung für weiterführende Lernziele, z. B. die Symptome von Dyskalkulie erklären können. Anschließend bearbeiteten alle Gruppenmitglieder in Einzelarbeit alle Lernziele in Auseinandersetzung mit wissenschaftlicher Literatur, um ihre jeweiligen Lernergebnisse anschließend zu diskutieren, zu dokumentieren und zu reflektieren. Zu Beginn der Intervention erhielten die Studierenden zusätzlich zum Fall Strukturierungshilfen (z. B. ein Arbeitsblatt zu POL-TF Schritten), anschließend wiederkehrende Übungsaufgaben (z. B. Hypothesenformulierung) und unterstützende Informationen in Form von Hinweisen zu Literatur sowie Literatur über Lernvoraussetzungen und -störungen. Die sechs verwendeten und von Expertinnen und Experten validierten Textfälle enthalten Informationen zu Schülerinnen und Schülern, z. B. einleitende Narrationen, Notenlisten, Tests und ausgefüllte Beobachtungsbögen. Ausgewählt wurden Lernstörungen mit den höchsten Prävalenzraten unter Schülerinnen und Schülern, z. B. Prüfungsangst oder Lese-Rechtschreibschwäche. Eine genauere Darstellung der Intervention findet sich an anderer Stelle (Wedel et al. 2019b).

Kontrollgruppe. In der Kontrollgruppe besuchten die Studierenden ein thematisch identisches Seminar mit theoretischen Inputs und instruierter Kleingruppen- und Einzelarbeit. In jeweils einer Seminarstunde gab ein theoretischer Input durch den Dozenten einen Überblick über die jeweilige Lernvoraussetzung oder -störung, wie etwa zu Ursache und Symptomen von Dyskalkulie sowie zu evidenzbasierten Interventionen. Im Anschluss wurde das verfügbare Wissen kritisch reflektiert, bspw. in methodischer und praktischer Hinsicht. In der verbleibenden Zeit leitete der Dozent oder eine Gruppe der Teilnehmenden eine Übung zur Anwendung des erworbenen Wissens an, z. B. ein Gruppenpuzzle zu aufmerksamkeitsförderlicher Klassenraumgestaltung. Den Teilnehmenden wurde dieselbe Literatur wie der Interventionsgruppe zur Verfügung gestellt. Die Kontroll- und Interventionsgruppe wurden durch denselben Dozenten angeleitet und ihnen wurden exakt dieselben Lernziele sowohl für das gesamte Seminar als auch die jeweiligen Themen mitgeteilt.

Messinstrumente

Folgende Konstrukte wurden mit einem papierbasierten Fragebogen in dieser Reihenfolge gemessen. Die Einstellung zu inklusivem Unterrichten wurden mit der Selbsteinschätzungsskala EFI-L von Seyfried und Heyl (2016) gemessen, wobei lediglich die Subskala zur Bereitschaft für inklusiven Unterricht ausgewählt wurde, da die Studierenden noch nicht im Lehrberuf eingemündet waren. Die verwendete Subskala EFI-L erfasst die Einstellung mit 5 Items auf einer 6-stufigen Antwortskala von *1 = stimme ganz und gar nicht zu* bis *6 = stimme voll und ganz zu*. Enthalten sind Items zur affektiven Einstellungskomponente, z. B. „Der Unterricht in einer inklusiven Klasse macht mir Angst.“ sowie zur kognitiven Einstellungskomponente, z. B. „Ich kann mir vorstellen, im kommenden Schuljahr in einer inklusiven Klasse zu unterrichten.“ Die Reliabilität betrug für die Interventionsgruppe $\alpha_{t1/t2} = .75/.62$ und für die Kontrollgruppe $\alpha_{t1/t2} = .68/.75$.

Das Selbstkonzept wurde mit der Lehrkräfte-Selbsteinschätzungsskala zum diagnostischen Selbstkonzept (LSD) von Wedel und Müller (2017) gemessen. 15 Items erfassen auf einer Antwortskala von *1 = trifft gar nicht zu* bis *4 = trifft genau zu*, wie die Studierenden ihre Fähigkeiten in pädagogischer Diagnostik einschätzen, z. B. „Ich kann Instrumente zur Erfassung von Lernvoraussetzungen einsetzen (z. B. Beobachtungsbögen, Tests).“ Die Reliabilität der Skala betrug für die Interventionsgruppe $\alpha_{t1/t2} = .94/.91$ und für die Kontrollgruppe $\alpha_{t1/t2} = .91/.91$.

Das Wissen zu pädagogischer Diagnostik wurde mit einem objektiven Wissenstest zum pädagogischen Diagnosewissen (PDW) von Wedel und Müller (2016) erfasst. Der PDW erfasst kognitive Fähigkeiten im Erkennen, Verstehen und Anwenden in den fünf aufeinander folgenden Inhaltsbereichen: Urteilsfehler, Feedback, Ablauf pädagogischer Diagnostik, individuelle Lernvoraussetzungen und Testgütekriterien. In der vorliegenden Studie werden die kognitiven Fähigkeiten im Verstehen analysiert, da diese den Kriterien hoher kognitiver Elaboration im ELM entsprechen. Das Verstehen der zentralen Inhaltbereiche des Diagnosewissens wurden mit 10 Multiple-Choice-Items (2 Items je Inhaltsbereich) erfasst, z. B. „Sie korrigieren einen schriftlichen Test Ihrer Klasse. Die ersten zehn Tests fallen durchgehend schlecht aus. Worauf achten Sie bei der weiteren Korrektur besonders?“. Jedes Item wurde entweder als richtig oder falsch codiert, so dass im Bereich

des Verstehens maximal 10 Punkte erreicht werden konnten (Range 0–10). Da der Test in zwei parallelen Versionen vorliegt, konnten Testwiederholungseffekte ausgeschlossen werden.

Analyseverfahren und messwiederholte Korrelation

Die Hypothesen wurden mit Korrelationsanalysen und Unterschiedstests (*t*-Tests für Korrelations- und Regressionskoeffizienten) geprüft. Eine methodische Herausforderung liegt dabei in der Kombination aus Zusammenhangsanalyse, der vorliegenden Datenstruktur und der kleinen Stichprobe.

Zusammenhangsanalysen (z. B. Korrelationen und Regressionen) basieren auf der Annahme unabhängiger Residuen bzw. Messfehler (Eid, Gollwitzer, & Schmidt, 2011). Mit anderen Worten dürfen zwar die gemessenen Konstrukte miteinander korrelieren, nicht aber die individuellen Abweichungen vom Erwartungswert, die bei der Datenerhebung entstehen. Diese Annahme kann in bestimmten Situationen allerdings nicht getroffen werden, bspw. wenn die Datenerhebung für einen Teil der Stichprobe systematisch durch Störgrößen beeinflusst wird oder wenn von ein- und derselben Person mehrfach Messwerte erhoben werden. Die paarweise Abhängigkeit durch Messwiederholung ist dabei der häufigste Fall und trifft auch für die vorliegende Studie zu.

Für größere Stichproben kann diese Autokorrelation als eigener Term in Modelle integriert und die eigentliche Zusammenhangsmessung davon bereinigt werden, z. B. in Multi-Level-Modellen, latenten Wachstumskurvenmodellen oder Autoregressiven Distributed Lag Modellen (ARDL). Da diese Modelle für kleine Stichproben nicht identifizierbar sind, braucht es andere Strategien, um mit einer Verletzung der Unabhängigkeit von Residuen umzugehen.

Nicht in der Standardliteratur behandelt ist das speziell für Einschränkungen in der Unabhängigkeit der Residuen entwickelte messwiederholte Korrelationsmaß r_{rm} (Bakdash & Marusich, 2017; Bland & Altman 1995). Da die Darstellung von r_{rm} in Bland und Altman (1995) nicht ausformuliert ist und in Bakdash und Marusich (2017) nachkorrigiert wurde, erfolgt an dieser Stelle die vollständige Fassung (siehe auch: Bakdash & Marusich, 2019).

Die messwiederholte Korrelation r_{rm} modelliert den Zusammenhang von Konstrukten als einfaktorielle Kovarianzanalyse (ANCOVA). Während in der Varianzanalyse üblicherweise Unterschiede zwischen Personengruppen getestet werden, sieht r_{rm} vor, dass jede Person eine ‚Gruppe‘ bzw. einen Faktor (*j*) bildet und die paarweise abhängigen Messwerte einer Person eines Konstrukts die ‚Gruppenmitglieder‘ bzw. Faktorstufen (*i*). Das zweite Konstrukt ist die Kovariate. Um einen für alle Personen geltenden Zusammenhang zu schätzen, wird eine, für *alle* Messwiederholungspaare am besten passende Beziehung zwischen den zwei korrelierenden Konstrukten geschätzt, d. h. eine Schar von Regressionsgeraden mit gleichem Anstieg, aber unterschiedlichem Schnittpunkt mit der *y*-Achse.

Mit anderen Worten wird die Varianz zwischen den Personen genutzt, um den durch die autokorrelative Datenstruktur veränderten Zusammenhang zu korrigieren. Es sei:

$$Y_{ij} = \mu + \tau_j + \beta(X_{ij} - \bar{X}) + \varepsilon_{ij} \quad (1)$$

mit

Y_{ij} ... der Messwert der abhängigen bzw. ersten Variable für den i -ten Messwert im j -ten Faktor, mit $j = 1 \dots p$,
 $i = 1 \dots q$

μ ... der Gesamtmittelwert $\bar{Y}_{..}$

τ_j ... der Effekt des j Faktors (Person)

β ... der Anstieg der Kovariate bzw. der zweiten Variable

X_{ij} ... der Messwert der Kovariate für den i -ten Messwert im j -ten Faktor

$\bar{X}$... der Mittelwert aller Werte X_{ij}

ε_{ij} ... der Messfehler

berechnet durch

$$\mu = \frac{\sum_{i=1}^q \sum_{j=1}^p (Y_{ij})}{qp}, \text{ mit } qp \dots \text{Anzahl der gemessenen Werte in } Y_{ij}$$

$$\tau_j = \sum_{i=1}^p (\bar{Y}_{.j} - \mu), \text{ mit } \bar{Y}_{.j} \dots \text{der Mittelwert des } j\text{-ten Faktors}$$

β = Schätzung durch Ordinary Least Squares (OLS) für alle Messwertpaare

$$\varepsilon_{ij} = \sum_{i=1}^q \sum_{j=1}^p (Y_{ij} - \bar{Y}_{.j})$$

Der Korrelationskoeffizient r_{rm} wird anschließend durch die Varianzanteile der Kovariate und der Residuen berechnet.

$$r_{rm} = \sqrt{\frac{SAQ \text{ der Kovariate}}{SAQ \text{ der Kovariate} + SAQ \text{ der Residuen}}} \quad (2)$$

mit

SAQ ... Summe der Abweichungsquadrate bzw. „sum of squares“ (SS)

berechnet durch

klassische Schritte der Kovarianzanalyse (Rasch & Schott, 2018: 495), die auch in statistischer Standardsoftware implementiert sind.

Das Vorzeichen des Anstiegs β der Kovariate bestimmt dabei das Vorzeichen des Korrelationskoeffizienten. Der Parameter r_{rm} bewegt sich zwischen -1 und $+1$ und ist analog zu Pearson's r zu interpretieren. Bei -1 liegt ein perfekter negativer Zusammenhang zwischen zwei Konstrukten vor, bei $+1$ ein perfekter positiver Zusammenhang und bei 0 kein Zusammenhang. Darüber hinaus zeigt der Anstieg β auch die Veränderung des Zusammenhangs für die jeweiligen Personen an.

Berechnungsschritte für r_{rm} in Statistiksoftware

Die Berechnung von r_{rm} kann sowohl in SPSS als auch in R erfolgen (hier dargestellt mit SPSS 25 und R 3.5.1). Für beide Programme müssen die Daten im Langformat vorliegen, d. h. die Messwertpaare der Konstrukte bekommen für jede Messwiederholung eine eigene Datenzeile, die um

die personenidentifizierende Variable, z. B. ein anonymisierter Code, ergänzt wird. Insgesamt sollten dementsprechend drei Datenspalten vorliegen (personenidentifizierende Variable, erste Variable [aV], zweite Variable [Kovariate]). Für jede Person entspricht die Anzahl der Datenzeilen dann der Anzahl der Messwiederholungen, wobei die personenidentifizierende Variable in jeder Datenzeile wiederholt wird.

In SPSS:

1. Analysieren > Allgemeines Lineares Modell > Univariat
2. Erste Variable als abhängige Variable eingeben, zweite Variable als Kovariate und personenidentifizierende Variable als festen Faktor. OK klicken.
3. Im Output der Tabelle „Test auf Zwischensubjekteffekte“ in der Spalte „Quadratsumme vom Typ III“ die Werte der Kovariate und Fehler entnehmen.
4. In Gleichung 2 einsetzen und Vorzeichen anpassen.

In R:

1. Paket *rmcorr* (Bakdash & Marusich, 2017) laden und Datensatz einlesen
2. Funktion *rmcorr* aufrufen (siehe Hilfedatei des Pakets) und Datensatz sowie Variablenamen einfügen. Ausführen.

9.4 Ergebnisse

Deskriptiv gaben die Lehramtsstudierenden eine eher positive Einstellung zu Inklusion an und hatten mittlere bis eher geringe kognitive Fähigkeiten in Bereich Verstehen des Diagnosewissens (siehe Tabelle 8).

Mann-Whitney-U-Tests ergaben, dass sich die Interventions- und Kontrollgruppe zu t_1 in den Einstellungen zu inklusivem Unterrichten unterschieden (EFI-L, $z = 2.084$, $p = 0.037$), nicht aber im Diagnosewissen (PDW-V, $z = 0.497$, $p = 0.497$) oder im diagnostischen Selbstkonzept (LSD, $z = -0.858$, $p = 0.391$). Zu t_2 unterschieden sich die Gruppen in Bezug auf EFI-L ($z = 2.463$, $p = 0.014$) und auf das PDW-V ($z = 3.718$, $p < 0.001$), aber nicht in Bezug auf LSD ($z = 0.778$, $p = 0.737$).

Tabelle 8

Mittelwerte und Standardabweichungen (in Klammern) für alle Teilnehmenden sowie Interventions- und Kontrollgruppe

	Alle Teilnehmenden		Interventionsgruppe		Kontrollgruppe	
	t_1	t_2	t_1	t_2	t_1	t_2
	$n = 63$	$n = 44$	$n = 23$	$n = 18$	$n = 40$	$n = 26$
EFI-L	3.35 (0.91)	3.38 (0.77)	3.69 (0.99)	3.73 (0.67)	3.15 (0.81)	3.14 (0.75)
PDW-V	3.86 (1.52)	4.28 (1.83)	4.04 (1.52)	5.44 (1.15)	3.75 (1.53)	3.44 (1.78)
LSD	2.59 (0.59)	2.87 (0.42)	2.49 (0.62)	2.92 (0.43)	2.65 (0.57)	2.83 (0.42)

Anmerkungen. t_1 = Erster Messzeitpunkt, t_2 = Zweiter Messzeitpunkt, EFI-L = Einstellungen zur Bereitschaft inklusiven Unterrichtens Skala 1–6, PDW-V = Wissen zu pädagogischer Diagnostik Verstehen Skala 1–10, LSD = Lehrkräfte Selbsteinschätzung zum diagnostischen Selbstkonzept Skala 1–4.

Zusammenhangsanalysen

Hypothese a betraf die Annahme eines positiven Zusammenhangs zwischen Einstellungen zu inklusivem Unterrichten, Diagnosewissen und dem diagnostischen Selbstkonzept. Die Prüfung erfolgte mit einer messwiederholten Korrelation r_{rm} in der Interventions- und Kontrollgruppe.

Die z. T. beträchtlichen Autokorrelationen zwischen den zwei Messzeitpunkten betrugen in der Kontrollgruppe EFI-L $r = .455$, $p = .044$; PDW-V $r = .270$, $p = .236$; LSD $r = .209$, $p = .376$ und in der Interventionsgruppe EFI-L $r = .288$, $p = .318$; PDW-V $r = -.006$, $p = .982$; LSD $r = .552$, $p = .041$. Dies ist eine typische Eigenschaft von reflektiven Konstrukten, die ohne Berücksichtigung im Schätzverfahren zu größeren Ungenauigkeiten führt. Die folgende Tabelle stellt die messwiederholte Korrelation, die diese autokorrelative Struktur berücksichtigt, möglichen Pearson Korrelationen, die diese Datenstruktur nicht berücksichtigt, zur Veranschaulichung gegenüber (siehe Tabelle 9).

Tabelle 9

Korrelationen in Pearsons r zwischen Einstellungen, Wissen und Selbstkonzept und zugehörige p -Werte (in Klammern)

		Kontrollgruppe		Interventionsgruppe	
		EFI-L	PDW-V	EFI-L	PDW-V
PDW-V	r_{ignore}	.167 (.218)		.076 (.658)	
	$r_{mean\ between}$	.171 (.312)		.171 (.321)	
	$r_{mean\ within}$	.170 (.294)		.003 (.989)	
	r_{rm}	.202 (.436)		.262 (.366)	
LSD	r_{ignore}	.391 (.001)	-.079 (.565)	.471 (.002)	.081 (.641)
	$r_{mean\ between}$	.398 (.002)	-.049 (.812)	.477 (.029)	-.099 (.696)
	$r_{mean\ within}$	.253 (.115)	-.166 (.313)	.563 (.005)	-.252 (.259)
	r_{rm}	.667 (.001)	.313 (.221)	-.126 (.652)	.505 (.065)

Anmerkungen. EFI-L = Einstellungen zur Bereitschaft inklusiven Unterrichtens Skala 1–6, PDW-V = Wissen zu pädagogischer Diagnostik Verstehen Skala 1–10, LSD = Lehrkräfte Selbsteinschätzung zum diagnostischen Selbstkonzept Skala 1–4.

Dafür wurden folgende Pearson Korrelationen gebildet, die sich auch in der Forschungsliteratur zu Interventionen für Einstellungsänderungen im Bereich schulischer Inklusion finden (z. B. Batsiou et al. 2008), jedoch zu fehlerhaften Ergebnissen führen können: Die Korrelation (r_{ignore}) behandelt jeden Datenpunkt als ein vollständiges Element der Population, d.h. es wird ignoriert, dass mehrere Datenpunkte von derselben Person stammen und daher autokorreliert sind. Die Korrelation (r_{mean_within}) behandelt die Mittelwerte der messwiederholten Datenpunkte jeder Person als vollständiges Element der Population, d. h. es werden für jede Person für jedes Konstrukt zuerst ein Mittelwert aller für diese Person verfügbaren Datenpunkte gebildet und anschließend diese Mittelwerte korreliert. Für die Korrelation ($r_{mean_between}$) werden zuerst Korrelationen der Konstrukte innerhalb jedes Messzeitpunktes berechnet und dann deren Mittelwert gebildet.

Eine besonders deutliche Abweichung besteht zwischen r_{ignore} und r_{rm} . Trägt man beide grafisch in einem Streudiagramm ab (siehe Abbildung 17), fällt insbesondere der Zusammenhang zwischen dem diagnostischen Selbstkonzept und der Einstellung zu Inklusion in der Interventionsgruppe sowie der Zusammenhang zwischen dem Diagnosewissen und dem diagnostischen Selbstkonzept in der Kontrollgruppe auf. Beide Zusammenhänge zeigen ein Simpson Paradoxon (Kievit, Frankenhuys, Waldorp & Borsboom, 2013), d. h. auf Populations- und auf Individualebene bestehen gegenläufige Zusammenhänge. In der Interventionsgruppe zeigte sich etwa, dass auf Ebene der Gesamtgruppe ein positiver Zusammenhang zwischen Einstellung und Selbstkonzept besteht. Korrekt auf Individualebene gemessen war dieser Zusammenhang jedoch negativ (siehe Abbildung 17).

In der Kontrollgruppe gab es einen mittleren, aber nicht signifikanten positiven Zusammenhang zwischen dem Diagnosewissen und den Einstellungen zu inklusivem Unterrichten ($r_{\text{rm}} = .202, p = .436$). Dagegen gab es einen großen, signifikanten positiven Zusammenhang der Einstellungen zu inklusivem Unterrichten mit dem diagnostischen Selbstkonzept ($r_{\text{rm}} = .667, p = .001$). Diagnosewissen und diagnostisches Selbstkonzept hatten einen mittleren positiven, aber nicht signifikanten Zusammenhang ($r_{\text{rm}} = .311, p = .221$).

In der Interventionsgruppe war der Zusammenhang zwischen dem Diagnosewissen und den Einstellungen zu inklusivem Unterrichten positiv und von mittlerer Größe, aber nicht signifikant ($r_{\text{rm}} = .262, p = .366$). Ein negativer, aber nicht signifikanter Zusammenhang bestand zwischen der Einstellung zu Inklusion und dem diagnostischen Selbstkonzept ($r_{\text{rm}} = -.126, p = .652$). Dagegen zeigte sich ein großer und marginal signifikanter positiver Zusammenhang zwischen dem diagnostischen Selbstkonzept und dem Diagnosewissen ($r_{\text{rm}} = .505, p = .065$).

Insgesamt kann Hypothese a trotz der positiven mittleren Effektgrößen nur für den Zusammenhang von Einstellungen zu inklusivem Unterrichten und dem diagnostischen Selbstkonzept für die Kontrollgruppe und für den Zusammenhang von Diagnosewissen und diagnostischem Selbstkonzept in der Interventionsgruppe bestätigt werden.

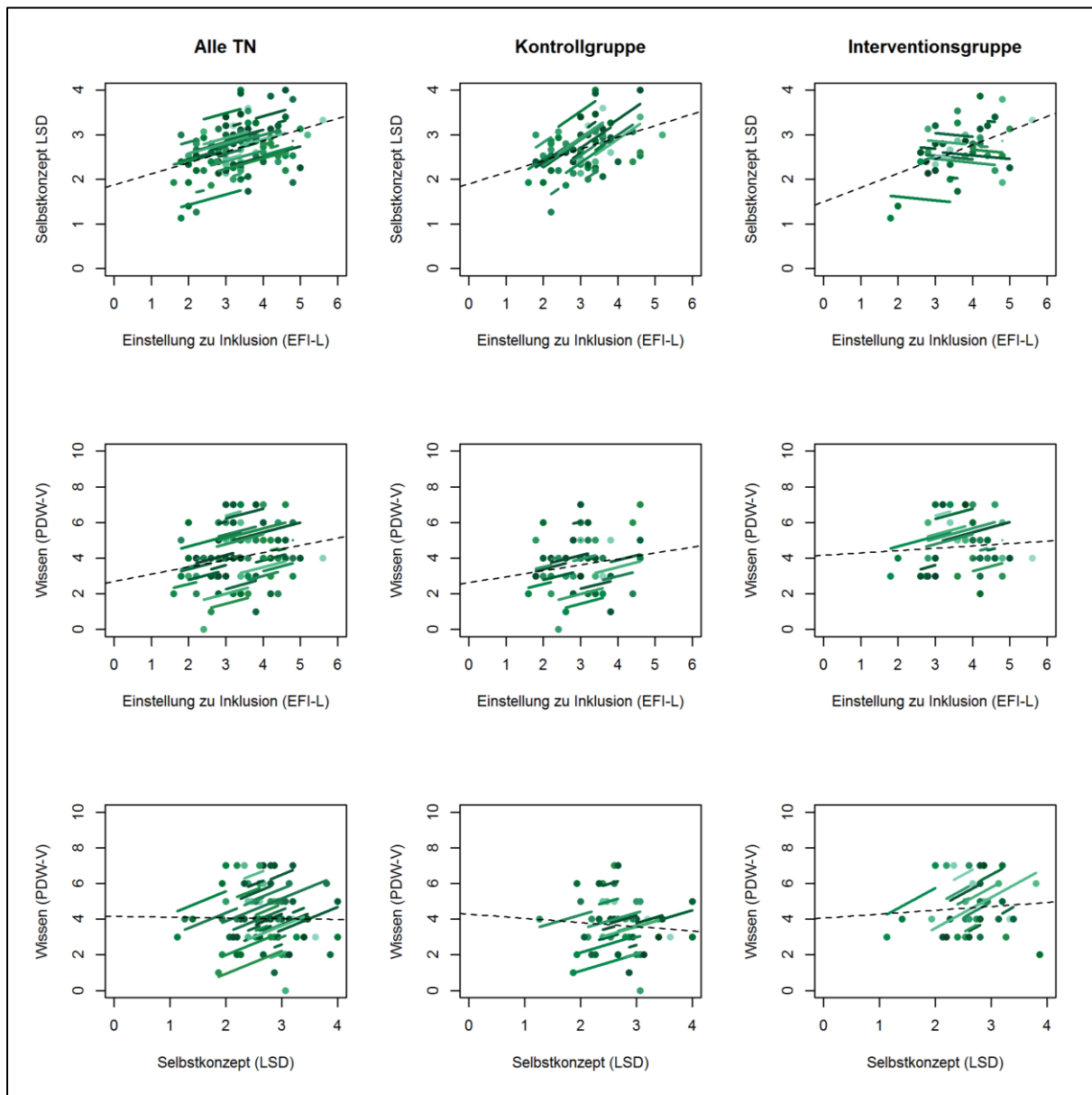


Abbildung 17. Streudiagramme für Einstellung zu Inklusion, Diagnosewissen und diagnostisches Selbstkonzept; in den Streudiagrammen stellt jeder Punkt einen Messwert einer Person zu einem Messzeitpunkt dar; die Messwerte, d. h. Personen sind durch Grünstufen unterschieden; Pearson-Korrelationen ohne Berücksichtigung der Datenstruktur sind mit einer gestrichelten Linie eingezeichnete Funktionsschaar der messwiederholten Korrelation ist mit durchgezogenen Linien eingezeichnet.

Unterschiedstests und Zusammenhangsveränderungen

Hypothese b trifft die Aussage, dass die Zusammenhänge in der Interventionsgruppe größer ausfallen, als in der Kontrollgruppe. Zur Prüfung wurden z-Tests für Korrelationsunterschiede zwischen Kontroll- und Interventionsgruppe berechnet (Eid, Gollwitzer, & Schmidt, 2011, 547).

Im Korrelationsmaß r_{rm} bestand kein Unterschied zwischen den Gruppen für den Zusammenhang von PDW-V und EFI-L ($z = -0.229$, $p = 0.410$), aber ein signifikanter Unterschied für den Zusammenhang von LSD und EFI-L ($z = -3.358$, $p < 0.001$). Für den Zusammenhang von PDW-V und LSD bestand kein Unterschied zwischen den Gruppen ($z = -0.836$, $p = 0.201$).

Insgesamt kann Hypothese b daher nur für den Zusammenhang von Einstellung zu Inklusion mit dem diagnostischen Selbstkonzept beibehalten werden.

Hypothese c trifft die Aussage, dass die Zusammenhänge in Interventionsgruppe stärker wachsen, als in der Kontrollgruppe. Zur Prüfung wurden Unterschiede zwischen der Interventions- und Kontrollgruppe in der Veränderung der Korrelationsmaße durch t-Tests für den Anstieg der Kovariate, d. h. für den β -Koeffizienten berechnet (Cohen, Cohen, West, & Aiken, 2003). Damit können Aussagen über Unterschiede in den Veränderungen von Zusammenhängen getroffen werden, was üblicherweise nur mit latenten Wachstumskurvenmodellen und entsprechend umfangreichen Stichproben möglich ist. In Simulationsstudien liegt die für eine ausreichende Teststärke benötigte Stichprobengröße für Wachstumskurvenmodelle bei 200 bis 500 Personen mit 4 bis 6 Messzeitpunkten (Hamilton & Hancock, 2003) und für ANCOVAs bei 30 bis 60 Personen (Seaman, Algina, & Oljenik 1985). Trotz dieses großen Vorteils hat das Verfahren des β -Vergleichs aber auch typische Einschränkungen manifester Modellspezifikation, bspw. werden Messfehler und Methodenfaktoren nicht explizit geschätzt.

Zwischen den Gruppen bestand ein Unterschied in der Veränderung im Zusammenhang von EFI-L und LSD ($t(67) = 2.104$, $p = 0.039$). In der Interventionsgruppe sank dieser Zusammenhang im Verlauf des Semesters ab ($\beta = -.08$), während er in der Kontrollgruppe anstieg ($\beta = .57$) (siehe Abbildung 1). Für die Veränderung des Zusammenhangs von PDW-V und EFI-L sowie von PDW-V und LSD bestand dagegen kein Unterschied zwischen den Gruppen ($\beta_{\text{Interventionsgruppe}} = .57$, $\beta_{\text{Kontrollgruppe}} = .53$, $t(57) = 0.046$, $p = 0.964$; $\beta_{\text{Interventionsgruppe}} = 1.72$, $\beta_{\text{Kontrollgruppe}} = .93$, $t(57) = 0.398$, $p = 0.692$).

Hypothese c kann daher nur für den Zusammenhang von Einstellung zu inklusivem Unterrichten und diagnostischem Selbstkonzept beibehalten werden.

9.5 Diskussion

Zusammenhänge von Diagnosewissen, Einstellung zu inklusivem Unterrichten und diagnostischem Selbstkonzept

Diagnosewissen und Einstellung zu inklusivem Unterrichten. In beiden Untersuchungsgruppen konnte ein mittlerer positiver, aber nicht signifikanter Zusammenhang festgestellt werden. Die Richtung dieses Zusammenhangs entspricht den Modellen professioneller Lehrkompetenz. Dabei bestand weder ein Gruppenunterschied in der Stärke des Zusammenhangs noch in dessen Anstieg im

Semesterverlauf. Die Ergebnisse entsprechen also nur eingeschränkt den Aussagen des ELM. Es kann zudem entgegen empirischen Ergebnissen des ELM zu selbstgesteuertem Lernen (Darke et al., 1998; Wenden, 1998) nicht davon ausgegangen werden, dass POL eine wirksamere Methode zur stärkeren Verknüpfung von Diagnosewissen und Einstellungen zu inklusivem Unterrichten darstellt als instruktionsbasiertes Lernen. Eine Erklärung dafür liegt in der Verfügbarkeit starker Argumente in Form wissenschaftlicher Evidenz in beiden Untersuchungsgruppen. Möglicherweise überschattete dies den Effekt des selbstgesteuerten Lernens.

Diagnostisches Selbstkonzept und Einstellung zu inklusivem Unterrichten. Während in der Kontrollgruppe ein starker signifikant positiver Zusammenhang bestand, konnte in der Interventionsgruppe ein schwacher negativer Zusammenhang festgestellt werden. Zwischen den Gruppen war sowohl dieser Unterschied signifikant als auch dessen Veränderung. Dabei wuchs der Zusammenhang in der Kontrollgruppe, während er in der Interventionsgruppe schrumpfte. Besonders letzteres Ergebnis widerspricht der Aussage der ‚functional matching‘ Hypothese des ELM. Eine mögliche Erklärung dafür ist die Zusammenarbeit in den POL-Kleingruppen. Die Lehramtsstudierenden waren mit der Methode nicht vertraut und zeigten anfänglich Schwierigkeiten in der Festlegung gemeinsamer Lernziele und der Strukturierung von Diskussionen. Wie Brownell und andere (2006) anmerken, braucht gelingende Kooperation zwischen Lehrkräften kontextspezifische Kommunikationsfähigkeiten, da sich pädagogische und diagnostische Handlungs- und Entscheidungsstrategien zwischen angehenden Lehrkräften unterscheiden können. Es ist plausibel anzunehmen, dass Diskussionen über Maßnahmen im Bereich schulischer Inklusion Konfliktpotenzial in sich tragen, was die wahrgenommene Gruppenzugehörigkeit verringern kann.

Diagnostisches Selbstkonzept und Diagnosewissen. In beiden Gruppen bestand ein theoriekonformer mittlerer bis großer positiver Zusammenhang, der jedoch nur in der Interventionsgruppe signifikant wurde. Dabei bestand kein Unterschied zwischen den Gruppen und ebenfalls kein Unterschied bzgl. der Zusammenhangsveränderung. Auch hier kann also nicht von einem differenziellen Effekt von POL ausgegangen werden.

Evidenz zu Diagnosewissen und dem Selbstkonzept als Basis der Einstellung zu inklusivem Unterrichten

Die evidenzbasierte Konstruktion von Lerngelegenheiten zu inklusivem Unterrichten stellt ein wichtiges Lernfeld für die Lehrkräftebildung dar. Die vorliegende Interventionsstudie trägt zwei zentrale Ergebnisse zu dieser Diskussion bei.

Die bisherigen Forschungsbefunde zum querschnittlichen Zusammenhang von Diagnosewissen und Einstellung zu inklusivem Unterrichten entsprachen auf deskriptiver Ebene den Ergebnissen der vorliegenden Studie. Während in bisherigen Studien oft auf Selbsteinschätzungen zum Wissen zurückgegriffen wurde (z. B. Sermier Dessemontet et al., 2011), zeigt die vorliegende Studie ähnliche deskriptive Zusammenhänge auch mit einem objektiven Wissenstest und einer stringenteren statistischen Analyse. Darüber hinaus konnte auf deskriptiver Ebene ein wachsender Zusammenhang dieser Konstrukte im Lernverlauf festgestellt werden. Dieser Befund unterstützt die konzeptionelle

Annahme der Modelle professioneller Lehrkompetenz, dass Diagnosewissen eine Grundlage der Einstellung zu inklusivem Unterrichten bildet. Dieses Ergebnis bestand unabhängig von dem Format, in dem Lehramtsstudierende Wissen zu pädagogischer Diagnostik erwerben.

Demgegenüber waren der Zusammenhang von Selbstkonzept und Einstellungen zu inklusivem Unterrichten sowie dessen Veränderung vom Lernformat abhängig. Der schrumpfende Zusammenhang in der Gruppe mit selbstgesteuertem Lernen zeigt exemplarisch, dass Lerngelegenheiten mit dem Ziel der Vermittlung von Diagnosewissen nicht in jedem Fall zu einer Verknüpfung von diagnostischem Selbstkonzept und Einstellungen zu inklusivem Unterrichten führen. Möglicherweise braucht es dazu besondere Instruktions- oder Reflexionsformen, die in den untersuchten Lerngelegenheiten nicht angeboten wurden. Ein Hinweis darauf ist etwa, dass eigene Vorerfahrungen mit inklusiven Kontexten zu den stärksten Einstellungsprädiktoren gehören (de Boer, Pijl, & Minnaert, 2011). Zu diesen Erfahrungen und ihrer Reflexion bieten etwa die Praxisanteile der Lehramtsausbildung ein großes Potenzial.

Studien zum Übergang angehender Lehrkräfte in die schulpraktische Arbeitswelt zeigen aber auch, dass sich häufig eine starke Orientierung an vorgefundenen Unterrichtsskripts, pädagogischen Praktiken und schulinternen Normen einstellt (Hargreaves, 2000; Santagata & Yeh, 2016). Diese Studien stellen für das Thema der Evidenzbasierung infrage, welche Wirksamkeitskriterien in der Entwicklung von Lerngelegenheiten angelegt werden. Im Hinblick auf diese Instabilität erworbener Kompetenzen besteht dringender Bedarf an wissenschaftlicher Evidenz, die über Angebots-Nutzen-Modelle für den Zeitraum der Kompetenzvermittlung in der ersten Phase der Lehrkräftebildung hinausgeht.

Messwiederholte Korrelation und manifeste Wachstumskurvenmodelle

Die Berechnung von Zusammenhangsveränderung, d.h. des Unterschieds oder Zusammenhangs von Zusammenhängen über die Zeit oder zwischen Gruppen stellt mit klassischen probabilistischen Modellen hohe Anforderungen an die Stichprobe und mit bayesianischen Modellen hohe Anforderungen an methodische Kenntnisse. Die messwiederholte Korrelation und das manifeste Wachstumskurvenmodell bieten einen Kompromiss an, der wünschenswerte Eigenschaften beider Modelle vereint und für die Anwendung in der Lehrkräftebildungsforschung hinnehmbare Einschränkungen aufweist. Die Berechnung und Prüfung des Wachsens oder Schrumpfens von Zusammenhängen in kleinen Stichproben eröffnet der Lehrkräftebildung die Untersuchung von Fragestellungen, die unter den strukturellen Bedingungen des Bildungssystems für angehende Lehrkräfte anderweitig nicht an jedem Standpunkt der Lehrkräftebildung möglich wäre. Da bisher nur wenige Untersuchungen im Bildungsbereich mit dieser Messgröße vorliegen, können Einschätzungen zur Effektgröße nur durch den Vergleich mit verwandten Korrelationsmaßen getroffen werden. Da die messwiederholte Korrelation in anderen Wissenschaftsbereichen (z. B. Medizin, Biologie) bereits zunehmend genutzt wird, ist dies jedoch auch in der empirischen Bildungsforschung absehbar.

Grenzen der Studie

Die Zusammenhangsanalysen betreffen einen sehr spezifischen Kontext, aus dem nicht ohne weiteres auf andere Lerngelegenheiten in der Lehrkräftebildung geschlossen werden kann. Wünschenswert wären weitere Stichproben, in denen Diagnosewissen mit anderen didaktischen Methoden vermittelt wird oder aber mit derselben Methode in einem anderen institutionellen Kontext. Es ist nicht auszuschließen, dass die ermittelten Zusammenhänge charakteristisch für die spezifische Stichprobe sind.

Die geringe Stichprobengröße bedeutet für viele der gezeigten Zusammenhänge eine Unsicherheit in Bezug auf ihre Signifikanz. Die berechneten Zusammenhänge fallen deskriptiv durchgängig in die theoretisch zu erwartende Richtung aus, sie können jedoch nicht mit hinreichender Sicherheit angenommen werden. Da die Effektgrößen jedoch z. T. als sehr groß eingestuft werden können, sollte der Signifikanzwert nicht als einziges Kriterium in die Interpretation einbezogen werden.

Darüber hinaus konnten die Teilnehmenden aus studienorganisatorischen Gründen nicht randomisiert zur Interventions- und Kontrollgruppe zugeordnet werden. Diese Limitation wirkt jedoch weniger gewichtig, da die Gruppen in den gemessenen Konstrukten zum ersten Messzeitpunkt eine gute Vergleichbarkeit zeigten.

Dass die Einstellungsskala nur akzeptable Reliabilitäten aufwies, ist möglicherweise auf die kleine Stichprobengröße im Verhältnis zur Anzahl der Items zurückzuführen. Da die Skala umfangreich validiert wurde, kann jedoch von ihrer prinzipiellen Zuverlässigkeit ausgegangen werden. Darüber hinaus wurde die Einstellung zu inklusivem Unterrichten vor dem Wissenstest abgefragt, um eine Konfundierung auszuschließen. Mit dem diagnostischen Selbstkonzept wurde nur eine Facette des professionellen Selbstkonzepts von Lehramtsstudierenden und Lehrkräften gemessen, die jedoch konzeptionell in einem wichtigen Zusammenhang mit dem Diagnosewissen steht. Vor dem Hintergrund der bisherigen Forschung, in der überwiegend nur eines dieser Konstrukte gemessen wird, stellt dies jedoch einen Erkenntnisgewinn dar.

Da keine weiteren Messzeitpunkte vorliegen, bleibt auch offen, inwiefern sich der Zusammenhang zwischen Diagnosewissen und Einstellung zu inklusivem Unterrichten durch Praxiserfahrungen im Schulkontext verändert. Denkbar wäre sowohl ein Anstieg des Zusammenhangs in Anlehnung an die Kontakthypothese (Allport, 1954), als auch ein Abfall durch den sogenannten ‚Praxischock‘.

Ausblick

Die Förderung inklusionsbejahender Einstellungen ist eine zentrale Voraussetzung für lernwirksames Unterrichten in heterogenen Klassen. Aktuelle Modelle professioneller Lehrkompetenzen konzipieren diagnostische Kompetenz als eine Voraussetzung für erfolgreiches Unterrichtshandeln in inklusiven Kontexten. Die vorliegende Studie zeigt z. T. starke positive, aber nicht signifikante Zusammenhänge mit Konstrukten aus dem Bereich pädagogischer Diagnostik. Im Bezug auf das Lernformat kann daher nur vorsichtig interpretiert werden, dass selbstgesteuertes Lernen dabei hilft, diesen Zusammenhang zu steigern. Daran anknüpfend stellt sich die generelle Frage

zum Effekt der Kooperation von Lehrkräften im Kontext schulischer Inklusion. Meta-Analysen zu Effekten von Kooperation unter Lehrkräften im allgemeinen Schulbetrieb deuten auf eine Reduktion von Stress und Resignation hin (Vangrieken, Meredith, Packer, & Kyndt, 2017). In Anbetracht der Befürchtungen vieler Lehrkräfte bzgl. inklusiven Unterrichtens könnte der positive Effekt von Kooperation möglicherweise bereits in der Lehrkräftebildung aufgebaut werden. Denkbar ist etwa die Implementation didaktischer Modelle zur Zusammenarbeit von Lehramtsstudierenden mit an inklusionsorientierten Schulen tätigen Lehrkräften. Weitere empirische Forschung im Kontext inklusiven Unterrichtens bietet das Potenzial zu einer grundlegend evidenzbasierten Professionalisierung der Lehrkräftebildung.

10 Studie 4: Welchen Effekt haben Intelligenz, Empathie, Textverstehen und Textmerkmale auf das spontane diagnostische Urteil von Lehramtsstudierenden?⁴

10.1 Einleitung und theoretischer Hintergrund

Lehramtsstudierende sind im Studium mehr denn je gefordert, sich auf inklusiven, differenzierenden Unterricht vorzubereiten, u. a. durch den Erwerb diagnostischer Kompetenz. Interventionsstudien in der ersten Phase der Lehrkräftebildung kommen übergreifend zu dem Schluss, dass sich insbesondere fallbasiertes Lernen zum Erwerb praxisnaher Wissensbestände und Fähigkeiten eignet (Kramer, König, Kaiser, Ligtvoet, & Blömeke, 2017). Für fallbasiertes Lernen pädagogischer Diagnostik werden sowohl Video- als auch Textfälle genutzt, aber Forschung zur Bedeutung von Textmerkmalen für den kognitiven Umgang Studierender mit Textfällen existiert bislang kaum (Syring et al., 2015). Dieser Erkenntnismangel ist jedoch sowohl praxis- als auch forschungsrelevant. Nicht nur ist der Aufwand zur Konstruktion, Implementation und Nachnutzung von Textfällen erheblich geringer als bei Videofällen, auch nutzen zahlreiche (quasi-)experimentelle Studien Textfälle in der Interventions- oder Kontrollgruppe (Kramer et al., 2017; Syring et al., 2015). Welche Bedeutung haben die konkrete Fallgestaltung und die kognitiven Grundfähigkeiten Studierender für den Umgang mit diesen Textfällen? In welcher Relation stehen beide Faktoren zueinander?

Die vorliegende Experimentalstudie untersucht die Effekte von Intelligenz, Empathie, Textverstehen und Textmerkmalen auf einen typischen Prozess der diagnostischen Fallarbeit: die spontane Urteilsbildung.

Diagnostische Urteilsbildung von Lehrkräften

Diagnostik im Schulkontext umfasst die systematische Informationssammlung und -integration der Lern- und Leistungsmerkmale von Schülerinnen und Schülern (Schrader, 2009). Diagnostik zielt u. a. auf die Erklärung von Beeinträchtigungen des Lernverhaltens (Klug, Bruder, & Schmitz, 2016), die Einteilung von Lerngruppen und die Steuerung von Bildungsverläufen durch Selektionsentscheidungen (Ingenkamp & Lissmann, 2008). Schülerinnen und Schüler akkurat zu beurteilen ist von grundlegender Bedeutung für die Unterstützung ihrer Lern- und Entwicklungsprozesse. Akkurate Urteile haben etwa positive Effekte auf das akademische Selbstkonzept (Urhahne, Chao, Florineth, Luttenberger, & Paechter, 2011), die Lern- und Leistungsmotivation (Urhahne, 2015) und die schulische Leistung (Rubie-Davies, Hattie & Hamilton, 2008).

Urteile sind Prozessmodellen pädagogischer Diagnostik zufolge insbesondere an zwei Punkten im Verlauf einer Diagnostik zentral (Klug, Bruder, Kelava, Spiel, & Schmitz, 2013; Herppich et al., 2018). Zum einen sind spontane Urteile aus eher unsystematischen Beobachtungen ein Anlass für eine

⁴ Diese empirische Studie ist zur Begutachtung in der Zeitschrift Psychologie in Erziehung und Unterricht eingereicht. Sie wurde auf der paEpsy 2019, Universität Leipzig, präsentiert. In der Zeitschrift werden keine exakten p -Werte berichtet. Signifikanzn werden entsprechend der spezifischen Vorgaben mit $p > .05$, $p < .05$, $p < .01$ und $p < .001$ angegeben.

systematische Informationssammlung. Sie münden in eine vorläufige Diagnosestellung bzw. Arbeitshypothese. Zum anderen erfolgt eine weitere Urteilsbildung nach empirischer Prüfung einer oder mehrerer Arbeitshypothesen. Dieses Urteil bildet wiederum die Grundlage für weitere diagnostische und pädagogische Entscheidungen, etwa für psychologische Differenzialdiagnosen oder Fördermaßnahmen.

Akkurate Urteile sind zum einen von kognitiven Fähigkeiten und zum anderen von Merkmalen der Beurteilungssituation abhängig (Ingenkamp & Lissmann, 2008). Für Lehrkräfte zeigte sich beispielsweise, dass eine akkurate Einschätzung leichter fällt, wenn sie viel spezifisches Diagnosewissen haben (Klug et al., 2016). Metaanalysen stellten jedoch insgesamt fest, dass Effekte von Lehrkräftemerkmalen auf die Urteilsgenauigkeit nach wie vor zu wenig untersucht wurden (Machts, Kaiser, Schmidt, & Möller, 2016; Südkamp et al., 2012). Für Lehramtsstudierende ist die Studienlage zum Effekt kognitiver Fähigkeiten auf die Urteilsgenauigkeit noch weitaus schmal und z. T. widersprüchlich. Einige Studien zeigten einen positiven Effekt von Diagnosewissen auf die Urteilsgenauigkeit in text- und videobasierten Szenariotests (Klug et al., 2016; Trittel, Bruder, & Schmitz, 2011; Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011). Kaiser, Helm, Retelsdorf, Südkamp und Möller (2012) zeigten darüber hinaus einen kleinen bis mittleren Effekt fluiden Intelligenz auf die Urteilsgenauigkeit im simulierten Klassenraum – allerdings wurden verwandte Konstrukte wie die Arbeitsgedächtniskapazität nicht kontrolliert. Huang, Liu und Chen (2018) zeigten in einer aktuellen Studie, dass Lehramtsstudierende bei spontaner Konfrontation mit einem Szenario über aggressives Sozialverhalten zunächst Empathie nutzen, um ihr weiteres Vorgehen zu informieren. Tettegah und Anderson (2007) fanden jedoch, dass Lehramtsstudierende auf ein solches Szenario kaum mit Empathie reagieren. Das heißt, sowohl die Rolle von Empathie in spontanen Urteilsprozessen als auch der Effekt von Intelligenz auf die Urteilsgenauigkeit Lehramtsstudierender ist bislang weitestgehend ungeklärt. Die vorliegende Studie schließt hier an.

Intelligenz und Empathie in der Urteilsbildung

Cattell (1987) definiert Intelligenz im Zwei-Faktoren-Modell als zwei analytisch trennbare Klassen kognitiver Fähigkeiten. Kristalline Intelligenz beschreibt die Anwendung kulturell erworbener Wissensbestände und kognitiver Fähigkeiten. Fluide Intelligenz beschreibt dagegen die Fähigkeit zur Anwendung kognitiver Operationen, die zur Lösung neuer und komplexer Problemsituationen benötigt werden. Dazu gehören u. a. das Erkennen abstrakter Gemeinsamkeiten und Unterschiede sowie logisches Schlussfolgern. Diese strukturelle Ähnlichkeit fluiden Intelligenz zu kognitiven Operationen in der diagnostischen Urteilsbildung macht fluide Intelligenz für die Forschung zu Diagnostik in der Lehrkräftebildung besonders interessant. Lehrkräfte müssen Erklärungen für Abweichungen einer gegenwärtigen und einer angestrebten Situation von Schülerinnen und Schülern finden. Dazu überführen sie Informationen aus Schul- bzw. Unterrichtssituationen in einen Abstraktionsgrad, der einen Vergleich über die konkrete Situation hinaus ermöglicht, bspw. mit anderen Personen oder diagnostischen Klassifikationen. Fluide Intelligenz sollte daher, wie auch bei Kaiser et al. (2012), positiv mit Urteilsgenauigkeit zusammenhängen.

Empathie ist die sozial-kognitive Fähigkeit zur stellvertretenden Erfahrung emotionaler Reaktionen anderer Personen und umfasst eine kognitive und eine affektive Komponente (Decety & Lamm, 2006). Kognitive Empathie ist die Fähigkeit, Gedanken und Gefühle anderer nachzuvollziehen (Perspektivenübernahme). Affektive Empathie ist die Fähigkeit, emotionale Reaktionen anderer nachzuempfinden. Kognitive und affektive Empathie werden sowohl als stabiles Trait-Merkmal als auch als situativ aktiviertes State-Merkmal untersucht. Aktuelle Modelle diagnostischer Kompetenz berücksichtigen Empathie bislang nicht, obwohl empirische Forschungsergebnisse das nahelegen.

So zeigten kognitionspsychologische Studien einen positiven Effekt kognitiver Trait-Empathie auf das korrekte Speichern und Abrufen personenbezogener Informationen (Gao, Ye, Shen, & Perry, 2016; Spreng & Mar, 2012; Stiller & Dunbar, 2007). Für affektive Trait-Empathie sind die Befunde uneinheitlich (Gao, Ye, Shen, & Perry, 2016; Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007; Wagner, Handke, & Walter, 2015). Dagegen zeigten Lombardo et al. (2007), dass affektive State-Empathie durch die Steigerung der Aufmerksamkeit ebenfalls einen positiven Effekt auf die Verarbeitung personenbezogener Informationen hat. Dementsprechend sollten kognitive Trait-Empathie sowie affektive State-Empathie einen positiven Effekt auf die Urteile von Lehramtsstudierenden haben.

Urteilsprozesse von Lehramtsstudierenden finden besonders in der ersten Studienphase häufig in didaktisch kontrollierten und zunehmend fallbasierten Settings statt (Kramer et al., 2017). Ein zentrales Merkmal dieses Settings – und damit der Beurteilungssituation – sind die Fälle selbst.

Diagnostische Fälle in der Lehrkräftebildung

L. Shulman (1992) definiert einen Fall als konkretes und greifbares Beispiel für ein theoretisches Konzept (z. B. Lernstörungen) oder eine Kategorie von Ereignissen (z. B. Elterngespräche). Ein diagnostischer Fall stellt etwa eine Unterrichtssituation dar, in der auffälliges Verhalten von Schülerinnen und Schülern beobachtet werden kann.

Diagnostische Fälle, theoretisch konstruiert oder aus reflektierter Unterrichtspraxis gewonnen, bilden das Kernstück zahlreicher Interventionen zur Steigerung diagnostischer Kompetenz von Lehramtsstudierenden. Dabei sind sie nicht nur häufig genutztes Lernmaterial (Klug, 2017; Kramer et al., 2017; Krammer, 2014; Santagata & Guarino, 2011; Seidel et al., 2011), sondern auch die Grundlage objektiver Testinstrumente, wie etwa dem Szenariotest von Klug et al. (2016), einigen Wissenstests (König & Klemenz, 2015; Klug, 2017) und Lehrbüchern für die Hochschullehre (Imhof, 2016; Zumbach & Mandl, 2007). Diagnostische Fälle haben den besonderen Vorteil, auch Lehramtsstudierenden ohne Unterrichtserfahrungen in einem geschützten Rahmen realitätsnahe Lern- und Prüfungssituationen anbieten zu können (Seidel et al., 2011).

Interventionsstudien mit Prä-Post-Design untersuchten überwiegend die Lernergebnisse diagnostischer Fallarbeit (s. o.). Dabei zeigte sich, dass Lehramtsstudierende Fälle abhängig von ihren kognitiven Lernvoraussetzungen unterschiedlich gut nutzen können. In einer Interventionsstudie mit Videos aus eigener Unterrichtspraxis hatte bspw. die Vorerfahrung mit Videoanalysen einen Effekt auf die Verbesserung der Wahrnehmungs- und Begründungsfähigkeiten (Seidel et al., 2011). In einer

Studie von Syring et al. (2015) zeigte sich zudem, dass Studierende kognitiv unterschiedlich auf Text- und Videofälle reagieren. Textfälle waren weniger motivierend, wurden aber auch als weniger belastend erlebt und lösten mehr Hoffnung auf Erfolg bei fallbezogenen Aufgaben aus. Zu den Lernprozessen selbst, d. h. mit welchen kognitiven Fähigkeiten sich Lehramtsstudierende Textfällen zuwenden, liegen jedoch nur wenige Befunde vor. Auf welche kognitiven Fähigkeiten greifen Studierende zurück, wenn sie mit Textfällen arbeiten?

Textverstehen und Textmerkmale

Lernprozesse mit Textfällen setzen zunächst das Verstehen der Textfälle voraus. Kintsch (1988) sowie Singer und Kintsch (2001) beschreiben in der Konstruktions-Integrations-Theorie das Textverstehen als Speicherung von Textinformationen in drei kognitiven Repräsentationsformen, die beim Lesen gleichzeitig gebildet werden, aber unterschiedlich stark ausgeprägt sein können. Die wortwörtliche Repräsentation beinhaltet exakte Worte und Sätze, die propositionale Repräsentation Bedeutungsgehalte und das Situationsmodell mit Gedächtnisinhalten und Schlussfolgerungen verknüpfte Textinhalte. Die propositionale Repräsentation und das Situationsmodell stehen in einem umgekehrten Zusammenhang, d. h. je stärker das Situationsmodell, desto schwächer die propositionale Repräsentation (Kintsch, 1988). Relevant für die vorliegende Studie ist insbesondere das Situationsmodell, da aus dem Situationsmodell Inferenzen gebildet werden können (Calvo, Castillo, & Schmalhofer, 2006), bspw. verhaltensbezogene Informationen in Relation zu einem theoretischen Aggressionsmodell setzen. Ein stark ausgeprägtes Situationsmodell sollte demnach einen positiven Effekt auf die Urteilsgenauigkeit haben.

Das Textverstehen ist u. a. von individuellen Merkmalen wie der Arbeitsgedächtniskapazität, aber auch von den Textmerkmalen wie dem Textstil abhängig. Dieser lässt sich nach Genette (1988) in Form von Fokalisierungen beschreiben. Fokalisierung bezeichnet die Darstellung der im Text vermittelten Informationen durch eine ausgewählte narrative Perspektive. Einen starken Kontrast bilden die psycho-narrative und die externale Fokalisierung (Stühling & Köppe, 2016). Im psycho-narrativen Textstil werden die zu vermittelnden Informationen als subjektive Wahrnehmungen eines Charakters dargestellt, bspw. durch Beschreibungen mit Hilfe von Gedanken und Gefühlen. Dagegen erfolgt die Darstellung von Informationen im external fokalisierten Stil in Form objektiver Beobachtungen.

Prinzipiell fördern narrative Textbestandteile die Ausprägung eines Situationsmodells (Dixon & Bortolussi, 2018; Salem, Westkott, & Holler, 2017). Allerdings wirken von Charakteren dargestellte Informationen, die inkongruent zum Handlungsverlauf sind, negativ auf das Erinnern dieser Informationen (Rapp & Kendeou, 2009). Da es in den verwendeten Textfällen um diagnostische Auffälligkeiten geht, bspw. ein Widerspruch zwischen einer Handlungsintention und einem Handlungsverlauf, ist eine sollte der psycho-narrative Textstil einen negativen Effekt auf die Urteilsgenauigkeit von Lehramtsstudierenden haben.

10.2 Fragestellung und Hypothesen

Die vorliegende Studie untersuchte die Effekte der kognitiven Grundfähigkeiten Intelligenz und Empathie, des Textverstehens und des Textstils textbasierter diagnostischer Fallszenarios auf die Urteilsgenauigkeit von Lehramtsstudierenden. Folgende Hypothesen wurden untersucht:

- (1) Die fluide Intelligenz hat einen positiven Effekt auf die Urteilsgenauigkeit.
- (2) Die kognitive Trait-Empathie hat einen positiven Effekt auf die Urteilsgenauigkeit.
- (3) Die affektive State-Empathie hat einen positiven Effekt auf die Urteilsgenauigkeit.
- (4) Das Situationsmodell hat einen positiven Effekt auf die Urteilsgenauigkeit.
- (5) Der psycho-narrative Textstil hat einen negativen Effekt auf die Urteilsgenauigkeit.

10.3 Methoden und Design

Durchführung

Die vorliegende Studie ist eine querschnittliche, experimentelle Onlinestudie. Zunächst beantworteten die Teilnehmenden Fragen zu Alter, Geschlecht, Studienfach etc. sowie zu affektiver und kognitiver Trait-Empathie. Anschließend lasen die Teilnehmenden entweder einen psycho-narrativen oder einen external-fokalisierten Fall über Dyskalkulie oder Prüfungsangst (2x2-Design). Danach beantworteten die Teilnehmenden Fragen zu affektiver und kognitiver State-Empathie und zu den Kontrollvariablen kognitiver Elaboration, Immersion, thematischem Interesse und wahrgenommener Ähnlichkeit zur Lehrkraft im Fall. Abschließend machten die Teilnehmenden einen Test zu fluider Intelligenz, einen Arbeitsgedächtnistest, einen Rekognitions- und Verifikationstest und trafen ein spontanes diagnostisches Urteil. Die Teilnahme fand während einer regulären Vorlesung im SoSe 2018 statt und dauerte ca. 20 Minuten.

Stichprobe

Die Teilnehmenden waren $N = 145$ Lehramtsstudierende (68,5 % weiblich) im Alter von 18–43 Jahren ($M = 22.51$, $SD = 3.97$) aus Jena ($N = 94$), Leipzig ($N = 39$) und Berlin ($N = 9$). Die Lehramtsstudierenden aus Jena und Berlin unterschieden im Alter ($M_{\text{Jena}} = 21.73$, $SD_{\text{Jena}} = 2.49$, $M_{\text{Leipzig}} = 23.26$, $SD_{\text{Leipzig}} = 4.57$, $M_{\text{Berlin}} = 29.00$, $SD_{\text{Berlin}} = 7.14$, $W(2, 18.58) = 6.11$, $p < .01$) und aus Jena und Leipzig in der Arbeitsgedächtniskapazität ($M_{\text{Jena}} = 0.63$, $SD_{\text{Jena}} = 0.12$, $M_{\text{Leipzig}} = 0.56$, $SD_{\text{Leipzig}} = 0.13$, $M_{\text{Berlin}} = 0.55$, $SD_{\text{Berlin}} = 0.14$, $W(2, 20.81) = 4.56$, $p < .05$). Die Teilnehmenden aus Jena, Leipzig und Berlin unterschieden sich nicht in den anderen gemessenen Variablen.

Instrumente

Fluide Intelligenz. Die fluide Intelligenz wurde mit einer 9-Item-Form des Raven's Standard Progressive Matrices Test von Bilker, Hansen, Brensinger, Richard, Gur und Gur (2012) gemessen. Die Teilnehmenden wurden gebeten, aus 6 oder 8 Teilstücken das Teilstück auszuwählen, das ein

vorgegebenes Muster richtig ergänzt. Die richtigen Antworten wurden summiert ($Range = 1 - 9$, $M = 6.56$, $SD = 1.72$).

Trait-Empathie. Die Trait-Empathie wurde mit der Basic Empathy Scale (BES) von Heynen et al. (2016) gemessen. Die Skala beinhaltet 5 Items für affektive Empathie ($Range = 0-3$, $M = 1.99$, $SD = 0.59$, $\alpha = .80$) und 5 Items für kognitive Empathie ($Range = 0-3$, $M = 2.38$, $SD = 0.42$, $\alpha = .74$), z. B. „Ich lasse mich leicht durch die Gefühle anderer beeinflussen.“ oder „Ich verstehe oft schon wie Menschen sich fühlen, bevor sie mir etwas darüber erzählt haben.“ Trait-Empathie wird insbesondere gegenüber ähnlich wahrgenommenen Personen gezeigt, daher wurde die wahrgenommene Ähnlichkeit zur Lehrkraft im Fall mit 4 Items des Measure for Perceived Homophily von McCroskey, Richmond und Daly (1975) kontrolliert ($Range = 0-3$, $M = 1.45$, $SD = 0.65$, $\alpha = .91$).

State-Empathie. Die State-Empathie wurde mit der State Empathy Scale von Shen (2010) gemessen. Die Skala beinhaltet 4 Items für affektive Empathie ($Range = 0-3$, $M = 2.07$, $SD = 0.58$, $\alpha = .84$) und 4 Items für kognitive Empathie ($Range = 0-3$, $M = 1.79$, $SD = 0.55$, $\alpha = .78$), z. B. „Ich hatte beim Lesen des Textes eine ähnliche Stimmung wie die Lehrerin/der Lehrer.“ und „Ich kann nachvollziehen, was die Lehrerin/der Lehrer im Text erlebt hat.“ State-Empathie wird besonders bei immersiven Texten aktiviert, daher wurde Immersion mit den 6 Items der Transportation Scale Short Form (TS-SF) von Appel, Gnambs, Richter und Green (2015) gemessen und kontrolliert ($Range = 0-6$, $M = 4.04$, $SD = 0.85$, $\alpha = .76$).

Textrepräsentation. Die Textrepräsentation wurde mit dem Rekognitions- und Verifikationstest (ReMem) von Wedel & Müller (2018) gemessen. Der ReMem beinhaltet 28 Sätze: 7 wortwörtliche, 7 paraphrasierte, 7 bedeutungsveränderte und 7 falsche Sätze. Die Teilnehmenden sollten für jeden Satz angeben, ob der Satz wortwörtlich im Text vorkam (Rekognition) und ob der Satz im Sinne des Textes korrekt war (Verifikation). Die richtigen Antworten wurden z-standardisiert und in d' -Werte mit einem $M = 0$ und $SD = 1.44$ transformiert ($Range_{d' \text{ wortwörtlich}} = -3.26-3.28$, $Range_{d' \text{ propositional}} = -4.14-2.72$, $Range_{d' \text{ situational}} = -2.91-4.66$). Die Textrepräsentation ist u. a. von der Arbeitsgedächtniskapazität, kognitiver Elaboration und thematischem Interesse am Text abhängig, daher wurden diese Konstrukte kontrolliert. Die Arbeitsgedächtniskapazität wurde mit einem Location-Letter-Test in Anlehnung an Wilhelm, Hildebrandt und Oberauer (2013) gemessen. Die Teilnehmenden sahen nacheinander für jeweils 500 ms mehrere 3x3-Gitter mit je einem Buchstaben ($ISI = 1000$ ms). Insgesamt gab es fünf Durchgänge mit 2–6 Gittern, also 2–6 Buchstaben. Die Teilnehmenden beantworteten nach jedem Durchgang entweder eine Letter-Frage, z. B. „Welcher Buchstabe stand auf Position 5?“ oder eine Location-Frage, z. B. „An welcher Position stand H?“. Die richtigen Antworten wurden je Durchgang und insgesamt gemittelt ($Range = 0.27-0.85$, $M = 0.60$, $SD = 0.13$). Kognitive Elaboration wurde mit den 4 Items einer Skala von Schiefele (1990) erfasst ($Range = 0-3$, $M = 1.97$, $SD = 0.57$, $\alpha = .63$). Thematisches Interesse wurde mit den 6 Items einer Skala aus Ramm et al. (2006) erfasst ($Range = 0-3$, $M = 1.76$, $SD = 0.59$, $\alpha = .85$).

Textstil und Textfälle. Textstil (psycho-narrativ, external fokalisiert) und Textinhalt (Dyskalkulie, Prüfungsangst) der vier eingesetzten Textfälle über eine aufgabenbezogene Gruppenarbeit unter Schülerinnen und Schülern wurden systematisch variiert (siehe Anhang 1). Der inhaltliche Fokus wurde durch systematische Unterschiede der Informationen über Friis hergestellt. Dazu gehörten Gesprächs-, Beobachtungs- und Testinhalte mit prototypischen Symptomen, durch die auf die jeweilige

Lernstörung von Friis geschlossen werden kann. Die Fälle wurden inhaltlich und formal so ähnlich wie möglich geschrieben (49–53 Sätze, 608–674 Wörter) und wurden den Teilnehmenden entsprechend des üblichen Aufbaus in der experimentellen Forschung zum Leseverstehen Satz für Satz präsentiert. Die Eignung der Fälle wurde in einer Validierung mit drei psychologischen Fachkräften und einer Lehrkraft festgestellt.

Urteil. Das spontane diagnostische Urteil wurde mit einer offenen Frage erfasst: „Nun geht es noch einmal um Friis. Was denken Sie, welche Lern- oder Leistungsbeeinträchtigung hat Friis?“ Die Antworten wurden als falsch, z. B. „Selbstwertprobleme“, als richtig mit hoher Spezifität, z. B. „Dyskalkulie“ oder als richtig mit niedriger Spezifität kodiert, z. B. „eine Art Lernbehinderung in Mathematik“. Das Kriterium bildete die Übereinstimmung mit der validierten Symptombeschreibung. 13 % der Teilnehmenden hatten ein richtiges Urteil mit hoher Spezifität, 15 % ein richtiges Urteil mit niedriger Spezifität und dementsprechend 28 % ein richtiges Urteil mit mindestens niedriger Spezifität. Ähnlich fanden auch Dünnebier, Krolak-Schwerdt und Gräsel (2009) Urteile unterschiedlicher Spezifität. Zur Kontrolle weiterer Voraussetzungen des Urteils wurden weiterhin die eigene Diagnose einer Lernstörung, die Diagnose einer Lernstörung von Freunden, Bekannten oder Verwandten und die Anzahl besuchter Veranstaltungen zu pädagogischer Diagnostik erfasst.

10.4 Ergebnisse

Um zu prüfen, ob fluide Intelligenz und Empathie die Wahrscheinlichkeit eines richtigen Urteils mit hoher und mit niedriger Spezifität vorhersagen, wurden zwei schrittweise logistische Regressionen mit den spontanen diagnostischen Urteilen als abhängiger Variable (0 = falsch, 1 = richtig) berechnet. Im ersten Schritt wurden Alter und Geschlecht, im zweiten Schritt fluide Intelligenz und Arbeitsgedächtnis, im dritten Schritt Trait-Empathie, im vierten Schritt Textstil und Textinhalt, im fünften Schritt die Textrepräsentation, im sechsten Schritt State-Empathie, kognitive Elaboration, Immersion, thematisches Interesse sowie wahrgenommene Ähnlichkeit und im siebten Schritt die eigene Diagnose sowie die Diagnose von Freunden, Bekannten und Verwandten ins Modell aufgenommen. Dargestellt in Tabelle 10 ist nur der siebte Schritt der Regressionen, die gesamte Tabelle findet sich in Tabelle 11 im Anhang.

Richtiges Urteil mit hoher Spezifität

Im ersten Regressionsschritt war Alter ein signifikanter Prädiktor für das richtige Urteil und die χ^2 -Tests zeigten eine gute Modellgüte (Model χ^2 (2) = 6.04, $p < .05$, Block χ^2 (2) = 6.04, $p < .05$, Nagelkerke R^2 = .08). Mit zunehmendem Alter stieg also die Wahrscheinlichkeit eines richtigen Urteils. Im zweiten Regressionsschritt war Alter nicht mehr signifikant, aber fluide Intelligenz. Mit zunehmender fluider Intelligenz stieg also die Wahrscheinlichkeit eines richtigen Urteils (χ^2 (4) = 17.77, $p < .01$, Nagelkerke R^2 = .21). Im dritten (χ^2 (6) = 19.81, $p < .01$, Nagelkerke R^2 = .24), vierten (χ^2 (8) = 20.72, $p < .01$, Nagelkerke R^2 = .25) und fünften (χ^2 (10) = 21.19, $p < .05$, Nagelkerke R^2 = .25) Regressionsschritt war die fluide Intelligenz weiterhin signifikant, aber die Hinzunahme von affektiver und kognitiver Trait-

Empathie, Textstil und Textinhalt sowie wortwörtlicher, propositionaler und situationaler kognitiver Textrepräsentation erhöhte nicht die Vorhersagekraft des Modells. Im sechsten und siebten Regressionsschritt waren zusätzlich zu fluider Intelligenz auch affektive State-Empathie und kognitive Elaboration signifikant. Die Hinzunahme der State-Variablen erhöhte die Vorhersagekraft des Modells ($\chi^2(16) = 37.92, p < .01$, Nagelkerke $R^2 = .43$), die Hinzunahme der eigenen Diagnose und der Diagnose von Freunden, Bekannten und Verwandten hingegen nicht ($\chi^2(18) = 38.43, p < .01$, Nagelkerke $R^2 = .43$).

Tabelle 10

Logistische Regression für richtiges Urteil hoher und niedriger Spezifität

Schritt	Konstrukt	Richtiges Urteil mit hoher Spezifität			Richtiges Urteil mit mindestens niedriger Spezifität		
		<i>B</i> (SE)	Walds χ^2, p	OR	<i>B</i> (SE)	Walds χ^2, p	OR
7	Alter	0.44 (.32)	1.95, > .05	1.56	0.45 (.25)	3.29, > .05	1.57
	Geschlecht	0.20 (.42)	0.22, > .05	1.22	0.27 (.26)	0.94, > .05	1.31
	FI	1.55 (.51)	9.19, < .01	4.72	0.20 (.24)	0.68, > .05	1.22
	AG	-0.05 (.39)	0.02, > .05	0.95	-0.07 (.24)	0.09, > .05	0.93
	TE affektiv	0.24 (.44)	0.30, > .05	1.27	0.25 (.29)	0.76, > .05	1.28
	TE kognitiv	-0.35 (.51)	0.47, > .05	0.71	0.25 (.30)	0.69, > .05	1.28
	Textstil	-0.03 (.35)	0.01, > .05	0.97	0.30 (.24)	1.51, > .05	1.35
	Textinhalt	-0.18 (.37)	0.24, > .05	0.84	0.08 (.24)	0.12, > .05	1.09
	d' wortwörtlich	0.32 (.36)	0.77, > .05	1.37	-0.03 (.24)	0.01, > .05	0.97
	d' propositional	1.08 (.60)	3.29, > .05	2.95	0.82 (.37)	5.10, < .05	2.28
	d' situational	0.62 (.61)	1.05, > .05	1.37	0.96 (.38)	6.57, < .05	2.62
	SE affektiv	1.35 (.51)	6.92, < .01	3.85	0.88 (.33)	7.01, < .01	2.42
	SE kognitiv	-0.25 (.54)	0.22, > .05	0.78	0.15 (.35)	0.20, > .05	1.17
	KE	1.04 (.45)	5.29, < .05	2.83	0.50 (.30)	2.85, > .05	1.65
	IM	0.11 (.60)	0.03, > .05	1.11	-0.25 (.35)	0.49, > .05	0.78
	TI	0.15 (.45)	0.11, > .05	1.16	0.01 (.27)	0.00, > .05	1.00
	WAE	-0.62 (.55)	0.03, > .05	0.54	-0.92 (.39)	5.41, < .05	0.40
	D _{Selbst}	0.09 (.53)	0.03, > .05	1.09	0.15 (.26)	0.31, > .05	1.16

Anmerkungen. FI = Fluide Intelligenz, AG = Arbeitsgedächtnis, TE = Trait-Empathie, d' = d' -Wert der kognitiven Textrepräsentation, SE = State-Empathie, KE = Kognitive Elaboration, IM = Immersion, TI = Thematisches Interesse, WAE = Wahrgenommene Ähnlichkeit, D = Diagnose; $N = 145$.

Richtiges Urteil mit mindestens niedriger Spezifität

Im ersten Regressionsschritt war Alter ein signifikanter Prädiktor für das richtige Urteil und die χ^2 -Tests zeigten eine gute Modellgüte ($\chi^2 (2) = 6.82, p < .05$, Nagelkerke $R^2 = .07$). Im zweiten Regressionsschritt war Alter weiterhin signifikant, aber χ^2 -Tests zeigten, dass es durch die Hinzunahme von fluider Intelligenz und Arbeitsgedächtnis keinen signifikanten Anstieg der Vorhersagekraft des Modells gibt ($\chi^2 (4) = 7.77, p > .05$, Nagelkerke $R^2 = .08$). Im dritten Regressionsschritt war Alter nicht mehr signifikant, aber kognitive Trait-Empathie. Das Modell hatte eine gute Modellgüte und die Hinzunahme von affektiver und kognitiver Trait-Empathie erhöhte die Vorhersagekraft des Modells signifikant ($\chi^2 (6) = 14.97, p < .05$, Nagelkerke $R^2 = .14$). Im vierten ($\chi^2 (8) = 15.92, p < .05$, Nagelkerke $R^2 = .15$) Regressionsschritt war kognitive Trait-Empathie weiterhin signifikant, aber keiner der anderen Prädiktoren. Im fünften Schritt ($\chi^2 (10) = 20.78, p > .05$, Nagelkerke $R^2 = .19$) war kognitive Trait-Empathie nicht länger signifikant, jedoch das Situationsmodell. Im sechsten ($\chi^2 (16) = 36.26, p < .05$, Nagelkerke $R^2 = .32$) und siebten ($\chi^2 (18) = 36.58, p > .05$, Nagelkerke $R^2 = .32$) Regressionsschritt wurden affektive State-Empathie, die propositionale und situationale Textrepräsentation sowie wahrgenommene Ähnlichkeit zu der dargestellten Lehrkraft signifikant, aber kognitive Trait-Empathie nicht mehr.

Daher wurde eine Mediationsanalyse mit kognitiver Trait-Empathie als unabhängiger, affektiver State-Empathie als mediierender und richtigem Urteil mit niedriger Spezifität als abhängiger Variable berechnet (siehe *Abbildung 18*). Kognitive Trait-Empathie hatte, ebenso wie affektive Trait-Empathie, kognitive State-Empathie, wahrgenommene Ähnlichkeit und propositionale sowie situationale Textrepräsentation, einen signifikanten positiven Effekt auf affektive State-Empathie ($F(18, 145) = 6.81, p < .001, R^2 = .49$), aber nicht auf das richtige Urteil. Affektive State-Empathie hatte wiederum einen signifikanten Effekt auf das richtige Urteil (Modell III (19) = 28.46, $p < .05$, Nagelkerke $R^2 = .32$). Der indirekte Effekt war mit $B = .25, SE = .82, BootLLCI = .05, BootULCI = .80, p < .05$ signifikant.

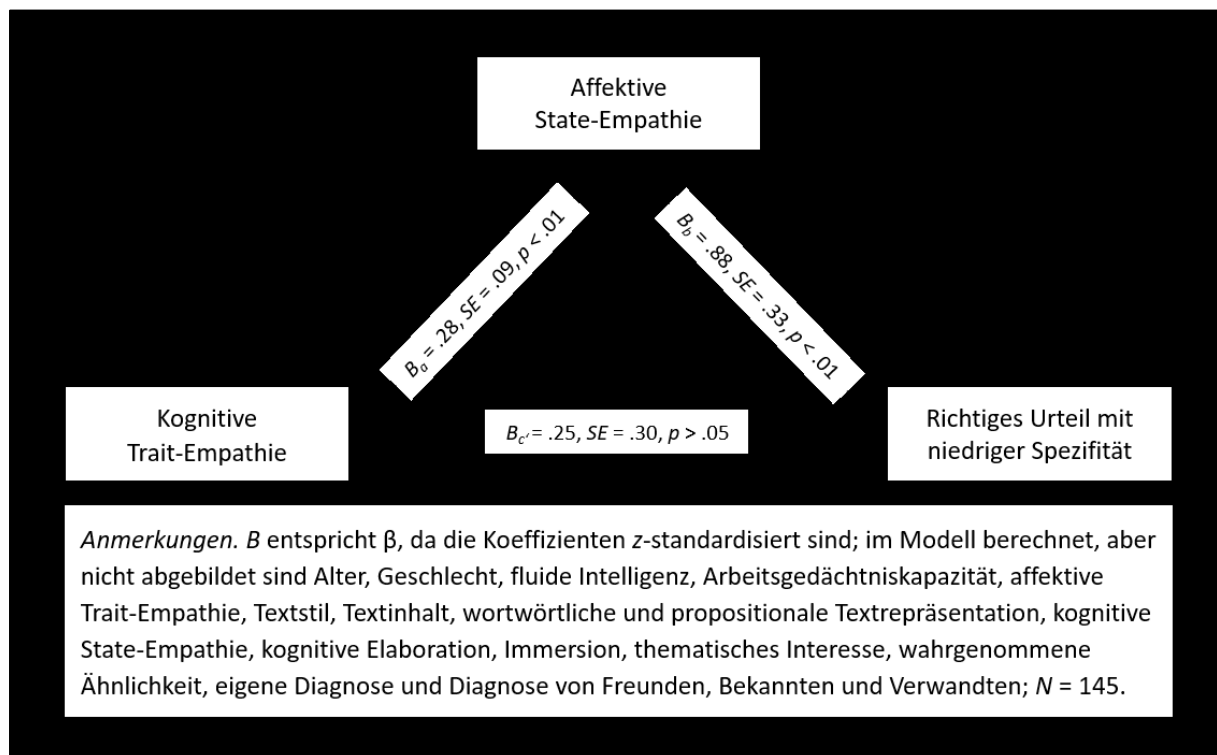


Abbildung 18. Mediation von kognitiver Trait-Empathie über affektive State-Empathie auf das richtige Urteil mit niedriger Spezifität.

10.5 Diskussion

Die vorliegende Studie untersuchte die Effekte kognitiver und sozial-kognitiver Grundfähigkeiten sowie von Textmerkmalen diagnostischer Fälle auf die Urteilsgenauigkeit von Lehramtsstudierenden. Diese lasen Textfälle zu Schülerinnen und Schülern mit Dyskalkulie oder Prüfungsangst und trafen im Anschluss in einem offenen Antwortformat ein diagnostisches Urteil. Wie bei Dünnebier et al. (2009) unterschieden sich die Urteile hinsichtlich ihrer Spezifität. So ließen sich richtige Urteile in richtige Urteile mit hoher und niedriger Spezifität einteilen, die sich durch den Grad an Symptomgeneralisierung unterschieden. Eine mögliche Erklärung für diesen Befund, die über die Prozessmodelle von Klug et al. (2016) und Herppich et al. (2018) hinaus geht, bietet das Linsenmodell. Im Linsenmodell wird die Urteilsgenauigkeit unter anderem durch die Gewichtung von Hinweisreizen für eine Diagnose erklärt. Da in den textbasierten Schulszenarien Hinweisreize mit unterschiedlicher Spezifität vorkamen, konnten diese in unterschiedlichem Maß gespeichert sowie auch in der Urteilsbildung gewichtet werden.

Positiver Effekt von Intelligenz bei hoher aber nicht bei mindestens niedriger Spezifität

Richtige Urteile mit hoher Spezifität wurden zum einen ähnlich wie bei Kaiser et al. (2012) durch die im engeren Sinne kognitionpsychologischen Konstrukte vorhergesagt. Vor allem fluide Intelligenz, aber unabhängig davon auch kognitive Elaboration unterstützten die Urteilsgenauigkeit, letztere möglicherweise durch das Erinnern an Wissensbestände oder den Abgleich mit symptombezogenen

Protopyen. Fluide Intelligenz hatte dagegen keinen Effekt auf richtige Urteile mit mindestens niedriger Spezifität. Das kann in Anlehnung an Forschung zu Zwei-Prozess-Modellen der Informationsverarbeitung (Chaiken & Trope, 1999) daran liegen, dass Personen mit höherer fluider Intelligenz eher zur systematischen Informationsverarbeitung neigen (Evans, 2011) und ihr spontanes Urteil daher eher an spezifischen Personeninformationen als an allgemeinen sozialen Beschreibungen orientieren. Das würde auch an vergleichbare Befunde der Expertiseforschung anschließen (Dünnebier et al., 2009; Krolak-Schwerdt, Böhmer, & Gräsel, 2012).

Zusammengefasst kann Hypothese 1 beibehalten werden. Zukünftige Forschung sollte jedoch unbedingt für die Spezifität des Urteils kontrollieren.

Positiver Effekt von Empathie bei hoher und mindestens niedriger Spezifität

Die Mediationsanalyse zeigte, dass kognitive Trait-Empathie zwar keinen direkten, aber vermittelt über affektive State-Empathie einen indirekten positiven Effekt auf die Bildung richtiger Urteile mit mindestens niedriger Spezifität hatte. Affektive State-Empathie hatte ebenfalls einen direkten positiven Effekt auf die Urteilsgenauigkeit. Dieser bestand sowohl für Urteile mit hoher als auch für Urteile mit mindestens niedriger Spezifität.

Eine Erklärungsmöglichkeit dafür bietet die Gefühl-als-Information Heuristik (Schwarz & Clore, 1983). Gefühle können als affektives Priming das Speichern, Abrufen und Verwenden von Informationen in Entscheidungssituationen beeinflussen. Eine Studie von Wagner, Handke und Walter (2015) zeigte bspw., dass ein höheres Maß an affektiver Empathie zur bevorzugten Speicherung sozialer gegenüber nicht-sozialen Informationen führt. Lehramtsstudierende, für die die Gedanken und Gefühle der Schülerinnen und Schüler in den textbasierten Schulszenarien salient waren, hatten möglicherweise eine Präferenz für die Speicherung personenbezogener Informationen und konnten dadurch genauere Urteile treffen.

Eine andere Erklärungsmöglichkeit für den Effekt affektiver State-Empathie auf die Urteilsgenauigkeit bietet die Risiko-als-Gefühl Heuristik (Loewenstein, Weber, Hsee & Welch, 2001). Ausgangspunkt dieser Heuristik ist der Effekt von Gefühlen auf die Einschätzung der Wahrscheinlichkeit möglicher Konsequenzen. Angst oder Hoffnung können in unsicheren Situationen zu einer Überbewertung eigentlich geringer Wahrscheinlichkeiten führen – mit anderen Worten ‚klingeln die Alarmglocken‘. Das Nachempfinden der Emotionen von Schülerinnen und Schülern in problematischen Situationen kann negative Gefühle bewusster machen, dadurch zu einer Überbewertung der Wahrscheinlichkeit negativer Konsequenzen und einer entsprechenden Steigerung der Aufmerksamkeit auf diese Schülerinnen und Schüler führen. Diese Erklärung wird dadurch unterstützt, dass auch die wahrgenommene Ähnlichkeit zur Lehrkraft ein signifikanter Prädiktor der Urteilsgenauigkeit bei mindestens niedriger Spezifität war. Möglicherweise unterstütze diese Wahrnehmung das Nachempfinden der Lage, in der sich die Lehrkraft im Text befindet.

Zusammengefasst muss Hypothese 2 verworfen werden. Hypothese 3 dagegen kann beibehalten werden. Insgesamt profitierten Lehramtsstudierende in der spontanen Urteilsbildung von sozial-kognitiven Fähigkeiten, besonders wenn sie die Emotionen der Schülerinnen und Schüler situativ

nachempfanden. Das heißt für den didaktischen Umgang mit Textfällen in der Lehrkräftebildung, dass Lehramtsstudierende zum Nachempfinden dargestellter Situationen aus Perspektive der Schülerinnen und Schüler aufgefordert werden könnten, was zu einer ‚günstigeren‘ Informationsverarbeitung führen könnte als auch die Fähigkeit, in diagnoserelevanten Situationen mit Empathie zu reagieren, trainieren könnte.

Kein Effekt von Textmerkmalen, aber von Textverstehen

Die Stärke des Situationsmodells hatte einen Effekt auf die Akkuratheit von Urteilen mit mindestens niedriger Spezifität, d. h. Hypothese 4 kann für diesen Spezifitätsgrad beibehalten werden. Dass die Bildung von Situationsmodellen keinen Effekt auf die Urteilsgenauigkeit bei hoher Spezifität hatte, ist überraschend, da einerseits Inferenzbildung Teil von Urteilsprozessen ist und andererseits Situationsmodelle die Inferenzbildung erleichtern. Eine mögliche Erklärung findet sich im Experimentaldesign. Calvo et al. (2006) schalteten dem Lesevorgang eine explizite Instruktion zur Inferenzbildung vor, in dem auch das Objekt der Inferenzbildung spezifiziert wurde. Dies war in der vorliegenden Studie nicht der Fall. Möglicherweise integrierten die Lehramtsstudierenden daher nur solche Informationen in ihr Situationsmodell, die eine Inferenz- und Urteilsbildung nicht unterstützten. Eine systematische Variation der mit der Informationsverarbeitung verbundenen Zielstellung könnte ein nächster Schritt für anschließende Studien sein.

Der Textstil hatte ebenfalls keinen Effekt auf die Urteilsgenauigkeit, d. h. Hypothese 5 muss verworfen werden. Eine mögliche Erklärung ist die Satz-für-Satz Präsentation des diagnostischen Textfalls und die dadurch entstehenden Leseпаusen. Allerdings ist die Satz-für-Satz Präsentation typisch für Studien zum Leseverstehen und aufgrund der Vergleichbarkeit mit anderen Studien anderen Formen der Präsentation vorzuziehen. Eine weitere Erklärung könnte sein, dass die Länge des Textfalls den Effekt des Textstils moderiert. Salem et al. (2017) nutzten bspw. einen Text mit 35 Sätzen, während die Textfälle in der vorliegenden Studie mit etwa 50 Sätzen deutlich länger waren. Darüber hinaus war der Unterschied zwischen den Textstilen im Experimentalmaterial von Salem et al. (2017) größer als in der vorliegenden Studie.

Aus didaktischer Sicht legen diese Ergebnisse nahe, dass die Urteilsgenauigkeit nicht über den Textstil gesteuert werden kann, was eine einfache didaktische Möglichkeit zur Förderung diagnostischer Kompetenz in Interventionen verschließt. Andererseits sind auch keine spezifischen Testeffekte des Textstils zu erwarten, was die Zuverlässigkeit vorhandener Forschungsinstrumente bekräftigt.

Grenzen und Stärken

Die vorliegende Studie hat einige Limitationen, die in der Ergebnisinterpretation berücksichtigt werden müssen. Erstens sind gegenüber einem realen Klassenraum die Anwendungsmöglichkeiten sozial-kognitiver Fähigkeiten in der Auseinandersetzung mit textbasierten Schulszenarien beschränkt, sodass eher von einer Unterschätzung der tatsächlichen Effekte ausgegangen werden muss. Zweitens

initiierten die Teilnehmenden den Urteilsprozess nicht selbst – was für eine möglichst ökologisch valide Studie aber wünschenswert wäre, da die Zielbildung einen positiven Effekt auf die Urteilsgenauigkeit hat (Klug et al., 2016; Dünnebieer et al., 2009). Drittens wurde nur ein einziges Urteil erfasst, sodass hierfür keine Reliabilität berechnet werden konnte. Diese Entscheidung wurde zugunsten der Komplexität der präsentierten Schulszenarien und der umfangreichen Kontrollvariablen getroffen. Viertens waren die Zusammensetzung und die Größe der Stichprobe nicht ideal, bspw. konnte keine Messinvarianz berechnet werden. Andererseits zeigten Simulationsstudien, dass die vorliegende Stichprobe auch bei der verhältnismäßig großen Anzahl an Prädiktoren zumindest für eine Mediationsanalyse eine sehr gute Teststärke aufweist (Fritz & MacKinnon, 2007).

Die vorliegende Studie hat auch eine Reihe von Stärken. Erstens wurden mehrere kognitive Grundfähigkeiten sowie relevante Kontrollvariablen (z. B. Arbeitsgedächtniskapazität) gemessen, sodass die Effektgrößen wesentlich genauer geschätzt werden konnten als in bisherigen Studien, die verwandte Konstrukte nicht kontrolliert haben. Zweitens wurde ein offenes Antwortformat zur Erfassung der Urteile angewendet, das im Vergleich zum häufig verwendeten Forced-Task-Ansatz, bei dem die einzuschätzenden Inhalte vorgegeben werden, ökologisch deutlich valider ist. Drittens nutzte die vorliegende Studie ein Experimentaldesign mit randomisierter Zuteilung der Teilnehmenden, was die Zuverlässigkeit der Ergebnisse zum Effekt von Textmerkmalen auf die Urteilsgenauigkeit erhöht.

Schlussbemerkung

Insgesamt zeigen die Ergebnisse, dass sowohl kognitive als auch sozial-kognitive Fähigkeiten und Prozesse eine Rolle für die Urteilsgenauigkeit spielen. Unter diesen ist Empathie besonders interessant für die Lehrkräftebildung, da es bereits wirksame Empathietrainings speziell für angehende Lehrkräfte im Studium (Withford & Emerson, 2018) gibt. In aktuellen Modellen diagnostischer Kompetenz werden sozial-kognitive Fähigkeiten und der soziale Referenzrahmen aber bislang nicht abgebildet (z. B. Herppich et al., 2018).

Wenngleich die vorliegende Studie keine differenzierten Aussagen über die Informationsverarbeitung und Entscheidungsbildung beim Diagnostizieren machen kann, werfen die Ergebnisse die Frage nach der Verwendung unterschiedlicher Urteilsheuristiken auf und eröffnen damit ein breites Untersuchungsfeld. Gigerenzer (2015) konnte in mehreren Studien zeigen, dass Urteilsheuristiken zu ähnlichen Urteilen wie systematische Denkvorgänge führen können, aber wesentlich ressourcensparender sind. Insbesondere in Unterrichtssituationen, in denen es Barrieren für systematische Diagnosevorgänge gibt, sind Lehrkräfte auf Urteilsheuristiken angewiesen. Die weitere Erforschung sozial-kognitiver Prozesse und Heuristiken von Lehrkräften bietet daher ein großes Erkenntnispotential für die Konzeption und Förderung diagnostischer Kompetenz.

DISKUSSIONSTEIL

11 Gesamtdiskussion und Ausblick

Ziel der Arbeit war die Identifikation förderlicher Bedingungen für die Entwicklung diagnostischer Kompetenz angehender Lehrkräfte. Dazu wurde eine kooperativ und problemorientiert angelegte Intervention mit Textfällen entwickelt und mehrfach in quasi-experimentellen Studien auf ihre Wirkungen hin untersucht. Daran anschließend stellte sich die Frage, welche Rolle grundlegende kognitive Voraussetzungen bzw. Fähigkeiten in der Verarbeitung diagnosebezogener Textfälle spielen.

Der kooperative Aspekt wurde auf Grund der zentralen Bedeutung kooperativen Handelns für die professionelle Weiterentwicklung und Qualitätssteigerung im Kontext diagnostischen Handelns gewählt (Hesjedal, Hetland, & Iversen, 2015; Jurkowski & Müller, 2018). In Kapitel 3.2 wurde ausführlich erläutert, dass kooperatives diagnostisches Handeln zugleich eine schulische Realität ist und ein deutlicher Bedarf an deren systematischer Förderung in und vor Aufnahme der alltäglichen Berufspraxis besteht. Demgegenüber zielen bisherige Interventionen zur Kompetenzentwicklung lediglich auf individuelles Handeln und nutzen dementsprechend primär individuelle Übungstätigkeiten (bspw. Klug, 2011; Krammer, 2014; Santagata & Guarino, 2011; Seidel, Blomberg, & Stürmer, 2010; Trittel, Gerich, & Schmitz, 2014). Eine mögliche Ursache dieser Aussparung liegt in den verfügbaren theoretischen Modellen und Definitionen diagnostischer Kompetenz, in denen nur die Interaktion von Lehrkräften mit Schülerinnen und Schülern, nicht aber die Interaktion von Lehrkräften untereinander und mit Personen aus anderen Professionen expliziert wird (Herppich et al., 2018; Leuders, Dörfler, Leuders, & Philipp, 2018; Schrader, 2017; ausführlich siehe Kapitel 3.1).

Der problemorientierte Aspekt wurde zum einen als Designgrundlage gewählt, da sich diagnostische Fälle als Probleme beschreiben lassen (siehe Kapitel 4.1). Diesbezüglich bestand die Annahme, dass Lehramtsstudierende in der Auseinandersetzung mit textbasierten Diagnosesituationen im Rahmen problemorientierten Lernens eher die Fähigkeit zur Elaboration und Anwendung ihres fallbezogenen Wissens entwickeln. Die Prüfung dieser Annahme war auch dadurch motiviert, dass Studien zur videogestützten Kompetenzentwicklung durchgehend einen geringen Elaborationsgrad des erworbenen Wissens finden (Bohl et al., 2016). Zum anderen sind aus der medizinischen Ausbildung zwar positive Effekte problemorientierten Lernens auf die Kompetenzentwicklung umfangreich belegt, in der Lehrkräftebildung existieren jedoch deutlich weniger äquivalente Bildungsangebote (Walker & Leary, 2009; Chernikova et al., 2019). Darüber hinaus existieren mit Ausnahme der Intervention von Klug (2011) keine Replikationen entwickelter Interventionen. In Bezug auf letztere wurden Interventionseffekte jedoch ebenfalls nicht in einem vergleichenden statistischen Modell geprüft. Dadurch bleibt unklar, inwiefern die Ergebnisse der bisherigen Interventionsforschung zu diagnostischer Kompetenz von Lehramtsstudierenden mit Kohorteneffekten konfundiert sind.

Die Definition von Kompetenz als funktionelle kognitive Struktur macht deutlich, dass die Entwicklung einzelner kognitiver Module im Sinne von Dispositionen (z. B. Wissen, Überzeugungen, Einstellungen) eine notwendige aber keine hinreichende Bedingung für kompetentes Handeln ist (Klieme, 2004; Klieme, Hartig, & Rauch, 2008). Vielmehr braucht es ein kontextspezifisches und

anwendungsbezogenes Zusammenwirken der Dispositionen, um effektiv im professionellen Kontext zu handeln. Interventionen zur Kompetenzentwicklung müssen sich also auch daran messen, inwiefern sie den Teilnehmenden ermöglichen, Zusammenhänge zwischen den notwendigen kognitiven Dispositionen zu entwickeln. An diesem Kriterium haben sich jedoch *keine* der bislang entwickelten Interventionen für diagnostische Kompetenz gemessen. In Kapitel 2.2 und 9.1 wurde diskutiert, dass dies keinesfalls ein selbstgewählter Zustand ist, sondern vielmehr den herausfordernden Studienbedingungen und eingeschränkten statistischen Modellierungsmöglichkeiten geschuldet ist. Daraus resultiert sowohl ein inhaltlicher Bedarf im Bereich der Diagnostik, als auch ein Bedarf an Exploration anwendbarer statistischer Verfahren.

Vor diesem Hintergrund tragen die Studien in Kapitel 7 bis 9 zur Befundlage über förderliche Bedingungen zur Entwicklung diagnostischer Kompetenz durch die Untersuchung der Frage bei, inwiefern in einer kooperativ und problemorientiert angelegten Intervention eine Kompetenzentwicklung in der frühen Phase der Lehramtsausbildung erfolgen kann.

Neben förderlichen Umweltbedingungen der Kompetenzentwicklung zielte die Forschungsarbeit auch auf die Untersuchung von Zusammenhängen kognitiver Grundfähigkeiten mit diagnostischer Kompetenz. Diese Motivation leitete sich aus der theoretischen Bestimmung des Kompetenzkonstrukts und einem Mangel an Studien zum Zusammenhang kognitiver Grundfähigkeiten mit kompetentem diagnostischem Handeln ab. So zeigten lediglich Kaiser et al. (2012) einen Zusammenhang fluider Intelligenz mit diagnostischer Urteilsgenauigkeit, kontrollierten dabei jedoch nicht die Arbeitsgedächtniskapazität oder verwandte Konstrukte, sodass die Ergebnisse nur sehr eingeschränkt interpretierbar sind. Eine Reihe kognitionspsychologischer Studien zeigten darüber hinaus, dass Empathie einen positiven Effekt auf die Verarbeitung personenbezogener Informationen hat und ein Prädiktor für eine höhere Interventionswahrscheinlichkeit von Lehrerinnen und Lehrern in Bezug auf Lernstörungen ist (bspw. Ha & Park, 2014; Huang, Liu, & Chen, 2018; Park & Hwang, 2013). Dass Empathie und damit verwandte kontextspezifische sozial-kognitive Fähigkeiten bislang nicht in grundlegenden Modellen professionellen Handelns von Lehrkräften integriert sind (z. B. Baumert & Kunter, 2006) ist nicht direkt nachvollziehbar.

Eine mögliche Ursache könnte die Fokussierung der Unterrichts- und Professionsforschung auf die Konstrukte des Wissens und der Überzeugungen sein, die enger mit genuin kognitiven Fähigkeiten verwandt und traditionelle Outcomevariablen des Bildungssystems sind. Die Expertiseforschung zeigt die große Bedeutung dieser traditionellen Konstrukte für professionelles Handeln auf, schließt jedoch andere Konstrukte keinesfalls aus (Berliner, 1991; Green & Gillhooly, 1992). In Bezug auf Wissen als kognitive Disposition besteht ebenfalls der klare Befund, dass zu fallbezogenem Wissen sedimentierte mentale Modelle eine zentrale Voraussetzung für kompetentes Handeln auf hohem Niveau sind (im Überblick: Bradley, Paul, & Seeman, 2006). Mit Blick auf die Identifikation förderlicher Umweltbedingungen für relevante kognitive Dispositionen besteht daher besonderes Interesse daran, inwiefern sich schon in frühen Phasen der Lehrkräftebildung differenzielle Effekte des Elaborationsgrades mentaler Modelle auf die diagnostische Urteilsgenauigkeit zeigen lassen.

Ein weiterer Beitrag der vorliegenden Arbeit ist daher mit der Experimentalstudie in Kapitel 10 die Untersuchung der Frage, inwiefern kognitive Grundfähigkeiten und der Elaborationsgrad mentaler Modelle einen Effekt auf die diagnostische Urteilsgenauigkeit angehender Lehrkräfte haben.

11.1 Zusammenfassung zentraler Befunde

Die Untersuchung der Forschungsfrage zur Förderung der Entwicklung diagnostischer Kompetenz angehender Lehrkräfte durch eine kooperativ und problemorientiert angelegte Intervention erbrachte folgende Ergebnisse:

- (1) Die Studie in Kapitel 7 zeigt, dass sich eine kontinuierliche kooperative Bearbeitung diagnostischer Fälle dazu eignet, kognitive Dispositionen diagnostischer Kompetenz zu entwickeln. Im quasi-experimentellen Vergleich zu einer Kontrollgruppe, die mit theoretischen Inputs und darauf bezogenen Anwendungsübungen arbeitete, zeigten sich kleine bis große positive Effekte auf das Diagnosewissen ($.042 < \eta^2_p < .098$) und ein großer positiver Effekt auf das Fähigkeitsselbstkonzept ($\eta^2_p = .210$, $p = .014$). Das Diagnosewissen wurde mit einem eigens dafür entwickelten Testinstrument im Erkennen, Verstehen und Anwenden gemessen, d. h. sowohl in Richtung deklarativer Wissensbestände als auch in Richtung prozeduralen Wissens (PDW; Wedel & Müller, 2016). Die Analysen ergaben deutliche differenzielle Effekte, mit einem nicht signifikanten Gruppenunterschied im Anstieg der Fähigkeit des Erkennens ($\eta^2_p = .042$, $p = .307$), und signifikanten großen Effekten auf den Anstieg kognitiver Fähigkeiten im Verstehen ($\eta^2_p = .167$, $p = .031$) und Anwenden ($\eta^2_p = .145$, $p = .050$) diagnostischer Wissensbestände. Die Untersuchung unterstützt die Annahme, dass systematisch kooperatives Handeln im diagnostischen Kontext bereits in einem frühen Stadium der Lehramtsausbildung zur Kompetenzentwicklung beitragen kann. Ob die Teilnehmenden auch ihre kooperativen Fähigkeiten weiterentwickelten, wurde nicht gemessen. Der Fokus lag auf der Untersuchung der individuellen Dispositionen, da diese grundlegend für kompetentes Handeln sind.
- (2) Dieses Ergebnis überprüfte die in Kapitel 8 vorgestellte Studie durch eine exakte Replikation. Darüber hinaus wurde das Untersuchungsdesign um Experience Sampling in Bezug auf das immersive Erleben und die Einschätzung der Fälle in Bezug auf Komplexität und Schwierigkeit erweitert. Insgesamt zeigten sich ebenfalls positive Effekt auf die kognitiven Dispositionen diagnostischer Kompetenz. Diese waren aber, wie für Replikationsstudien zu erwarten, kleiner als in der ersten Studie (Nosek et al., 2015). So bestand ein mittlerer positiver, aber nicht signifikanter Effekt auf das diagnostische Fähigkeitsselbstkonzept ($\eta^2_p = .080$, $p = .118$) und kleine bis mittlere, aber nicht signifikante Effekte auf das Diagnosewissen ($.019 < \eta^2_p < .086$). Unerwartet war dagegen, dass sich bezüglich der differenziellen Effekte auf das Wissen auch ein anderes Befundmuster zeigte. So ergab die Varianzanalyse einen größeren aber nicht signifikanten Effekt der Intervention auf die kognitive Fähigkeit des Erkennens theoretisch gesicherter Wissensbestände ($\eta^2_p = .086$, $p = .088$) im Vergleich zum wissensbasierten Verstehen von Fallvignetten ($\eta^2_p = .052$, $p = .186$) und der fallbezogenen Anwendung diagnostischen Wissens ($\eta^2_p = .019$, $p = .431$). Die Diskussion der Studie stellt explorativ Erklärungen dieser Unterschiede vor, wie etwa systematische Unterschiede der untersuchten Gruppen in den Lernzielen, die sich die Studierenden im Rahmen der Intervention für jeden Fall selbst setzten. Das Experience Sampling ergab, dass die entwickelten Fälle in Bezug auf ihre Komplexität und Schwierigkeit erwartungsgetreu eingeschätzt wurden, sodass eine

unterschiedliche diesbezügliche Wahrnehmung der Fälle durch die jeweiligen Gruppen unwahrscheinlich ist. Die Replikation erlaubte weiterhin die Schätzung multivariater meta-analytischer Fixed- und Random-Effects Regressionsmodellen.

Im Fixed-Effects Modell, das auf Grund geringerer Freiheitsgrade einen besseren relativen Fit der geschätzten Parameter aufwies und dessen Annahme multipler ‚wahrer‘ Interventionseffekte methodologisch plausibler ist, zeigten sich durchschnittlich (1) ein signifikanter positiver Effekt auf den Gesamtzuwachs der kognitiven Fähigkeiten im Bereich des Diagnosewissens ($\beta = 1.176$, $p = .002$) und insbesondere auf das Verstehen ($\beta = 1.131$, $p = .002$) sowie (2) auf das diagnostische Selbstkonzept ($\beta = .734$, $p = .019$) und (3) erstaunlicherweise auch auf die Motivation ($\beta = .629$, $p = .035$). Die Messwerte zur Motivation zeigten zwar in beiden einzelnen Analysen einen stärkeren Zuwachs in der Interventionsgruppe, dieser wurde jedoch in den messwiederholten ANOVAs nicht signifikant. Insgesamt korrigiert die Replikationsstudie den Erwartungswert der Interventionseffekte z. T. nach unten, zeigt jedoch auch gemessen am Signifikanzkriterium, dass positive Effekte der Intervention auf die Kompetenzentwicklung angehender Lehrkräfte zu erwarten sind.

- (3) Die in Kapitel 9 dargestellte Studie konzentrierte sich weiter auf die Entwicklung kognitiver Dispositionen und untersuchte, inwiefern die Intervention den Zuwachs des intra-individuellen Zusammenhangs von Diagnosewissen, inklusionsbejahenden Einstellungen und diagnostischem Selbstkonzept während der Interventionslaufzeit fördern kann. Gruppenunterschiede in den Zusammenhangsveränderungen wurden mit dem in Kapitel 5.2 vorgestellten statistischen Modell geprüft. Die Analysen zeigten einen Gruppenunterschied in der Veränderung des Zusammenhangs von Einstellungen und Selbstkonzept ($\beta_{KG} = .57$, $\beta_{IG} = -.08$, $t(67) = 2.104$, $p = 0.039$). Keine Unterschiede zeigten sich dagegen in der Zusammenhangsveränderung von Wissen und Einstellungen ($\beta_{KG} = .53$, $\beta_{IG} = .57$, $t(57) = 0.046$, $p = .964$) wie auch von Wissen und Selbstkonzept ($\beta_{KG} = .93$, $\beta_{IG} = 1.72$, $t(57) = 0.398$, $p = .692$). Insgesamt lässt sich feststellen, dass die Intervention einen mit einer Kontrollgruppe vergleichbaren Zuwachs der Zusammenhänge von Diagnosewissen mit inklusionsbejahenden Einstellungen und dem diagnostischen Selbstkonzept ermöglicht. Dies ist ein Hinweis auf die förderliche Wirkung kooperativen und problemorientierten Lernens auf die Bildung funktioneller Zusammenhänge zwischen kognitiven Modulen, die als notwendige Voraussetzungen kompetenten diagnostischen Handelns gelten. Der Zusammenhang zwischen Selbstkonzept und Einstellungen schrumpfte dagegen in der Interventionsgruppe. Als mögliche Ursache dafür wurden die zusätzlichen Anforderungen in der Interventionsgruppe an multiperspektivische Fallbetrachtung und gemeinsamen Wissensaufbau diskutiert.

- (4) Die in Kapitel 10 dargelegte Studie vertiefte die Frage nach Grundlagen diagnostischer Kompetenz und dem Kompetenzerwerb mit Textfällen weiter. Sie fokussiert zum einen den bislang kaum untersuchten Zusammenhang kognitiver und sozial-kognitiver Grundfähigkeiten mit der diagnostischen Urteilsgenauigkeit angehender Lehrkräfte. Diese Untersuchung ist von Bedeutung für die Bewertung von Urteilsgenauigkeit als Indikator diagnostischer Kompetenz. Ohne Zweifel ist Urteilsgenauigkeit ein bedeutender Faktor effizienten diagnostischen Handelns. Dies lässt aber nicht den direkten Schluss zu, dass sie auch ein geeigneter Indikator für Kompetenz ist. Letztere ist als Konstrukt definiert, das in formaler Bildung entwickelt werden kann. Sollten kognitive und sozial-kognitive Grundfähigkeiten einen substanziellen Effekt auf die Urteilsgenauigkeit haben, müsste die Gültigkeit dieses häufig verwendeten Indikators für Kompetenz mit aller gebotenen Vorsicht in Frage gestellt werden. Dies wurde bislang jedoch nicht ausreichend beforscht. Darüber hinaus adressierte die Studie die Frage, inwiefern sich Textmerkmale und unterschiedlich stark elaborierte mentale Modelle auf die Urteilsgenauigkeit auswirken. Da mentale Modelle in der Expertiseforschung als Maß für fallbezogenes Wissen im Sinne einer kognitiven Disposition für diagnostische Kompetenz verwendet werden, sollten stärker elaborierte mentale Modelle einen größeren Effekt auf die Urteilsgenauigkeit haben, als schwächer elaborierte mentale Modelle.

Insgesamt zeigen die Ergebnisse, dass die manipulierten Textmerkmale keinen Effekt auf die Urteilsgenauigkeit hatten. Dagegen ergab die Analyse einen positiven und substanziellen Effekt von Intelligenz auf akkurate Urteile mit hohem Spezifitätsgrad ($OR = 4.72$, $B = 1.55$, $p = .002$) und von affektiver State-Empathie für akkurate Urteile mit hohem und mindestens niedrigem Spezifitätsgrad ($OR_{hoch} = 3.85$, $B_{hoch} = 1.35$, $p = .009$; $OR_{niedrig} = 2.42$, $B_{niedrig} = 0.88$, $p = .008$) – auch unter Kontrolle der Arbeitsgedächtniskapazität, einem Indikator für das Ausmaß wahrgenommener formaler Bildungsangebote zu Diagnostik und anderen personen- sowie situationsbezogenen Kontrollvariablen. Affektive State-Empathie wirkte in einer Mediationsanalyse als Mediator des positiven Effekts kognitiver Trait-Empathie auf ein Urteil mit hohem Spezifitätsgrad ($c': B = .25$, $p = .047$). Darüber hinaus hatten mentale Modelle mit einem hohen Elaborationsgrad einen positiven Effekt auf richtige diagnostische Urteile mit mindestens niedriger Spezifität ($OR = 2.62$, $B = 0.96$, $p = .010$). Diese Ergebnisse zeigen, dass Urteilsgenauigkeit entgegen der weit verbreiteten Annahme keinesfalls als einziger Indikator für diagnostische *Kompetenz* (im Sinne der Definition von Weinert [1990]) verwendet werden kann, da Zusammenhänge mit kognitiven Basisfähigkeiten bestehen. Zum Kompetenzbegriff, der in Kapitel 2.4 entwickelt wurde und funktionelle Zusammenhänge mit kognitiven Basisfunktionen umfasst, besteht kein Widerspruch. Vorausgesetzt, in der Studie nicht gemessene kognitive Dispositionen für diagnostische Kompetenz haben auch einen Effekt auf die Urteilsgenauigkeit, wie er in den zahlreichen Studien besteht, die nicht für kognitive Grundfähigkeiten kontrollieren, ist insbesondere der Befund zum positiven Effekt von Empathie bedeutsam. Er zeigt, dass Empathie als sozial-kognitive Fähigkeit eine wirksame Disposition in diagnostischen Anforderungssituationen ist. Da Empathie angehender Lehrkräfte trainierbar ist (Whitford & Emerson, 2019) und als State-Konstrukt eine kontextspezifische Aktivierung zeigt, erfüllt das Konstrukt die Anforderungen einer kognitiven

Disposition von Kompetenz. Mit der Studie liegt daher ein deutlicher Hinweis darauf vor, dass bisherigen Modellen diagnostischer Kompetenz der Bezug zu Empathie fehlt. Weitergefasst eröffnet sie die Frage, wie Modelle professionellen Handelns von Lehrkräften sozial-kognitive Fähigkeiten in Zukunft einbeziehen. Darüber hinaus ist der positive Effekt gebildeter Situationsmodelle auf die Urteilgenauigkeit ein erster empirischer Beleg für die theoretischen Modelle diagnostischer Kompetenz von Lehrkräften, die mentale Modellierung als einen zentralen kognitiven Prozess beim Diagnostizieren annehmen.

11.2 Einordnung der Ergebnisse in Interventionsforschung und Theorie zu diagnostischer Kompetenz in der Lehrkräftebildung

Eignet sich kooperativ angelegtes problemorientiertes Lernen mit Textfällen zur Entwicklung von Diagnosewissen in einer frühen Phase des Lehramtsstudiums?

Kooperation hat sowohl im diagnostischen Handeln als auch in der Entwicklung diagnostischer Kompetenz eine zentrale Bedeutung. Die Ergebnisse der Studien in Kapitel 7 und 8 entsprechen dem ausgearbeiteten Forschungsstand zu positiven Effekten von Kooperation auf die professionelle Weiterentwicklung von Lehrkräften (Vangrieken, Meredith, Packer, & Kyndt, 2017) und insbesondere in Form problemorientierten Lernens auf die Entwicklung des Diagnosewissens von Lehramts- und Medizinstudierenden (Walker & Leary, 2009). So fanden Chernikova et al. (2019) in ihrer diesbezüglichen Meta-Analyse mit $g = .58$ einen mittleren positiven Effekt von kooperativ und fall- bzw. problembasiert angelegter Interventionen auf die Entwicklung des Diagnosewissens bei Lehramtsstudierenden. Die Schätzung beruhte auf fünf Studien, von denen zwei textbasierte statt videobasierter Fälle nutzten. Die Analysen der vorliegenden Forschungsarbeit in Kapitel 7 und 8 ergaben kleine bis große Effektstärken ($.28 < g < .90$) mit einem Median von $g = .53$, sodass die Ergebnisse vergleichbar sind. Der Zusammenhang zwischen gemessenem Diagnosewissen und einem validierten Szenariotest für diagnostische Kompetenz liegt unter Kontrolle ähnlicher (Selbstkonzept vs. Selbstwirksamkeitserwartungen) bzw. gleicher Konstrukte (Motivation) etwas höher ($\beta \approx .53$) als in der Analyse von Klug et al. (2013) ($\beta \approx .30$), was wahrscheinlich einerseits darauf zurückzuführen ist, dass in der vorliegenden Studie das Diagnosewissen über deklarative Wissensbestände hinaus auch in Form angewandten Wissens getestet wurde und andererseits darauf, dass mit dem problemorientierten Aspekt der Intervention aus der vorliegenden Arbeit eine größere Ähnlichkeit zwischen Lern- und Testformat bestand, als in der Gelegenheitsstichprobe von Klug et al. (2013).

Klug et al. (2016) fanden für kooperative und textbasierte Fallarbeit mit berufstätigen Lehrkräften im Vergleich zu einer Wartegruppe mittlere bis große Effekte auf den Wissenszuwachs ($.11 < \eta^2_p < .22$), gemessen mit einem Szenariotest. In den vorliegenden Studien mit Lehramtsstudierenden und einer ebenfalls aktiv, aber in einem anderen Format lernenden Kontrollgruppe, fielen die Effekte im selben Szenariotest deutlich kleiner aus ($.03 < \eta^2_p < .08$). Die Studien unterscheiden sich jedoch in der Zielgruppe, der Kontrollgruppe und der konkret angewandten Intervention, sodass ein direkter Vergleich nur vorsichtig interpretiert werden kann.

Diese Ergebnisse weisen darauf hin, dass eine kooperative Fallarbeit in späteren Phasen der professionellen Entwicklung noch ergiebiger sein kann, als sie es in frühen Phasen ist. Einen Befund zur Bekräftigung dieser Hypothese erbringen Kraft et al. (2018) in ihrer Meta-Analyse zu Weiterbildung berufstätiger Lehrkräfte. Ein kooperativ und fallbasiert angelegtes Coaching in Kleingruppen hatte in ihrer Schätzung im Vergleich zu Eins-zu-Eins Coaching einen stärkeren Effekt auf die Instruktionskompetenz im Unterricht ($\beta = .31/- .16$) und die Leistung von Schülerinnen und Schülern ($\beta = .12/- .04$). Wenngleich die Interventionsforschung in der Lehrkräftebildung eine deutliche Bereicherung des institutionellen Lernangebots durch kooperative Fallarbeit nachweist, mangelt es im Bereich der Diagnostik nach wie vor an einem umfassenden und systematischen Vergleich verschiedener Lernformate zu verschiedenen Zeitpunkten der professionellen Entwicklung – insbesondere hinsichtlich zu erwartender Lerngewinne, ihrer Nachhaltigkeit und der notwendigen Lernvoraussetzungen.

Begünstigt problemorientiertes Lernen die Entwicklung theoretischen oder anwendungsbezogenen Diagnosewissens?

Die wohl bedeutsamste und in der empirischen Bildungsforschung am besten untersuchte kognitive Disposition ist das Wissen, d. h. im Gedächtnis gespeicherte Informationen. Die Ergebnisse der vorliegenden Forschungsarbeit zeigen differenzielle Effekte problemorientierten Lernens auf den Wissenserwerb und tragen damit zur Diskussion über die Entwicklung von Diagnosewissen bei. In den Kapiteln 2.2 und 4.1 wurde diskutiert, dass die Expertiseforschung zwischen (1) theoretischem Wissen, das in formaler Bildung aufgebaut werden kann und (2) erfahrungsbasiertem Wissen, das in informellen Bildungsprozessen entlang praktischer Anforderungssituationen erworben wird, unterscheidet (Berliner, 2001; Bromme, 2008). Auch im Modell evidenzbasierten Handelns von Sackett et al. (1996) bilden fallbasiertes Wissen aus praktischen Erfahrungen und Anwendungen sowie deklaratives Wissen die zentralen Säulen kompetenten Handelns. Im Kontext von Diagnostik beschreiben Brush et al. (2017), dass anwendungsbezogenes Wissen von Expertinnen und Experten in Form skriptförmiger Informationen über Zustandsveränderungen von Personen, daran gekoppelte institutionelle Abläufe und kausale Zusammenhänge im konkreten Fall gespeichert ist. Chernikova et al. (2019) bezeichnen letzteres in Anschluss an Polanyi (1966) als prozedurales Wissen und trennen es von strategischem Wissen über Abläufe ab. Meta-Analysen zu problemorientiertem Lernen im Medizinstudium sowie Wissenstests zu pädagogischem und didaktischem Wissen von Lehrkräften unterscheiden zwischen deklarativem und anwendungsbezogenem Wissen, wobei letzteres häufig durch die Lösung fallbasierter Testaufgaben (Vignetten) erfasst wird (König et al., 2018; Kunter et al., 2017).

Die Entwicklung prozeduralen Wissens setzt eine Auseinandersetzung mit Fällen voraus. Die zunehmende Einführung fallbasierter Lernumgebungen in die Lehrkräftebildung schließt erfahrungsbasiertes Lernen nun auch für formale Bildungskontexte auf, was die Art der Lerngewinne in den Fokus rückt. Gijbels, Dochy, van de Bossche und Segers (2005) fanden in ihrer Meta-Analyse zum Vergleich problemorientierten Lernens mit Vorlesungen im Medizinstudium einen kleinen negativen Effekt auf die Entwicklung deklarativen Wissens ($g = -.11$) und einen großen positiven Effekt

auf angewandtes Wissen ($g = .66$). D. h., beide Gruppen gewannen an Wissen, die problemorientiert lernende Gruppe entwickelte jedoch stärker anwendungsbezogenes Wissen. Ähnlich fanden Bergstrom et al. (2016) im Vergleich derselben Lernumwelten keinen Effekt auf den deklarativen Wissenserwerb aber einen großen positiven Effekt der problembasierten Lernumwelt auf anwendungsbezogenes Wissen ($\eta^2_p = .271$).

Für Lehrkräfte ist die Studienlage eher spärlich. So fanden Walker und Leary (2009) in ihrer Meta-Analyse 133 Studien aus dem Medizinstudium und eine Studie aus der Lehrkräftebildung. Borhan (2014) identifizierte in einem Review keine quantitativen Vergleichsstudien zu problemorientiertem Lernen in der Lehrkräftebildung. Auch Chernikova et al. (2019) fanden lediglich fünf Studien zu Lehramtsstudierenden, die zwar sogar Diagnosewissen untersuchten, jedoch keine Differenzierung zwischen deklarativem und prozeduralem Wissen vornahmen und sehr unterschiedliche Arten problembasierten Lernens berichteten. Eine in den o. g. Studien nicht aufgegriffene vergleichende Untersuchung zum Effekt problemorientierten Lernens im Bereich diagnostischer Kompetenz von Lehrkräften von Erdogan und Senemoglu (2017) berichtete keinen differenziellen Effekt auf den Anstieg deklarativen Wissens, aber einen großen Effekt auf anwendungsbezogenes Wissen ($\eta^2_p = .630$). Die Ergebnisse der vorliegenden Studien bestärken diese Befunde, werfen jedoch auch neue Fragen auf.

Die Studie in Kapitel 7 zeigte in Übereinstimmung mit der Forschungslage einen kleinen, nicht signifikanten Effekt der Intervention auf deklaratives Wissen ($\eta^2_p = .042$), aber einen großen Effekt auf anwendungsbezogenes Wissen ($.145 < \eta^2_p < .167$). In der exakten Replikation sanken die Effekte jedoch ab und das Befundmuster drehte sich um, sodass die Analysen einen mittleren, aber nicht signifikanten Effekt auf deklaratives Wissen ergaben ($\eta^2_p = .086$) und kleine bis mittlere, aber ebenfalls nicht signifikante Effekte auf anwendungsbezogenes Wissen ($.019 < \eta^2_p < .052$). Diese Diskrepanz wirft die Frage auf, welche Faktoren im selbstgesteuerten Lernen der Lehramtsstudierenden mit unterschiedlichen Lerngewinnen assoziiert sind. Zukünftige Studien zu problemorientiertem Lernen in der Lehrkräftebildung könnten die Empfehlungen zur Ergebnisdokumentation von Ravitz (2009) nutzen, um spezifische Wirkmechanismen problemorientierten Lernens identifizierbar zu machen. Diese zielen insgesamt auf eine umfassendere theoretische Herleitung und Dokumentation der implementierten Interventionen, insbesondere der Problemtypen in den verwendeten Fällen und der geplanten Bearbeitungsschritte. Die Diskrepanz der Ergebnisse ist jedoch auch ein Hinweis darauf, dass problemorientiertes Lernen als Format selbstgesteuerten Lernens genutzt wird. Gerade in beruflichen Orientierungsphasen wie dem Studium kann selbstgesteuertes Lernen die Reflexion eigener Kompetenzen und Entwicklungsbedarfe sowie die Vorbereitung auf die weitere professionelle Entwicklung und Weiterbildung im Beruf unterstützen (Zlatkin-Troitschanskaia et al., 2015).

Hängt der Erwerb von Diagnosewissen mit einem Zuwachs positiver inklusionsbezogener Einstellungen zusammen?

Diagnostik im schulischen Kontext zielt u. a. darauf, ein Bewusstsein über die Lernvoraussetzungen und eventuellen Lernstörungen von Schülerinnen und Schülern zu gewinnen, um inklusives Lernen in heterogenen Gruppen zu ermöglichen (Ingenkamp & Lissmann, 2008). Die theoretische Diskussion des

Kompetenzkonstrukts in Kapitel 2 ergab dahingehend, dass die Anwendung dispositionaler Wissens- und Fähigkeitsbestände zum Erreichen dieses Ziels nicht automatischer erfolgt, sondern durch handlungsförderliche Überzeugungen, Werte und Einstellungen moderiert wird. Diese werden im funktionellen Ansatz der Kompetenzforschung im Gegensatz zu Motivation als elementarer Bestandteil von Kompetenz betrachtet (Weinert, 2001; Klieme & Hartig, 2006).

Reviews zu förderlichen Bedingungen der Entwicklung inklusionsbejahender Einstellungen von Lehrkräften und Lehramtsstudierenden identifizieren einheitlich Wissen über Lernstörungen als einen begünstigenden Faktor positiver Einstellungsbildung (Mieghem, Verschueren, Petry, & Struyf, 2018; de Boer, Pijl, & Minnaert 2011). Diese Einschätzung unterliegt jedoch deutlichen, aber nicht explizit reflektierten Einschränkungen, da nur zwei der darin rezipierten Studien Korrelationskoeffizienten berichten (Batsiou et al. 2008; Ghanizadeh et al., 2006). Die anderen Studien berichten entweder einen Anstieg positiver Einstellungen durch eine Trainingsmaßnahme oder längere Berufspraxis, ohne darin erworbenes Wissen tatsächlich zu messen (z. B. Avramidis & Kalyva 2007; Lifshitz, Glaubman, & Issawi, 2004; Sari, 2007) oder sie weisen lediglich auf einen deskriptiven parallelen Anstieg beider Konstrukte hin (z. B. Kurniawati et al. 2017; Liu et al., 2016; Mulholland 2016; Youssef & Youssef, 2015). Für die Stärke des Zusammenhangs geben Batsiou et al. (2008) eine Korrelation von $r = .26$ an und Ghanizadeh et al. (2006) berichten mit $r = .23$ ebenfalls einen mittleren positiven Zusammenhang. Beide Studien arbeiteten allerdings mit Selbsteinschätzungen zum Wissen, sodass diese Ergebnisse ebenfalls nur eingeschränkt in Bezug auf die Fragestellung interpretierbar sind, da Selbsteinschätzungen kein geeignetes Instrument zur Erfassung von Wissensbeständen sind (König, Kaiser, & Felbrich, 2012). Die Studie in Kapitel 9 nutzte einen objektiven Wissenstest und zeigte Zusammenhänge ähnlicher Größe mit $r_{KG} = .20$ und $r_{IG} = .26$, diese waren jedoch nicht signifikant. In beiden untersuchten Gruppen stieg dieser Zusammenhang im Verlauf des Wissenserwerbs jedoch an ($\beta_{KG} = .53$, $\beta_{IG} = .57$). Daher kann die Studie ebenfalls nur mit Einschränkungen in der Teststatistik als empirischer Hinweis darauf gewertet werden, dass der Erwerb von Diagnosewissen in einem funktionellen Zusammenhang mit einer positiven Veränderung inklusionsbejahender Einstellungen steht. Insgesamt liegen entgegen der Einschätzungen o. g. Reviews bei genauer Betrachtung nicht ausreichend empirische Untersuchungen vor, um eine gesicherte Aussage über die einstellungsförderliche Wirkung des Erwerbs von Diagnosewissen treffen zu können. Im Hinblick auf die Anforderung des Kompetenzkonstrukts zu funktionellen Zusammenhängen kognitiver Strukturen könnten insbesondere Wachstumskurvenmodelle zu aussagekräftigeren Ergebnissen kommen.

Entstehen durch kooperatives Diagnostizieren kategorial andere Anforderungssituationen als durch individuelles Diagnostizieren?

Definitionen diagnostischer Kompetenz berücksichtigen kooperative Anforderungssituationen bislang entgegen ihrer handlungsbezogenen Relevanz nicht systematisch (Herppich et al., 2018; Klug et al., 2016; Leuders et al., 2018). Sowohl internationale als auch nationale Studien zur beruflichen Situation und dem Handeln von Lehrkräften dokumentieren ein Ressentiment gegenüber professioneller und multiprofessioneller Zusammenarbeit, und zugleich einen diesbezüglichen Weiterbildungsbedarf (OECD, 2018; Richter & Pant, 2016). Die vorliegende Dissertation zeigte darüber

hinaus, dass gezielte Zusammenarbeit im Kontext diagnostischen Handelns sowohl in der Forschung als auch in zugehörigen theoretischen Modellen eine Leerstelle darstellt. So stellte die Auseinandersetzung mit Kompetenz in Kapitel 2 heraus, dass domänenspezifische Anforderungssituationen sowohl Kompetenzen inhaltlich definieren als auch notwendiger Bestandteil von Prozessen der Kompetenzentwicklung sind (Alexander, 2004; Rauner & Zymer, 1991).

Die Fallarbeit in einem kooperativ angelegten Lernangebot hält nicht nur Anforderungen aus dem Inhalt der bearbeiteten Fälle bereit, sondern fordert auch die Bewältigung gruppenbezogener kontextspezifischer Anforderungen heraus, wie u. a. die gemeinsame Zielbildung für das weitere diagnostische und interventive Vorgehen in einem Fall (siehe Kapitel 3.2). Die Interventionsstudien der vorliegenden Forschungsarbeit zeigten, dass eine kontinuierliche Auseinandersetzung mit individuellen *und* kooperativen Anforderungen zu Lerngewinnen in kognitiven Dispositionen für diagnostisch kompetentes Handeln führen. Zum Zeitpunkt der vorliegenden Arbeit konnten keine Testinstrumente ausfindig gemacht werden, die spezifische Dispositionen kooperativen diagnostischen Handelns messbar machen, so dass die Untersuchung dieser Annahme noch mit theoretischen und methodischen Herausforderungen verbunden ist. Zwar kann aus den durchgeführten Interventionsstudien nicht eindeutig empirisch darauf geschlossen werden, dass kooperatives Diagnostizieren ein genuines Set an Anforderungen in Abgrenzung von individuellem Diagnosehandeln umfasst, das Kompetenzentwicklung eröffnet. Diese Möglichkeit kann jedoch vor dem Hintergrund der durchgeführten Studien auch nicht mit Sicherheit ausgeschlossen werden, sodass die Studien in Kapitel 7 und 8 zumindest einen Anlass zur weiteren Untersuchung gruppenförmiger diagnostischer Handlungssituationen sowie der kognitiven Dispositionen, die Lehrkräften darin zu Handlungsfähigkeit verhelfen, darstellen. Inwiefern eine analytische Differenzierung individueller und kooperativer Anforderungen möglich ist, könnte mit Forschungsdesigns untersucht werden, die das Diagnostizieren in Problem Spaces (Newell & Simon, 1972) untersuchen, die systematisch die Interaktion von einerseits Lehrerinnen und Lehrern mit Schülerinnen und Schülern, und andererseits von Lehrkräften untereinander trennen.

Wirken Empathie und Intelligenz auf die textbasierte diagnostische Urteilsgenauigkeit?

Urteilsgenauigkeit gilt als zentraler Indikator diagnostischer Kompetenz. Dabei wird Urteilsgenauigkeit häufig auf Grundlage der Verarbeitung textbezogener Informationen gemessen, z. B. mit dem simulierten Klassenraum (per Mausklick können Textbausteine über ausgewählte Schülerinnen und Schüler aufgerufen werden, so etwa in Kaiser, Helm, Retelsdorf, Südkamp, & Möller, 2012) oder Schülerprofilen (Glock, Krolak-Schwerdt, Klapproth, & Böhmer, 2013). Auch in der Schulpraxis bilden textbasierte Informationen eine wichtige Urteilsgrundlage, z. B. Klassenarbeiten, Hausaufgaben oder Übergangsempfehlungen. Die Ergebnisse der Studie in Kapitel 10 zeigen jedoch – entgegen der impliziten Annahme eines Indikatorverhältnisses zwischen Urteilsgenauigkeit und diagnostischer *Kompetenz* im Sinne von Helmke und Schrader (1989) sowie Herppich et al. (2018) und anderen theoretischen Modellen – dass kognitive Basisfähigkeiten einen Effekt auf die Genauigkeit textbasierter Urteile hatten. Dieser war zudem stärker als der Effekt von Proxyvariablen für Wissen und Erfahrungen hinsichtlich der zu identifizierenden Lernstörung. So hatten in der Studie in Kapitel 10

fluide Intelligenz einen direkten positiven Effekt auf die Urteilsgenauigkeit und kognitive Trait-Empathie hatte einen indirekten positiven Effekt auf die Urteilsgenauigkeit. Die einzige weitere Studie zum Verhältnis von Urteilsgenauigkeit und Intelligenz von Kaiser et al. (2012) zeigt ebenfalls einen positiven Zusammenhang.

Die Ergebnisse der vorliegenden Forschungsarbeit haben zwei zentrale Implikationen. Dies betrifft erstens die Annahme, Urteilsgenauigkeit sei ein besonders geeigneter Kompetenzindikator. Die vorliegenden empirischen Ergebnisse stützen die diesbezüglich kritische Diskussion in Kapitel 3.1. In der Forschung zu Urteilsgenauigkeit von Lehrkräften wurden bislang nur solche Prädiktoren untersucht, mit denen die Indikatorannahme nicht angezweifelt oder widerlegt werden konnte. Dazu gehören Wissen, Motivation, Einstellungen und selbstbezogene Überzeugungen (Südkamp et al., 2012; Machts et al., 2016). Diese Tendenz, nur solche Prädiktoren zu untersuchen, die zur Bestätigung der kompetenztheoretischen Annahme beitragen können, dass Urteilsgenauigkeit ein Indikator für ein latentes Konstrukt sei, das nur durch formale Bildung trainierbar wäre, müsste einer genaueren und umfangreichen Prüfung unterworfen werden. Damit sind Ergebnisse bisheriger Studien, die Urteilsgenauigkeit als Indikator für diagnostische Kompetenz verwendeten, keinesfalls entwertet, da akkurate diagnostische Urteile nichtsdestotrotz eine bedeutsame Richtgröße in diagnostischen Handlungszusammenhängen ist. Es stünde aber zur Debatte, inwiefern es sinnvoll ist, einen Indikator zur Beurteilung einer Kompetenz zu nutzen, wenn dieser auch sensitiv für Unterschiede in kognitiven Dispositionen ist, die nicht ohne Weiteres im Rahmen der Lehrkräftebildung trainierbar sind. Möglicherweise könnten weitere Indikatoren ergänzend herangezogen werden. Alternative Performanzen zu Urteilsgenauigkeit werden bereits diskutiert (Südkamp & Praetorius, 2017).

Zweitens ist das Ergebnis der Studie ein Hinweis darauf, dass sozial-kognitive Fähigkeiten eine größere Bedeutung für kompetentes diagnostisches Handeln haben, als ihnen in bisherigen theoretischen Modellen zukommt. Weder die in Kapitel 3.1 vorgestellten Modelle diagnostischer Kompetenz, noch das häufig angewandte COACTIV-Modell beinhalten sozial-kognitive Fähigkeiten. Kim (2017) stellt in einer Befragung von Lehrkräften fest, dass diese jedoch explizit Empathie als Begründung von Unterrichts- und Förderentscheidungen anführen. Empathie ist auch von Bedeutung für die Prävention, Diagnose und Intervention sozialer Probleme und Lernstörungen. Lehrerinnen und Lehrer mit höheren Empathiewerten intervenieren mit höherer Wahrscheinlichkeit bei Bullying (Huang, 2018) und können traditionelles Bullying und Viktimisierung mit höherer Wahrscheinlichkeit erkennen sowie entsprechende Interventionen einleiten (Tsoulopas et al., 2014). Darüber hinaus finden Ha und Park (2014) sowie Park und Hwang (2013) positive Effekte von stärker ausgeprägten Empathiefähigkeiten bei Lehrkräften auf die Durchführung von Interventionen mit Schülerinnen und Schülern mit ADHS sowie auf das eigene Copingverhalten in Bezug auf stressauslösende Unterrichtssituationen im Kontext von inklusiven Klassen mit Schülerinnen und Schülern mit ADHS. Weiterhin stellen Hellmann (2015) und Knutson-Miller (2001) fest, dass kognitive Empathie im Zusammenhang mit der differenzierten Wahrnehmung fachbezogener Schülerinnen- und Schülervorstellungen und der Unterrichtsqualität steht. Die Ergebnisse der Studie in Kapitel 10 unterstützen den skizzierten Nutzen von Empathiefähigkeiten als kognitive Dispositionen für wirksames Unterrichtshandeln und Diagnostizieren.

Die Studienlage zu Empathiefähigkeiten von Lehramtsstudierenden ist jedoch vergleichsweise ernüchternd. Tettegah und Anderson (2007) stellten fest, dass untersuchte Lehramtsstudierende in einer frühen Phase ein geringes Maß an Empathiefähigkeiten zeigten. Zusammengenommen mit den Ergebnissen zum positiven Effekt von Empathiefähigkeiten auf akkurate diagnostische Urteile, ist damit Anlass für diesbezügliche Interventionsforschung in der Lehrkräftebildung gegeben, wie auch für die weitere und grundlegende Diskussion der Bedeutung sozial-kognitiver Fähigkeiten für Modelle professionellen Handelns von Lehrkräften.

Wie wirken mentale Modelle auf die textbasierte diagnostische Urteilsgenauigkeit?

Die Ergebnisse der Studie in Kapitel 10 zeigen, dass mentale Modelle diagnostischer Situationen mit hohem Elaborationsgrad einen Effekt auf die diagnostische Urteilsgenauigkeit von Lehramtsstudierenden hatten. Dieser Befund schließt direkt an die Ergebnisse der Expertiseforschung zum Zusammenhang von mentaler Modellierung und kompetentem Handeln an (siehe Kapitel 4.1). Zur Einordnung der Ergebnisse ist jedoch die Entstehung dieser mentalen Modelle zu berücksichtigen. So bildeten die Untersuchungspersonen in der Studie in Kapitel 10 ihre kognitiven Repräsentationen spontan und nicht vor dem Hintergrund mehrjähriger berufspraktischer Erfahrung. Bei Expertinnen und Experten besteht ein Zusammenhang zwischen mentalen Modellen und fallbasiertem Wissen (siehe Kapitel 4.1), während Novizinnen und Novizen nur eingeschränkt auf fallbezogenes Wissen zurückgreifen können. Ein stärkerer Zusammenhang mentaler Modelle mit fallbezogenem Wissen könnte bspw. die Aktivierung funktioneller heuristischer Urteilsprozesse wie im Modell von Brush et al. (2017) unterstützen und die Akkuratheit von Urteilen fördern. Urteilsheuristiken von Lehrkräften sind bislang jedoch noch nicht untersucht. Die Ergebnisse der Studie in Kapitel 10 sind zum Zeitpunkt der Anfertigung der vorliegenden Dissertation der erste quantitative empirische Hinweis darauf, dass Modelle diagnostischer Kompetenz, in denen mentale Modelle zu Fähigkeiten und Schwierigkeiten von Schülerinnen und Schülern als zentrale Kompetenzfacette konzipiert sind (z. B. Ostermann, 2018), bereits in frühen Phasen der Lehrkräftebildung angewendet werden können.

Quantitative Interventionsforschung zu Kompetenzentwicklung braucht klare Evaluationskriterien.

Mit dem Ziel, die Bildungschancen und das Bildungsniveau auf gesellschaftlicher Ebene zu steigern, sucht die Forschung zur Lehrkräftebildung kontinuierlich nach Möglichkeiten der Qualitätsverbesserung der Ausbildung von Lehrerinnen und Lehrern (Bauer & Prenzel, 2012). Das Konstrukt der Kompetenz hat dabei die Bedeutung einer Regel- und Stellgröße bekommen, auf Grund derer – neben anderen Faktoren – Ausbildungsstrukturen verändert und konsolidiert werden (Zlatkin-Troitschanskaia et al., 2016). Interventionsforschung stellt die dafür nötige Evidenz durch die Beurteilung der Effektivität von Maßnahmen für Kompetenzentwicklung her. Erwartungen an die Qualität von Interventionsstudien müssten daher besonders hoch sein. Tatsächlich existieren auch eine Reihe anwendbarer Standards für Interventionsstudien in der Lehrkräftebildung. Dazu gehören die APA Reporting Standards (American Psychological Association; Applebaum, Cooper, Kline, Mayo-Wilson, Nezu, & Rao, 2018), die sehr ähnlichen und an den medizinischen Kontext angelehnten

CONSORT-SPI 2018 (Consolidated Standards of Reporting Trials for Social and Psychological Interventions; Grant et al., 2018), die wesentlich kürzere WIDER Checklist (Workgroup for Intervention Development and Evaluation Research; Albrecht, Archibald, Arseneau, & Scott, 2013) und die SPR Standards (Society of Prevention Research, Gottfredson et al., 2015). Letztere differenzieren zwischen Standards für experimentelle Forschung (Efficacy Standards), quasi-experimentelle Forschung (Effectiveness Standards) und Implementationsprozesse (Dissemination Standards). Obwohl das Konstrukt der Kompetenz deutliche theoretische Unterschiede zu anderen lernbezogenen Outcomes hat, thematisiert keiner dieser Standards diese Besonderheiten. Eine exemplarische Diskussion verdeutlicht die Notwendigkeit eines entsprechenden Bewusstseins oder entsprechender Erweiterungen, die in dieser Forschungsarbeit auch nur ansatzweise umgesetzt werden konnten.

Erstens ist eine Anforderung des häufig angewendeten Kompetenzbegriffs im Sinne durch formale Bildung erlernbarer kognitiver Dispositionen (Weinert, 2001; Klieme & Hartig, 2006) die Unabhängigkeit von kognitiven Basisfähigkeiten (kritisch dazu siehe Kapitel 2.4). Diese Annahme wird jedoch aus forschungspragmatischen Gründen in nahezu keiner Interventionsstudie geprüft, was auch die Interventionsforschung der vorliegenden Dissertation einschließt (bspw. keine der Studien in Chernikova et al., 2019). Nicht für den Varianzanteil kognitiver Grundfähigkeiten zu kontrollieren, erhöht deutlich das Risiko einer Überschätzung diesbezüglicher Effekte oder auch vollständiger Fehler erster Art, d. h. Effekte statistisch zu finden, die real nicht gegeben sind. Ein möglicher Standard zur Qualitätsbeurteilung wäre die Anführung von direkten (Messung) oder indirekten Belegen (empirische Vergleichsstudien) für die Unabhängigkeit der gemessenen Konstrukte von kognitiven Basisfähigkeiten.

Zweitens existiert eine Annahme über die Struktur von Kompetenz in der Organisation kognitiver Module in funktionellen Zusammenhängen (Weinert, 1999). Eine Aussage über Kompetenzentwicklung zu treffen ist daher strenggenommen nur dann möglich, wenn entweder (1) eine Steigerung von kognitiven Dispositionen bei konstanten funktionellen Zusammenhängen zu verzeichnen ist (Messinvarianzkriterium) oder (2) der funktionelle Zusammenhang der Dispositionen steigt (Zusammenhangskriterium [1]) oder (3) der Zusammenhang dieses latenten Zusammenhangs mit verhaltensbezogener Leistung, der Performanz, steigt (Zusammenhangskriterium [2]). Dies bedeutet, dass mit entsprechenden Aussagen ein enormer Aufwand für die Datenerhebung und statistische Modellierung verbunden ist. Dementsprechend begnügt sich die Mehrzahl kompetenzbasierter Interventionsstudien mit der Veränderungsmessung einzelner kognitiver Dispositionen und beansprucht die Gültigkeit der allgemeinen theoretischen Annahmen zu Kompetenz, ohne zu prüfen, ob die gesammelten Daten diesen Annahmen entsprechen. Ein möglicher Standard zur Qualitätsbeurteilung wäre die transparente Darlegung der optimalen Modellierung und der Schritte, mit denen diese eingeschränkt wurde, bspw. in einem Flussdiagramm (Levin, 1994).

Drittens variieren Interventionseffekte in und besonders außerhalb experimenteller Settings. Kapitel 5.1 beschreibt die große Bedeutung von Replikationsstudien für eine beurteilbare Schätzung von Effekten. Der Großteil der publizierten Interventionsstudien in der Lehrkräftebildungsforschung (z. B. alle in dieser Forschungsarbeit zitierten Studien) legt keine Daten zu Replikationen vor. Die Ergebnisse der Studie 8 in dieser Forschungsarbeit zeigen jedoch, dass gerade bei quasi-experimentellen Designs, die in der Lehrkräftebildung das oftmals einzig umsetzbare Design sind,

bedeutsame und interpretationsbedürftige Unterschiede zwischen Original- und Replikationsstudien entstehen können.

Weitere prüfbare Annahmen des Kompetenzkonstrukts bestehen in der Definition über Anforderungssituationen (direkt experimentell manipulierbar oder indirekt durch Vergleichsstudien beschreibbar) und der Domänen- bzw. Kontextspezifität. Da bislang keine angepassten Standards zur Erforschung von Maßnahmen zur Kompetenzentwicklung vorhanden sind, variiert die Qualität der verfügbaren Interventionsstudien enorm, was sowohl darauf aufbauende Forschung vor Herausforderungen stellt als auch die institutionelle Entscheidungsbildung für Strukturveränderungen der Lehrkräftebildung. Von einer rein signifikanzbasierten Qualitätsbeurteilung wurde bereits deutlich abgeraten (Sedlmeier, 1996; Press & Harris, 1994), was die Entwicklung von Standards der Interventionsforschung in der Lehrkräftebildung umso dringlicher erscheinen lässt.

11.3 Grenzen und Stärken der Forschungsarbeit

Die Einordnung der Forschungsergebnisse verdeutlicht, dass der Interpretation der Befunde aus den Studien der Kapitel 7 bis 9 auch Limitationen gesetzt sind. Bedingt durch das Design der Studien und der entwickelten Intervention lassen sich die Befunde weder direkt auf andere Formen der Lehramtsausbildung übertragen noch auf alle Inhaltsbereiche der Diagnostik. Auch durch die verfügbaren und verwendeten Messinstrumente sowie die vorliegende Datenstruktur sind den Erkenntnissen der Forschungsarbeit Grenzen gesetzt. Deren Diskussion ermöglicht eine präzisere Interpretation und kann gemeinsam mit einer Darstellung der Stärken eine Grundlage weiterführender Forschung bilden.

Eine erste Grenze stellt *die entwickelte Intervention* selbst dar. In dieser wurden nur bestimmte Lernvoraussetzungen und -störungen sowie diagnostische Vorgehensweisen behandelt, wobei sich die Auswahl einerseits an den (hohen) Prävalenzraten und andererseits an gesetzlichen Richtlinien und Empfehlungen orientierte. So fordern die KMK-Standards bspw. explizit „kooperieren bei der Diagnostik“ als Studieninhalt ein (2004, 12). Von einer allgemeingültigen Definition der Inhalte diagnostischen Handelns und diagnostischer Kompetenz von Lehrkräften sind Forschung und Praxis aber weit entfernt. Spinath (2005) schlug infolge sehr uneinheitlicher Ergebnisse in der Einschätzung lernrelevanter Konstrukte durch Lehrkräfte vor, von Kompetenzen im Plural, statt von Kompetenz im Singular zu sprechen. Auch aktuelle Studien zeigen in den Einschätzungen von Schülerinnen und Schülern große Unterschiede zwischen Lehrkräften sowie ‚innerhalb‘ von Lehrkräften zwischen verschiedenen Konstrukten (Südkamp et al., 2012; Machts et al., 2016). Zwar geben die Richtlinien der KMK (2004, 2015) der Lehrkräftebildung einen Rahmen, dieser ist jedoch recht offen formuliert, sodass Unterschiede in den Ausbildungsinhalten zu erwarten und je nach lokalen Gegebenheiten auch angemessen sind (z. B. regional gehäuft auftretende Lernstörungen). Durch die Überarbeitung des DSM (Schulte-Koerne, 2014) sind Lehrkräfte nun auch stärker in der Diagnostik ganz bestimmter Lernstörungen – Entwicklungsstörungen schulischer Fähigkeiten – gefordert. Es existiert eine Reihe an Inhalten, die in der entwickelten Intervention nicht thematisiert wurden (z. B. Erstellung von Gutachten, relevante institutionelle Strukturen), sodass keine Aussage darüber getroffen werden

kann, inwiefern kooperatives und problemorientiertes Lernen mit textbasierten Fällen auch für diese Inhalte ein geeignetes Lernformat darstellt.

Die Fallkonstruktion und die Durchführung der Intervention folgten einem strengen und dokumentierten Protokoll (Wedel, 2019), sodass exakte Replikationen möglich waren und sind. Darüber hinaus wurden die konstruierten Fälle von Lehramtsstudierenden und Lehrkräften mit diagnostischer Expertise in einer zusätzlichen Studie erfolgreich validiert. Die Lehramtsstudierenden ($n = 12$) zeigten eine hohe bis sehr hohe inter-individuelle Übereinstimmung in der Bewertung der wahrgenommenen Schwierigkeit der Fälle ($ICC_{3k} = .85$), der erlebten Immersion ($ICC_{3k} = .84$) und der wahrgenommenen Komplexität der verfügbaren Informationen im Fall ($ICC_{3k} = .91$). Auch der intra-individuelle Zusammenhang von Informationskomplexität und Schwierigkeit im Längsschnitt der Intervention, der entsprechend der Cognitive Load Theorie (Sweller, 1994) bestehen sollte, konnte über vier Messzeitpunkte hinweg durch ein Experience Sampling festgestellt werden ($r_{rm} = .80$, $p < .001$). Somit konnte ein gewichtiger Varianzfaktor in der Intervention kontrolliert werden. In einer weiteren Validierungsstudie schätzten Lehrkräfte ($n = 5$) mit Arbeitsschwerpunkt Sonderpädagogik die Fälle hinsichtlich der Realitätsnähe ($.83 < \alpha < .95$) und Plausibilität ($.88 < \alpha < .99$) ein. Die Realitätsnähe wurde mit sehr hoher Übereinstimmung ($ICC_{3k} = .76$) auf einer vierstufigen Skala als hoch bis sehr hoch eingeschätzt ($3.03 < M < 3.69$; $.26 < SD < .62$). Ähnlich wurde die Plausibilität der Fälle mit hoher Übereinstimmung ($ICC_{3k} = .63$) auf einer vierstufigen Skala als hoch bis sehr hoch eingeschätzt ($3.06 < M < 3.78$; $.40 < SD < .68$). Die entwickelten Fallvignetten spiegeln also realitätsnah diagnostisch relevante Situationen aus dem Schulkontext wider. Da sich die Intervention an die klassischen Schritte problemorientierten Lernens hält (Hmelo-Silver, 2004), ist ebenfalls eine gute Vergleichbarkeit mit anderen Studien gegeben, die diese didaktische Methode anwenden.

Durch das *Experimentaldesign* der Studie in Kapitel 10 konnte systematisch die Wirkung von Textmerkmalen auf die textbezogene Informationsverarbeitung und Urteilsbildung untersucht werden. Die Untersuchung von Eigenschaften und Wirkungen des in Interventionen eingesetzten Fallmaterials und insbesondere von Textfällen erfolgte bislang häufig über die Kontrastierung text- und videobasierter Fälle und in quasi-experimentellen Studiendesigns (Bohl et al., 2016; Frommelt et al., 2019; Syring et al., 2015). Diese Studien fanden durchgängig Unterschiede in kognitiven Prozessen während der Fallbearbeitung und in den Lernergebnissen aus der Fallarbeit. Daher liegt die Vermutung nahe, dass Einschätzungen von Lehrkräften auf Grundlage von Verhaltensbeobachtungen z. T. andere Dispositionen und Prozesse zugrunde liegen als Einschätzungen und Urteilsbildungen auf Grundlage textbasierter Informationen. Insofern sollten die Studienergebnisse nicht ohne Weiteres hinsichtlich der Urteilsprozesse abseits textbasierter Einschätzungen interpretiert werden. Weiterhin bestand das experimentelle Material nicht aus Originaltexten von Schülerinnen und Schülern, sodass die Ergebnisse nicht auf diese unterrichtstypische Situation übertragen werden können. Das Experimentaldesign greift vielmehr die Situation einer textbasierten Fallbearbeitung und diagnostischen Urteilsbildung durch angehende Lehrkräfte auf.

Eine zweite Grenze sind die *psychometrischen Messungen und Messinstrumente*. In den Interventionsstudien erfolgte eine Messung des Diagnosewissen, des diagnostischen Selbstkonzepts, der Motivation und weiterer Konstrukte, die etwa 40 Minuten in Anspruch nahm. Die Messwerte mussten auf Grund der Stichprobengröße in manifester Form modelliert werden, sodass Messfehler

aus der Interaktion von Personen mit den Messinstrumenten nicht explizit geschätzt werden konnten. Darüber hinaus erfolgte nur in der ersten Studie eine performanzbasierte Messung diagnostischer Kompetenz mit einem Szenariotest (Klug et al., 2013). Eine weitere Einschränkung betrifft die Messung von Fähigkeiten im kooperativen diagnostischen Handeln. Diese wären z. B. durch Verhaltensbeobachtungen oder zusätzliche Befragungen zwar prinzipiell möglich gewesen. Jedoch existieren zum einen keine diesbezüglichen objektiven Messinstrumente, die spezifisch für den diagnostischen Kontext konstruiert wurden. Zum anderen verdeutlicht der Forschungsstand die große Bedeutung der gewählten Konstrukte für diagnostische Kompetenz (Klug et al., 2016; Machts et al., 2016), sodass die Entscheidung auf die Messung dieser Konstrukte fiel. Eine möglicherweise bedeutsame Variable, die Lernzeit außerhalb der Intervention, wurde ebenfalls nicht kontrolliert, was in der Interpretation der Befunde zu berücksichtigen ist. Dies wurde überraschenderweise auch in anderen einschlägigen Interventionsstudien bislang nicht umgesetzt (z. B. Klug et al., 2013; Kumschick et al., 2017; Seidel et al., 2011; Syring et al., 2015). In den vorliegenden Interventionsstudien fand zwar ein Experience Sampling statt, das sich jedoch auf das Erleben der Fälle konzentrierte. Eine anekdotische Beobachtung war, dass die Studierenden auch über die Kohorten hinweg auf Nachfrage hin relativ konstant von etwa einer Stunde Recherche- und Lernzeit pro Woche berichteten. Zukünftige Interventionsstudien in der Lehrkräftebildungsforschung könnten Interventionseffekte durch Kontrolle dieser Variable noch präziser schätzen.

Auf der anderen Seite zeigten die verwendeten Messinstrumente gute bis sehr gute Reliabilitätswerte und es wurde eine Reihe objektiver Testverfahren genutzt, sodass die Messwerte mit hoher Wahrscheinlichkeit verlässlich sind. Dazu gehörten der Test zum pädagogischen Diagnosewissen (Wedel & Müller, 2016), ein Rekognitions- und Verifikationstest zur Messung mentaler Modelle (Wedel & Müller, 2017), ein Test für fluide Intelligenz von Bilker et al. (2012) sowie ein konstruierter Test zur Messung der Arbeitsgedächtniskapazität in Anlehnung an Wilhelm, Hillenbrand und Oberauer (2013). Spätestens nachdem König, Kaiser und Felbrich (2012) niedrige Korrelationen von Selbsteinschätzungen durch Lehramtsstudierende und Performanz in einem Wissenstest feststellten, können Selbsteinschätzungen zu Wissen und Kompetenzen nicht mehr im Sinne von Indikatoren für die eigentlich zu messenden Konstrukte verstanden werden. Der eingesetzte Test zum Diagnosewissen korrelierte sehr hoch mit dem Szenariotest von Klug et al. (2013), was als Hinweis auf konvergente Validität des Wissenstests verstanden werden kann.

Eine dritte Grenze bilden *die Stichproben und Analyseverfahren*. Aus inhaltlicher Perspektive ist in der Ergebnisinterpretation zu berücksichtigen, dass in den Interventionsgruppen Lehramtsstudierende aus der beruflichen Bildung teilnahmen und in den Kontrollgruppen Studierende für das Lehramt im Fach Wirtschaft-Arbeit-Technik (Sekundarstufe I). Eine randomisierte Verteilung der Studierenden auf Interventions- und Kontrollgruppe war aus studienorganisatorischen und projektbezogenen Gründen nicht möglich. Zwar ergaben die statistischen Analysen keine Unterschiede der Gruppen in den veränderungsrelevanten Konstrukten, es lässt sich jedoch nicht ausschließen, dass bspw. systematische Unterschiede im Lernstil oder anderen nicht erfassten aber möglicherweise relevanten Kontrollvariablen existierten. Der Bias, der einer institutionsspezifischen Durchführung einer Intervention entspringt, ließe sich durch eine engere Kooperation mit anderen Standorten der

Lehrkräftebildung reduzieren. Dementsprechend steht das Interventionsprotokoll mit allen Materialien als Open Educational Resource zur Verfügung (Wedel, 2019).

Die Interventionsstudien aus den Kapiteln 7 bis 9 weisen eine kleine Stichprobengröße auf, die zwar für Interventionsforschung nicht unüblich ist (Chernikova et al., 2019), jedoch Einschränkungen in den statistischen Analysen nach sich zieht. So ergab die Replikationsstudie zwar mittlere Effektgrößen, diese waren in der Varianzanalyse jedoch nicht signifikant. Da eine exakte Replikation jedoch auch die Anwendung der exakt selben Analysemethoden der zu replizierenden Studie verlangt, konnte hier nicht auf möglicherweise geeignetere Verfahren (z. B. robuste Schätzer) zurückgegriffen werden. Auch ließ die Stichprobengröße nicht die Berechnung latenter Wachstumskurvenmodelle zu, die jedoch auf Grund einer geringeren Wahrscheinlichkeit des Auftretens von Fehlern erster Art besser zur Überprüfung von Zusammenhangsveränderungen (als Kriterium von Kompetenzentwicklung) geeignet sind als multiple Zusammenhangsanalysen. Darüber hinaus konnten nur jeweils zwei Messzeitpunkte realisiert werden, sodass keine nicht-linearen Veränderungen prüf- und abbildbar waren, obwohl sie Kompetenzentwicklung z. T. realitätsgetreuer darstellen können (Alexander, 2004). Die Realisierung weiterer Messzeitpunkte wären auch dahingehend wünschenswert, dass eine Beurteilung der Stabilität von Lerngewinnen aus der Intervention erfolgen könnte. So stellen etwa Buholzer und Zulliger (2013) differenzielle Verläufe der Entwicklung verschiedener Kompetenzfacetten im Studienverlauf angehender Lehrerinnen und Lehrer fest. Auch könnte dadurch untersucht werden, inwiefern kooperativ angelegtes Lernen den Erhalt von Wissensbeständen beim Übergang in eine kooperative Praxis während des Praxissemesters oder in der zweiten Phase der Lehramtsausbildung stärker fördert als dies Lernformen mit Fokus auf individuelle Aktivitäten tun. Weiterhin lag durch die Kleingruppenarbeit eine hierarchische Datenstruktur mit einer Nestung von Messzeitpunkten in Studierenden in Kleingruppen in den Quasi-Experimentalgruppen in den Kohorten vor, was bedingt durch die Größe der Stichprobe nicht statistisch modelliert werden konnte. Zwar könnte für eine genaue Beurteilung der statistischen Voraussetzungen ein mehrebenenanalytisches Nullmodell zum Vergleich des Effekts der Varianzkomponenten herangezogen werden, jedoch würden die dafür minimal bestehenden Freiheitsgrade bei der Schätzung des Modells für nur ein gemessenes Konstrukt bereits die Stichprobengröße in den Kleingruppen ($n \approx 5$) übertreffen, sodass die Schätzungen einen erheblichen Bias hätten. Generell wäre zur Berücksichtigung dieser fünf (!) Ebenen eine Stichprobe notwendig, welche die Studierendenzahlen einer jeden Universität überträfe. Eine Ziehung weiterer Universitäten würde weitere Ebenen in das Modell einführen etc. sodass diese Modellierung auch prinzipiell nur in Simulationsstudien und nicht in der Praxis umgesetzt werden kann.

Den methodischen Einschränkungen wurde durch die Anwendung meta-analytischer und eigens entwickelter statistischer Verfahren begegnet (Kapitel 5). So erfolgte durch die Anwendung eines multivariaten meta-analytischen Regressionsmodells zumindest eine Schätzung der kohortenspezifischen Varianzkomponente der Interventionsstudien unter Kontrolle der Korrelationen zwischen den gemessenen Konstrukten (Berkey et al., 1998). Da das Diagnosewissen und das diagnostische Selbstkonzept eine substantielle Korrelation aufwiesen und sich die jeweilige Varianz zwischen den Kohorten unterschied, trug diese Analyse zur Präzision der Ergebnisse bei. Die Entwicklung eines Verfahrens zur Berechnung von Zusammenhangsveränderungen über die Zeit mit kleinen Stichproben stellt einen Beitrag zur Interventionsforschung dar, der in bereits geplanten

Simulationsstudien weiter ausgebaut wird. Damit konnte eine Alternative zur Schätzung latenter Wachstumskurvenmodelle erschlossen werden, mit der die Prüfung eines zentralen Kriteriums von Kompetenzentwicklung auch mit für Interventionsforschung üblichen Stichprobengrößen möglich ist.

11.4 Anschließende Fragestellungen und zukünftige Forschung

Die Befunde und Grenzen der vorliegende Forschungsarbeit werfen eine Reihe anschließender Fragestellungen auf, die bislang noch nicht untersucht wurden. Der in Kapitel 3.2 herausgearbeiteten Bedeutung von inner- und interprofessioneller Kooperation im diagnostischen Handeln von Lehrkräften steht bisher eine größere Forschungslücke in Bezug auf die *institutionelle Ebene* der Diagnostik im Schulwesen gegenüber. Gerade die Diskrepanzen in den Ergebnissen der durchgeführten Interventionsstudien weisen darauf hin, dass Interaktions- und Kommunikationsmuster von Bedeutung für die Ergebnisse des kooperativen diagnostischen Handelns sind. Es schließt sich die Frage an, inwiefern unterschiedliche Verhaltensmuster auf Gruppenebene auffindbar sind, ob diese einen systematischen Effekt auf die Effektivität von Diagnostik haben, welche Faktoren zur Bildung bestimmter Kommunikations- und Interaktionsmuster führen und inwiefern diese mit positiven Effekten auf die schulische Entwicklung und Leistung von Schülerinnen und Schülern zusammenhängen. Entsprechende Muster könnten aus in Schulen gewonnenen Gruppendiskussionen Gruppen mit der dokumentarischen Methode als explizite und konjunktive Kommunikationsstrukturen herausgearbeitet und mit Typenbildung klassifiziert werden, um anschließend in quantitativen Untersuchungen, idealerweise mit einem Längsschnittdesign, ihre Wirkung auf Lern- und Leistungsmerkmale von Schülerinnen und Schülern zu prüfen.

Auf *individueller Ebene* stellt die Forschungsarbeit dar, dass bereits ein breiter Forschungsstand zum Diagnosewissen besteht. Auch die Untersuchung relevanter Wahrnehmungsvorgänge im Rahmen der Signaldetektionstheorie schreitet bereits voran (Oudman, van de Pol, Bakker, Moerbeek, & van Gog, 2018). Die Ergebnisse der Studie in Kapitel 10 weisen darauf hin, dass es Aspekte diagnostischer Urteile und ihrer Bildung gibt, die noch weiterer Untersuchung bedürfen. Die Befunde zu positiven Effekten von Empathie auf textbasierte Urteile legen nahe, dass während der Urteilsbildung neben systematischer Informationsverarbeitung noch weitere kognitive Prozesse ablaufen. Interessant sind hier insbesondere Urteilsheuristiken (Gigerenzer & Goldstein, 1996), die in der Medizin bereits Gegenstand von Modellen diagnostischer Kompetenz sind (Brush et al., 2017), in Bezug auf das diagnostische Handeln von Lehrkräften jedoch bislang nicht erforscht wurden. Da Lehrkräften im Unterricht möglicherweise eher spontane als systematische Urteile abverlangt werden, könnte die Identifikation funktionaler diagnostischer Heuristiken auch zur Erklärung der großen Unterschiede in der Urteilsgenauigkeit von Lehrkräften beitragen, die in der Forschung trotz immer zuverlässigerer Messung relevanter kognitiver Dispositionen wiederholt festgestellt werden (Kaiser et al., 2012; Spinath, 2005; Südkamp et al., 2012).

Darüber hinaus verweisen allgemeine Kompetenzmodelle zwar auf die prinzipielle Funktion von Überzeugungen und Motivation zur Ziel- und Volitionsbildung (Baumert & Kunter, 2006; Weinert, 1999), wann genau sie im diagnostischen Prozess auf welche Art und Weise und in welchem Umfang eine Rolle spielen, ist bislang jedoch noch nicht eingehend untersucht. Eine Möglichkeit der

Erforschung ihrer Bedeutung besteht in der Modellierung mentaler Zustände, die während des diagnostischen Handelns typischer- oder idealerweise auftreten. Damit werden schemaartige kognitive Aktivierungsmuster bezeichnet, die in Verbindung mit verhaltensbezogenen Anforderungen und Zielen auftreten (Haynes & Rees, 2006). Wären typische mentale Zustände beim Diagnostizieren bekannt und mit geringem Aufwand messbar, könnte weitere Forschung daran ansetzen, förderliche und hinderliche Faktoren für das Erreichen zielführender mentaler Zustände zu identifizieren. Eine zentrale Herausforderung liegt dabei in der Entwicklung geeigneter Messverfahren und Instrumente, die in der Forschung im medizinischen Kontext bereits begonnen hat (Lajoie & Gube, 2018). In Anschluss an die vorliegende Forschungsarbeit startete diesbezüglich bereits ein Projekt (TUB Teaching 2.0: Diagnostik), das auf die Analyse von Log-Files aus der Bearbeitung speziell konstruierter diagnostischer Simulationsumgebungen mit latenten Markov-State-Trait Modellen hinarbeitet.

11.5 Implikationen für die Praxis der Lehrkräftebildung

In der Ausbildung von Lehrerinnen und Lehrern gewinnt fallbasiertes Lernen zunehmend an Bedeutung (Chernikova et al., 2019). Aus den Ergebnissen der vorliegenden Forschungsarbeit können hochschuldidaktische Implikationen hinsichtlich der *Verwendung* von Textfällen und der *Instruktion* während der Arbeit mit Textfällen abgeleitet werden.

Universitäre Lernangebote können Lehrkräften grundlegende Wissensbestände und Fähigkeiten für kompetentes Handeln im Bereich der Diagnostik vermitteln, wenn sie eine kognitiv anregende Umwelt bereitstellen. Eine wichtige Entscheidung in Instructional Design Modellen betrifft die zielbezogene Auswahl oder Erstellung des Lern- und Arbeitsmaterials (van Merriënboer et al., 2002). Mit den Ergebnissen der Replikationsstudie in Kapitel 8 können theoretisch konstruierte Textfälle als geeignetes und effektiv adaptier- und herstellbares Lernmaterial zur Aneignung diagnostischer Wissensbestände empfohlen werden. Insbesondere in Anbetracht des „Digitalisierungswahns“ (Blomert, 2017) ermutigen die Befunde dieser Arbeit zum Erhalt medialer Pluralität. Konstruierte Textfälle sind im Vergleich zu videobasierten Fällen weder in der Erstellung noch in der Verwendung mit besonderen Anforderungen in Bezug auf Datenschutz, technische Ausstattung oder Schauspielpersonal verbunden. Da die Arbeit mit Videofällen spezifisch positive Effekte auf die unterrichtsbezogenen Wahrnehmungsfähigkeiten hat (Seidel et al., 2011), können diese gut als Ergänzung zu textbasierter Fallarbeit verwendet werden.

Problembasiertes Lernen führt mit der Offenheit für eigene Lernziele ein Element der Unbestimmtheit in den Lehr- und Lernprozess ein, in dem Studierende Verantwortung für ihren eigenen Lernprozess übernehmen. Wenngleich selbstgesteuertem Lernen in Vorbereitung auf die professionelle Weiterentwicklung eine wichtige Rolle zukommt, sollte die tertiäre Bildung auch das Erreichen von Standards der Lehrkräftebildung sicherstellen (KMK, 2004). Eine dementsprechende Anforderung in der Arbeit mit problemorientiertem Lernen liegt in der Lernbegleitung. So beschreiben Evans und Tylor (1996), welche institutionellen Veränderungsprozesse an einer ausgewählten Universität bei der Umstellung auf ein Curriculum mit hohem Anteil an problembasiertem Lernen notwendig waren und stellen Personalentwicklung als einen erfolgskritischen Punkt heraus. Auch

Barrows (1986) betont in einem Review die hohen Anforderungen an meta-kognitive und kommunikative Fähigkeiten der Lernbegleitenden. Im Prozess der Implementation der entwickelten Intervention oder anderer problemorientierter Lernformate sollten daher auch institutionelle Angebote wie Weiterbildungen und kollegiale Reflexionsformate zur Verfügung gestellt und wahrgenommen werden (Eder & Scholkmann, 2011).

Eine praktische Konsequenz der Experimentalstudie in Kapitel 10 ist weiterhin, in textbasierter Fallarbeit den förderlichen Effekt von affektiver State-Empathie und kognitiver Trait-Empathie auf eine urteilsbezogen effektive Informationsverarbeitung zu nutzen. Die Instruktion zur Bearbeitung der Textfälle könnte demnach beinhalten, sich in die Personen im Fall hineinzuversetzen und sich ihre Gefühlslage vorzustellen. Insgesamt zeigt die vorliegende Forschungsarbeit, dass Diagnostik und die Entwicklung diagnostischer Kompetenz durch die Integration sozial-kognitiver und kooperativer Prozesse an Qualität gewinnen.

LITERATUR

- Abs, H. J. (2007). Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern. In M. Lüders (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 63-84). Münster: Waxmann.
- Albrecht, L., Archibald, M., Arseneau, D., & Scott, S. D. (2013). Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations. *Implementation Science*, 8(1), 52. Doi: 10.1186/1748-5908-8-52
- Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In D. Y. Dai & R. J. Sternberg (Eds.), *The educational psychology series. Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 273-298). Mahwah: Lawrence Erlbaum Associates.
- Alexander, P. A., & Murphy, P. K. (1998). Profiling the differences in students' knowledge, interest, and strategic processing. *Journal of Educational Psychology*, 90(3), 435. Doi: 10.1037//0022-0663.90.3.435
- Alheit, P. (1995). „Biographizität“ als Lernpotenzial. Konzeptionelle Überlegungen zum biographischen Ansatz in der Erwachsenenbildung. In H. H. Krüger, & W. Marotzki (Hrsg.), *Erziehungswissenschaftliche Biographieforschung* (S. 276-307). Opladen: Leske + Budrich.
- Allmendinger, J. (2003). Soziale Herkunft, Schule und Kompetenzen. *Politische Studien: Zweimonatsschrift für Politik und Zeitgeschehen*, 54(03), 79-90.
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge: Addison-Wesley. Doi: 10.2307/2573151
- Altmann, A. F., & Nückles, M. (2017). Empirische Studie zu Qualitätsindikatoren für den diagnostischen Prozess. In A. Südkamp, & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 142-150). Münster: Waxmann.
- Amhag, L., & Jakobsson, A. (2009). Collaborative learning as a collective competence when students use the potential of meaning in asynchronous dialogues. *Computers & Education*, 52(3), 656-667. Doi: 10.1016/j.compedu.2008.11.012
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 261-295. Doi: 10.1016/S0022-5371(83)90201-3
- Artelt, C. (2015). Teacher judgments and their role in the educational process. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1-16. Doi: 10.1002/9781118900772.etrds0402
- Artelt, C., & Karing, C. (2014). Urteilsgenauigkeit von Lehrer(inne)n im emotional-motivationalen Bereich und im Leistungsbereich. In M. Mudiappa, & C. Artelt (Hrsg.), *BiKS – Ergebnisse aus den Längsschnittstudien* (S. 111-118). Bamberg: University of Bamberg Press.
- Arthaud, T. J., Aram, R. J., Breck, S. E., Doelling, J. E., & Bushrow, K. M. (2007). Developing collaboration skills in pre-service teachers: A partnership between general and special education. *Teacher Education and Special Education*, 30(1), 1-12. Doi: 10.1177/088840640703000101
- Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education*, 27(1), 10-20. Doi: 10.1016/j.tate.2010.08.007

- Avramidis, E., & Kalyva, E. (2007). The influence of teaching experience and professional development on Greek teachers' attitudes towards inclusion. *European Journal of Special Needs Education*, 22(4), 367-389. Doi: 10.1080/08856250701649989
- Avramidis, E., Bayliss, P., & Burden, R. (2002). Inclusion in action: an in-depth case study of an effective inclusive secondary school in the south-west of England. *International Journal of Inclusive Education*, 6(2), 143-163. Doi: 10.1080/13603110010017169
- Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration/inclusion: a review of the literature. *European Journal of Special Needs Education*, 17(2), 129-147. Doi: 10.1080/08856250210129056
- Backes-Haase, A., & Klinkisch, E. M. (2015). Das kompetente Subjekt? – Implizite Tendenzen zur Verdinglichung in der beruflichen Bildung. *bwp@ Berufs- und Wirtschaftspädagogik – online*, 29, 1-24.
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556-559. Doi: 10.1126/science.1736359
- Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, 40(1), 8-34. Doi: 10.1007/s11747-011-0278-x
- Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, 8, 456. Doi: 10.3389/fpsyg.2017.00456
- Bakdash, J. Z., & Marusich, L. R. (2019). Corrigendum: Repeated measures correlation. *Frontiers in Psychology*, 10, 1201. Doi: 10.3389/fpsyg.2019.01201
- Baker, J., & Young, B. (2014). 20 years later: deliberate practice and the development of expertise in sport. *International Review of Sport and Exercise Psychology*, 7(1), 135-157. Doi: 10.1080/1750984X.2014.896024
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, 533(7604), 452. Doi: 10.1038/533452a
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407. Doi: 10.1177/0022487108324554
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 248-287. Doi: 10.1016/0749-5978(91)90022-I
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815-824. Doi: 10.1016/j.paid.2006.09.018
- Barrows, H. S. (1985). *How to design a problem-based curriculum for preclinical years*. New York: Springer.
- Barrows, H. S. (1986). A taxonomy of problem-based learning methods. *Medical Education*, 20(6), 481-486. Doi: 10.1111/j.1365-2923.1986.tb01386.x
- Barrows, H. S. (2000). *Problem-based learning applied to medical education*. Springfield, IL: Southern Illinois University School of Medicine.
- Batsiou, S., Bebetos, E., Panteli, P., & Antoniou, P. (2008). Attitudes and intention of Greek and Cypriot primary education teachers towards teaching pupils with special educational needs in mainstream schools. *International Journal of Inclusive Education* 12(2), 201-219. Doi: 10.1080/13603110600855739

- Bauer, J., & Prenzel, M. (2012). European teacher training reforms. *Science*, 336(6089), 1642-1643. Doi: 10.1126/science.1218387
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469-520. Doi: 10.1007/s11618-006-0165-2
- Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenzen von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 29-53). Münster: Waxmann. Doi: 10.1007/s35834-011-0017-x
- Baumert, J., Köller, O. & Schnabel, K. (2000). Schulformen als differentielle Entwicklungsmilieus – eine ungehörige Fragestellung? In Gewerkschaft Erziehung und Wissenschaft GEW (Hrsg.), *Messung sozialer Motivation. Eine Kontroverse*. Schriftenreihe des Bildungs- und Förderungswerks der GEW, Band 14 (S. 28-68). Frankfurt am Main: Bildungs- und Förderungswerk der GEW.
- Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research*, 19(4), 425-430. Doi: 10.2307/3151716
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41(2), 257-278. Doi: 10.1111/j.2044-8317.1988.tb00901.x
- Behrmann, L., & Glogger-Frey, I. (2017). Produkt- und Prozessindikatoren diagnostischer Kompetenz. In A. Südkamp, & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 134-142). Münster: Waxmann.
- Behrmann, L., & van Ophuysen, S. (2017). Das Vier-Komponenten-Modell der Diagnosequalität. In A. Sdkamp, & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (S. 38-42). Münster: Waxmann.
- Behrmann, L., & Souvignier, E. (2013). The relation between teachers' diagnostic sensitivity, their classroom behavior, and their students' achievement gains in reading. *Zeitschrift für Pädagogische Psychologie*, 27, 283-293. Doi: 10.1024/1010-0652/a000112
- Bergstrom, C. M., Pugh, K. J., Phillips, M. M., & Machlev, M. (2016). Effects of Problem-Based Learning on Recognition Learning and Transfer Accounting for GPA and Goal Orientation. *The Journal of Experimental Education*, 84(4), 764-786. Doi: 10.1080/00220973.2015.1083521
- Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, 17(22), 2537-2550. Doi: .1002/(SICI)1097-0258(19981130)17:22<2537::AID-SIM953>3.0.CO;2-C
- Berliner, D. C. (1985). Laboratory settings and the study of teacher education. *Journal of Teacher Education*, 36(6), 2-8. Doi: 10.1177/002248718503600601
- Berliner, D. C. (1991). Educational psychology and pedagogical expertise: New findings and new opportunities for thinking about training. *Educational Psychologist*, 26(2), 145-155. Doi: 10.1207/s15326985ep2602_6
- Berliner, D. C. (1992). The nature of expertise in teaching. In F.K. Oser, A. Dick, & J.-L. Patry (Eds.), *Effective and Responsible Teaching* (pp. 227-248). San Francisco, CA: Jossey-Bass.

- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463-482. Doi: 10.1016/S0883-0355(02)00004-6
- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, 24(3), 200-212.
- Bilker, W. B., Hansen, J. A., Brensing, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, 19(3), 354-369. doi:10.1177/1073191112446655
- Bilz, L., Steger, J., & Fischer, S. M. (2016). Die Genauigkeit des Lehrerurteils bei der Identifikation von an Mobbing beteiligten Schülerinnen und Schülern. *Psychologie in Erziehung und Unterricht*, 63(2), 1-15. Doi: 10.2378/peu2016.art08d
- Bland, J. M., & Altman D. G. (1995). Calculating correlation coefficients with repeated observations: part 1 Correlation within subjects. *BMJ*, 310, 446. Doi: 10.1136/bmj.310.6977.446
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. (2015). Beyond dichotomies: Viewing competence as a continuum. *Zeitschrift für Psychologie*, 223(1), 3-13. Doi: 10.1027/2151-2604/a000194
- Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F. J. Hsieh, G. Kaiser, & W. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn. Teachers education and development study in mathematics (TEDS-M)* (pp. 19-47). Dordrecht: Springer. Doi: 10.1007/978-94-007-6437-8_2
- Blömeke, S., & Kaiser, G. (2017). Understanding the development of teachers' professional competencies as personally, situationally and socially determined. In D. J. Clandinin, & J. Husudoi (Eds.), *The SAGE Handbook of Research on Teacher Education* (pp. 783-802). Doi: 10.4135/9781526402042.n45
- Blömeke, S., Suhl, U., & Kaiser, G. (2011). Teacher education effectiveness: Quality and equity of future primary teachers' mathematics and mathematics pedagogical content knowledge. *Journal of Teacher Education*, 62(2), 154-171. Doi: 10.1177/0022487110386798
- Blömeke, S., & Zlatkin-Troitschanskaia, O. (2015). Kompetenzen von Studierenden. Einleitung. In S. Blömeke, & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden. 61. Beiheft Zeitschrift für Pädagogik* (S. 7-10). Weinheim: Juventa Beltz.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill. Doi: 10.3102/00028312014001073
- Blossfeld, P. N., Blossfeld, G. J., & Blossfeld, H.-P. (2015). Educational expansion and inequalities in educational opportunity: Long term changes for East and West Germany. *European Sociological Review*, 31, 144-160. Doi: 10.1093/esr/jcv017
- Boehnke, K., Sillbereisen, R. K., Reynolds, C. R., & Richmond, B. O. (1986). What I think and feel – German experience with the revised form of the children's manifest anxiety scale. *Personality and Individual Differences*, 7(4), 553-560. Doi: 10.1016/0191-8869(86)90132-7
- Bohl, T., Schneider, J., Kleinknecht, M., Kuntze, S., Rehm, M., & Syring, M. (2016). Unterricht analysieren und reflektieren mit unterschiedlichen Fallmedien: Ist Video wirklich besser als Text? *Unterrichtswissenschaft: Zeitschrift für Lernforschung*, 44(4), 474-490. Doi: 10.3262/UW1604474

- Bohner, G., & Wänke, M. (2002). *Attitudes and Attitude Change: Social Psychology*. Hove, UK: Psychology Press. Doi: 10.4135/9781446247198.n12
- Böhnisch, L., Lenz, K., & Schröer, W. (2009). *Sozialisation und Bewältigung: eine Einführung in die Sozialisationstheorie der zweiten Moderne*. Weinheim: Beltz Juventa.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York: Psychology Press.
- Borhan, M. T. (2014). Problem based learning (PBL) in teacher education: A review of the effect of PBL on pre-service teachers' knowledge and skills. *European Journal of Educational Sciences*, 1(1), 76-87. Doi: 10.19044/ejes.v1no1a9
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473-498. Doi: 10.2307/1162861
- Borko, H., & Putnam, R. T. (1996). Learning to teach. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 673-708). New York: Macmillan Library Reference; London: Prentice Hall International.
- Bos, W., Wendt, H., Köller, O., & Selter, C. (2012). TIMSS 2011: *Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern im internationalen Vergleich*. Münster: Waxmann.
- Bradley, J. H., Paul, R., & Seeman, E. (2006). Analyzing the structure of expert knowledge. *Information & Management*, 43(1), 77-91. Doi: 10.1016/j.im.2004.11.009
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217-224.
- Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16(3), 201-213. Doi: 10.1016/j.jesp.2013.10.005
- Brinkmann, J., & Twiford, T. (2012). Voices from the field: Skill sets needed for effective collaboration and co-teaching. *International Journal of Educational Leadership Preparation*, 7(3), 1-13. Doi: 10.1080/13603124.2011.638401
- Brochard, R., Dufour, A., & Despres, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, 54(2), 103-109. Doi: 10.1016/S0278-2626(03)00264-1
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Psychologie des Unterrichts und der Schule* (S. 177-212). Göttingen: Hogrefe.
- Bromme, R. (2001). Teacher Expertise. In: N. Smelser, & P. Baltes (Eds.), *International encyclopedia of the behavioral sciences. Volume: Education* (pp. 15459-15465). London: Pergamon. Doi: 10.1016/B0-08-043076-7/02447-5
- Bromme, R. (2008). Lehrerexpertise. In W. Schneider, & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 159-167). Göttingen: Hogrefe.

- Bromme, R., Prenzel, M., & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 17(4), 3-54. Doi: 10.1007/s11618-014-0514-5
- Bromme, R., Prenzel, M., & Jäger, M. (2016). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik: Zum Zusammenhang von Wissenschaftskommunikation und Evidenzbasierung in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 19(1), 129-146. Doi: 10.1007/s11618-016-0703-5
- Brownell, M. T., Adams, A., Sindelar, P., Waldron, N., & Vanhover, S. (2006). Learning from collaboration: The role of teacher qualities. *Exceptional Children*, 72(2), 169-185. Doi: 10.1177/001440290607200203
- Brownell, M. T., Ross, D. D., Colón, E. P., & McCallum, C. L. (2005). Critical features of special education teacher preparation: A comparison with general teacher education. *The Journal of Special Education*, 38(4), 242-252. Doi: 10.1177/00224669050380040601
- Brunswik, E. (1940). Thing constancy as measured by correlation coefficients. *Psychological Review*, 47, 69-78. Doi: 10.1037/h0057921
- Brush Jr., J. E., Sherbino, J., & Norman, G. R. (2017). How expert clinicians intuitively recognize a medical diagnosis. *The American Journal of Medicine*, 130(6), 629-634. Doi: 10.1016/j.amjmed.2017.01.045
- Buholzer, A., & Zulliger, S. (2013). Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 186-197. Doi: 10.3726/978-3-653-01714-4/9
- Bundesärztekammer (2005). *Stellungnahme zur „Aufmerksamkeitsdefizit- / Hyperaktivitätsstörung (ADHS)“*. Langfassung. Berlin. Zugriff am 10.10.2019 unter https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/ADHSLang.pdf.
- Burkhardt, J. M., Détienne, F., & Wiedenbeck, S. (1997). Mental representations constructed by experts and novices in object-oriented program comprehension. In S. Howard, J. Hammond, & G. Lindgaard (Eds.), *Human-Computer Interaction INTERACT'97* (pp. 339-346). Boston: Springer.
- Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment*, 23(4), 381-394. Doi: 10.1177/073428290502300406
- Byrne, M. (2017). Why change: How additional teaching resources are allocated to schools? *Reach Journal of Special Needs Education in Ireland*, 30(2), 76-84.
- Cairns, E., McWhirter, L., Duffy, U., & Barry, R. (1990). The stability of self-concept in late adolescence: Gender and situational effects. *Personality and Individual Differences*, 11(9), 937-944. Doi: 10.1016/0191-8869(90)90275-V
- Calvo, M. G., Castillo, M. D., & Schmalhofer, F. (2006). Strategic influence on the time course of predictive inferences in reading. *Memory & Cognition*, 34(1), 68-77. Doi: 10.3758/bf03193387
- Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, 3(1), 14-25. Doi: 10.1002/bs.3830030103

- Carbonell, K. B., Stalmeijer, R. E., Könings, K. D., Segers, M., & van Merriënboer, J. J. (2014). How experts deal with novel situations: A review of adaptive expertise. *Educational Research Review*, 12, 14-29. Doi: 10.1016/j.edurev.2014.03.001
- Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, 21(10), 1363-1368. Doi: 10.1177/0956797610383437
- Casale, G., Hennemann, T., & Grosche, M. (2015). Zum Beitrag der Verlaufsdiagnostik für eine evidenzbasierte sonderpädagogische Praxis am Beispiel des Förderschwerpunktes der emotionalen und sozialen Entwicklung. *Zeitschrift für Heilpädagogik*, 7, 325-334.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. New York: Elsevier.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(2), 752-766. Doi: 10.1037/0022-3514.39.5.752
- Chaiken, S., & Trope, Y. (Eds.) (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55-81. Doi: 10.1016/0010-0285(73)90004-2
- Chen, F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95, 1005-1018. Doi: 10.1037/e514412014-064
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods and Research*, 36(4), 462-494. Doi: 10.1177/0049124108314720
- Chen, Z., & Siegler, R. S. (2013). Young children's analogical problem solving: Gaining insights from video displays. *Journal of Experimental Child Psychology*, 116(4), 904-913. Doi: 10.1016/j.jecp.2013.08.009
- Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2019). Facilitating diagnostic competences in higher education – a meta-analysis in medical and teacher education. *Educational Psychology Review*, 1-40. Doi: 10.1007/s10648-019-09492-2.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple correlation/regression analysis for the social sciences*. Mahwah: Lawrence Erlbaum.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141. Doi: 10.1037//0022-0663.78.2.141
- Commission of the European Communities. (2000). *A memorandum on lifelong learning*. Retrieved September 6, 2019, from <http://hdl.voced.edu.au/10707/151503>
- Cook, D. A., Sherbino, J., & Durning, S. J. (2018). Management reasoning: beyond the diagnosis. *Jama*, 319(22), 2267-2268. Doi: 10.1001/jama.2018.4385

- Crayen, C., Eid, M., Lischetzke, T., & Vermunt, J. K. (2017). A continuous-time mixture latent-state-trait Markov model for experience sampling data. *European Journal of Psychological Assessment*, 33, 296-311. Doi: 10.1027/1015-5759/a000418
- Croskerry, P. (2017). A model for clinical decision-making in medicine. *Medical Science Educator*, 27(1), 9-13. Doi: 10.1007/s40670-017-0499-9
- Darke, P. R., Chaiken, S., Bohner, G., Einwiller, S., Erb, H. P., & Hazlewood, J. D. (1998). Accuracy motivation, consensus information, and the law of large numbers: Effects on attitude judgment in the absence of argumentation. *Personality and Social Psychology Bulletin*, 24(11), 1205-1215. Doi: 10.1177/01461672982411007
- Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education*, 40(3), 291-309. Doi: 10.1080/02619768.2017.1315399
- Davison, M. L., & Sharma, A. R. (1994). ANOVA and ANCOVA of pre-and post-test, ordinal data. *Psychometrika*, 59(4), 593-600. Doi: 10.1007/bf02294394
- De Boer, A., Pijl, S. J., & Minnaert, A. (2011). Regular primary schoolteachers' attitudes towards inclusive education: A review of the literature. *International Journal of Inclusive Education*, 15(3), 331-353. Doi: 10.1080/13603110903030089
- Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The Scientific World Journal*, 6, 1146-1163. Doi: 10.1100/tsw.2006.221
- Demirel, M., & Dağyar, M. (2016). Effects of Problem-Based Learning on Attitude: A Metaanalysis Study. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(8), 2115-2137. Doi: 10.12973/eurasia.2016.1293a
- Deutscher Bundestag (BRK). (1973). *Bildungsgesamtplan*. Drucksache 7/1474. Bonn: Hans Heger.
- Dickhäuser, O., Schöne, C., Spinath, B., & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept: Konstruktion und Überprüfung eines neuen Instrumentes. *Zeitschrift für differentielle und diagnostische Psychologie*, 23(4), 393-405. Doi: 10.1024//0170-1789.23.4.393
- Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53(3), 497. Doi: 10.1037//0022-3514.53.3.497
- Dietrich, J. (2019). Methoden der Veränderungsmessung. In B. Kracke, & P. Noack (Hrsg.), *Handbuch Entwicklungs- und Erziehungspsychologie* (S. 1-23). Berlin: Springer.
- Ditton, H. (2013). Bildungsverläufe in der Sekundarstufe. Ergebnisse einer Längsschnittstudie zu Wechseln der Schulform und des Bildungsgangs. *Zeitschrift für Pädagogik*, 59(6), 887-911. Doi: 10.1007/s11618-014-0536-z
- Dixon, P., & Bortolussi, M. (2018). Readers' processing of perceptual perspective and stance. *Discourse Processes*, 56(7), 513-529. Doi:10.1080/0163853x.2018.1512829
- Doane, S. M., Sohn, Y. W., & Jodlowski, M. T. (2004). Pilot ability to anticipate the consequences of flight actions as a function of expertise. *Human Factors*, 46(1), 92-103. Doi: 10.1518/hfes.46.1.92.30386

- Dochy, F., Segers, M., van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, 13(5), 533-568. Doi: 10.1016/s0959-4752(02)00025-7
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer-Lehrbuch. Berlin: Springer. doi:10.1007/978-3-642-41089-5
- Dreher, A., Lindmeier, A., Heinze, A., & Niemand, C. (2018). What kind of content knowledge do secondary mathematics teachers need? *Journal für Mathematik-Didaktik*, 39(2), 319-341. Doi: 10.1007/s13138-018-0127-2
- Dreyfus, H. L., & Dreyfus, S. E. (2004). The ethical implications of the five-stage skill-acquisition model. *Bulletin of Science, Technology & Society*, 24(3), 251-264. Doi: 10.1177/0270467604265023
- Dunlap, W. P., Jones, M. B., & Bittner, A. C. (1983). Average correlations vs. correlated averages. *Bulletin of the Psychonomic Society*, 21(3), 213-216. Doi: 10.3758/bf03334690
- 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. Doi: 10.1024/1010-0652.23.34.187
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, 621. Doi: 10.3389/fpsyg.2015.00621
- Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of 'diagnostic competence'. *Language Testing*, 21(3), 259-283. Doi: 10.1191/0265532204lt284oa
- Edelsbrunner, P. A., & Dablander, F. (2019). The psychometric modeling of scientific reasoning: A review and recommendations for future avenues. *Educational Psychology Review*, 31(1), 1-34. Doi: 10.1007/s10648-018-9455-5
- Eder, F., Dämon, K., & Hörl, G. (2011). Das „Autonomie-Paritäts-Muster“: Vorberuflich erlerntes Stereotyp, Bewältigungsstrategie oder Ergebnis der beruflichen Sozialisation? *Zeitschrift für Bildungsforschung*, 1(3), 199-217. Doi: 10.1007/s35834-011-0021-1
- Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden* (2. Aufl.). Weinheim: Beltz.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169-189. Doi: 10.1207/s15326985ep3403_3
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.
- Elstein, A. S. (2009). Thinking about diagnostic thinking: A 30-year perspective. *Advances in health sciences education*, 14(1), 7-18. Doi: 10.1007/s10459-009-9184-0
- Elstein, A. S. (2016). Revisiting 'Measuring the process of solving clinical diagnostic problems'. *Medical Education*, 50(2), 155-159. Doi: 10.1111/medu.12804
- Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *BMJ*, 324(7339), 729-732. Doi: 10.1136/bmj.324.7339.729
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge: Harvard University Press.
- Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56(3), 495-515. Doi: 10.1007/BF02294487

- Erdogan, T., & Senemoglu, N. (2017). PBL in teacher education: Its effects on achievement and self-regulation. *Higher Education Research & Development*, 36(6), 1152-1165. Doi: 10.1080/07294360.2017.1303458
- Ericsson, K. A., Delaney, P. F., Weaver, G., & Mahadevan, R. (2004). Uncovering the structure of a memorist's superior "basic" memory capacity. *Cognitive Psychology*, 49(3), 191-237. Doi: 10.1016/j.cogpsych.2004.02.001
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363. Doi: 10.1037//0033-295X.100.3.363
- Erikson, R. (2016). Is it enough to be bright? Parental background, cognitive ability and educational attainment. *European Societies*, 18(2), 117-135. Doi: 10.1080/14616696.2016.1141306
- Erpenbeck, J. (2002). Kompetenz und Performanz im Bild moderner Selbstorganisationstheorie. Bundesinstitut für Berufsbildung (Hrsg.), *Berufsbildung für eine globale Gesellschaft. Perspektiven im 21. Jahrhundert* (S. 1-12). Bielefeld: Bertelsmann.
- Erpenbeck, J. (2014). Stichwort Kompetenzen. *DIE Zeitschrift für Erwachsenenbildung*, 3, 20-21.
- Erpenbeck, J., & Heyse, V. (1997). Berufliche Weiterbildung und berufliche Kompetenzentwicklung. In Arbeitsgemeinschaft Qualifikation-Entwicklung-Management (Hrsg.), *Kompetenzentwicklung 1996* (S. 15-152), Bielefeld: Bertelsmann.
- Esser, G., & Schmidt, M. H. (1993). Die langfristige Entwicklung von Kindern mit Lese-Rechtschreibschwäche. *Zeitschrift für Klinische Psychologie*, 22, 100-116.
- Esser, G., Wyschkon, A., & Schmidt, M. H. (2002). Was wird aus Achtjährigen mit einer Lese- und Rechtschreibstörung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 31(4), 235-242. Doi: 10.1026/0084-5345.31.4.235
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278. Doi: 10.1146/annurev.psych.59.103006.093629
- Evans, J. S. B. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, 31(2-3), 86-102. Doi: 10.1016/j.dr.2011.07.007
- Everett, J. A. C., & Earp, B. D. (2015). A tragedy of the (academic) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers. *Frontiers in Psychology*, 6, 1152. Doi: 10.3389/fpsyg.2015.01152
- Ewing, D. L., Monsen, J. J., & Kielblock, S. (2018). Teachers' attitudes towards inclusive education: A critical review of published questionnaires. *Educational Psychology in Practice*, 34(2), 150-165. Doi: 10.1080/02667363.2017.1417822
- Falkai, P., & Wittchen H.-U. (Hrsg.) (2015). *Diagnostisches und statistisches Manual psychischer Störungen DSM-5*. Göttingen: Hogrefe.
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56-83. Doi: 10.1080/10705519909540119
- Fehm, L., & Fydrich, T. (2011). Prüfungsangst. Fortschritte der Psychotherapie (Bd. 44). Göttingen: Hogrefe.

- Fernet, C., Senécal, C., Guay, F., Marsh, H., & Dowson, M. (2008). The work tasks motivation scale for teachers (WTMST). *Journal of Career Assessment*, 16, 256-279. Doi: 10.1177/1069072707305764
- Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, 98, 19-38. Doi: 10.1016/j.brat.2017.05.013
- Fischbach, A., Schuchardt, K., Brandenburg, J., Kleszczewski, J., Balke-Melcher, C., Schmidt, C., ... Hasselhorn, M. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien. *Lernen und Lernstörungen*, 2, 65-76. Doi: 10.1024/2235-0977/a000035
- Fischer, C., Kopmann, H., Rott, D., Veber, M., & Zeinz, H. (2014). Adaptive Lehrkompetenz und pädagogische Haltung. Lehrerbildung für eine inklusive Schule. *Jahrbuch für allgemeine Didaktik*, (4), 16-33.
- Fischer, F., Kollar, I., Stegmann, K., Wecker, C., Zottmann, J., & Weinberger, A. (2013). Collaboration scripts in computer-supported collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. Chan, & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (S. 403-419). New York: Routledge.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 5, 28-45. Doi: 10.1007/s11251-016-9380-2
- Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, 48(1), 3-26. Doi: 10.1007/BF02314674
- Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., ... Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, 137(4), 517. Doi: 10.1037/a0023304
- Fleming, M. A., & Petty, R. E. (2000). Identity and persuasion: An elaboration likelihood approach. In D. J. Teddy, & M. A. Hogg (Eds.), *Applied social research. Attitudes, behavior, and social context: The role of norms and group membership* (pp. 171-199). Mahwah, NJ: Lawrence Erlbaum Associates Publishers. Doi: 10.4324/9781410603210-10
- Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, 7(6), 585-594. Doi: 10.1177/1745691612459520
- Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., & Bolhuis, J. J. (2017). Language, mind and brain. *Nature Human Behaviour*, 1(10), 713-722. Doi: 10.1038/s41562-017-0184-4
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18(3), 233-239. Doi: 10.1111/j.1467-9280.2007.01882.x
- Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice*, 18(3), 172-186. Doi: 10.1111/1540-5826.00073
- Füchter, A. (2011). Pädagogische und didaktische Diagnostik: Eine schulische Entwicklungsaufgabe mit hohem Professionalitätsanspruch. In A. Füchter, & K. Moegling (Hrsg.), *Diagnostik und Förderung Teil I: Didaktische Grundlagen* (S. 45-83). Immenhausen: Prolog.
- Gagné, R. M. (1965). *Conditions of learning*. New York: Holt, Reinehart & Winston.

- Gagne, R. M. (1980). Learnable aspects of problem solving. *Educational Psychologist*, 15(2), 84-92. Doi: 10.1080/00461528009529218
- Gao, Z., Ye, T., Shen, M., & Perry, A. (2016). Working memory capacity of biological movements predicts empathy traits. *Psychonomic Bulletin & Review*, 23(2), 468-475. Doi: 10.3758/s13423-015-0896-2
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Geiser, C. (2011). *Datenanalyse mit Mplus*. Wiesbaden: VS Verlag.
- Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, 48(3), 432-435. Doi: 10.1198/004017005000000661
- Gelman, R., & Greeno, J. G. (1989). On the nature of competence: Principles for understanding in a domain. In L. B. Resnick (Eds.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 125-186). Hillsdale: Lawrence Erlbaum Associates.
- Genette, G. (1988). *Narrative discourse revisited*. Ithaca, NY: Cornell University Press.
- Ghanizadeh, A., Bahredar, M. J., & Moeini, S. R. (2006). Knowledge and attitudes towards attention deficit hyperactivity disorder among elementary school teachers. *Patient Education and Counseling*, 63(1-2), 84-88. Doi: 10.1016/j.pec.2005.09.002
- Ghousseini, H., & Herbst, P. (2016). Pedagogies of practice and opportunities to learn about classroom mathematics discussions. *Journal of Mathematics Teacher Education*, 19(1), 79-103. Doi: 10.1007/s10857-014-9296-1
- Gigerenzer, G. (2015). *Oxford series in evolution and cognition. Simply rational: Decision making in the real world*. New York: Oxford University Press.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4), 650. Doi: 10.1037//0033-295x.103.4.650
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98(4), 506. Doi: 10.1037//0033-295X.98.4.506
- Gijbels, D., Dochy, F., van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, 75(1), 27-61. Doi: 10.3102/00346543075001027
- Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013). Beyond judgment bias: How students' ethnicity and academic profile consistency influence teachers' tracking judgments. *Social Psychology of Education*, 16(4), 555-573. Doi: 10.1007/s11218-013-9227-5
- Goldschmidt, G. (1997). Capturing indeterminism: Representation in the design problem space. *Design Studies*, 18(4), 441-455. Doi: 10.1016/s0142-694x(97)00011-2
- Gonczi, A. (1994). Competency based assessment in the professions in Australia. *Assessment in Education: Principles, Policy & Practice*, 1(1), 27-44. Doi: 10.1080/0969594940010103
- Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, 10(3), 281-287. Doi: 10.1111/j.1467-7687.2007.00584.x
- Graber, M. L., Kissam, S., Payne, V. L., Meyer, A. N., Sorensen, A., Lenfestey, N., ... Singh, H. (2012). Cognitive interventions to reduce diagnostic error: A narrative review. *BMJ Quality & Safety*, 21(7), 535-557. Doi: 10.1136/bmjqs-2011-000149

- Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage*, 38(2), 346-356. Doi: 10.1016/j.neuroimage.2007.07.041
- Grant, S., Mayo-Wilson, E., Montgomery, P., Macdonald, G., Michie, S., Hopewell, S., & Moher, D. (2018). CONSORT-SPI 2018 Explanation and Elaboration: Guidance for reporting social and psychological intervention trials. *Trials*, 19(1), 406. Doi: 10.1186/s13063-018-2735-z
- Gräsel, C., Fussangel, K., & Parchmann, I. (2006). Lerngemeinschaften in der Lehrerfortbildung. *Zeitschrift für Erziehungswissenschaft*, 9(4), 545-561. Doi: 10.1007/s11618-006-0167-0
- Gräsel, C., Fußangel, K., & Pröbstel, C. (2006). Lehrkräfte zur Kooperation anregen – eine Aufgabe für Sisyphos? *Zeitschrift für Pädagogik*, 52(2), 205-219. Doi: 10.1515/zpt-1986-0604
- Gräsel, C., Krolak-Schwerdt, S., Nölle, I., & Hörstermann, T. (2010). Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der Übergangsempfehlung. Eine Analyse aus der Perspektive der sozialen Urteilsbildung. Projekt Diagnostische Kompetenz. In E. Klieme, D. Leutner, & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 286-295). Weinheim: Beltz.
- Green, A. J. K., & Gilhooly, K. J. (1992). Empirical advances in expertise research. In M. T. Keane, & K. J. Gilhooly (Eds.), *Advances in the Psychology of Thinking, Volume 1* (pp. 45-70), Hemel Hempstead: Harvester Wheatsheaf.
- Greenwood, J., & King, M. (1995). Some surprising similarities in the clinical reasoning of 'expert' and 'novice' orthopaedic nurses: Report of a study using verbal protocols and protocol analyses. *Journal of Advanced Nursing*, 22(5), 907-913. Doi: 10.1111/j.1365-2648.1995.tb02642.x
- Greiten, S., Geber, G., Gruhn, A., & Königer, M. (2018). Inklusion als Aufgabe für die Lehrerbildung. Theoretische, institutionelle, curriculare und didaktische Herausforderungen für Hochschulen. In S. Greiten, G. Geber, A. Gruhn, & M. Königer (Hrsg.), *Lehrerbildung für Inklusion: Fragen und Konzepte zur Hochschulentwicklung* (S. 14-36). Münster: Waxmann.
- Grohs, J. R., Kirk, G. R., Soledad, M. M., & Knight, D. B. (2018). Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. *Thinking Skills and Creativity*, 28, 110-130. Doi: 10.1016/j.tsc.2018.03.003
- Gronau, Q. F., van Erp, S., Heck, D. W., Cesario, J., Jonas, K. J., & Wagenmakers, E. J. (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: The case of felt power. *Comprehensive Results in Social Psychology*, 2(1), 123-138. Doi: 10.1080/23743603.2017.1326760
- Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology*, 5, 980. Doi: 10.3389/fpsyg.2014.00980
- Guerra, M. G., Kupczynski, L. P., Mundy, M. A., & Garza, C. (2017). The diagnostic competencies: Perceptions of educational diagnosticians, administrators, and classroom teachers. *International Journal of Special Education*, 32(4), 708-722.

- Ha, D., & Park, W. (2014). The relationship among teachers' empathy, communication style and coping type for students with attention deficit hyperactivity disorder. *Journal of Korean Academy of Psychiatric and Mental Health Nursing*, 23(2), 103-112. Doi: 10.12934/jkpmhn.2014.23.2.103
- Habermas, J. (1971). *Erkenntnis und Interesse* (2. Aufl.). Frankfurt am Main: Suhrkamp.
- Habermas, J. (1984). *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*. Frankfurt am Main: Suhrkamp.
- Hager, P., & Gonczi, A. (1996). What is competence? *Medical Teacher*, 18(1), 15-18. Doi: 10.3109/01421599609040255
- Hagiwara, M., Shogren, K. A., Thompson, J. R., Burke, K. M., Uyanik, H., Amor, A. M., ... Aguayo, V. (2019). International trends in inclusive education intervention research: A literature review. *Education and Training in Autism and Developmental Disabilities*, 54(1), 3-17. Doi: 10.1080/13603116.2018.1445304
- Hambrick, D., Burgoyne, A., Macnamara, B., & Ullen, F. (2018). Toward a multifactorial model of expertise: Beyond born versus made. *Annals of the New York Academy of Sciences*, 1423(1), 284-295. Doi: 10.1111/nyas.13586
- Hamilton, J., Gagne, P. E., & Hancock, G. R. (2003). *The effect of sample size on latent growth models*. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago.
- Hankin, B. L., & Abela, J. R. (Eds.) (2005). *Development of psychopathology: A vulnerability-stress perspective*. London: Sage.
- Hargreaves, A. (2000). Four ages of professionalism and professional learning. *Teachers and Teaching*, 6(2), 151-182. Doi: 10.1080/713698714
- Hartig, J. (2006). Kompetenzen als Ergebnis von Bildungsprozessen. *DIPF Informiert*, 10, 2-7.
- Hartig, J., & Frey, A. (2013). Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl“ für die Modellierung von Kompetenzen? *Zeitschrift für Erziehungswissenschaft*, 16(1), 47-51. Doi: 10.1007/s11618-013-0386-0
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127-143). Berlin: Springer.
- Harvey, M. W., Yssel, N., Bauserman, A. D., & Merbler, J. B. (2010). Preservice teacher preparation for inclusion: An exploration of higher education teacher-training institutions. *Remedial and Special Education*, 31(1), 24-33. Doi: 10.1177/0741932508324397
- Haynes, J. D., & Rees, G. (2006). Neuroimaging: Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7), 523. Doi: 10.1038/nrn1931
- Heckhausen, H., & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and Emotion*, 11(2), 101-120. Doi: 10.1007/bf00992338
- Heinisch, J., Sonnleitner, M., & Rank, A. (2018). Professionalisierung für Inklusion – Situiertes Lernen in der universitären LehrerInnenbildung. *Jahrbuch Grundschulforschung*, 22, 296-302. Doi: 10.1007/978-3-658-13502-7_37

- Hellmann, K. A. (2015). *Diagnostische Kompetenz als Perspektivenübernahmefähigkeit. Der Einfluss von Lehrerfahrung, Vertrautheit mit dem Urteilsobjekt und Aufgabendesign auf die Urteilsgröße*. Unveröffentlichte Dissertation. Albert-Ludwigs-Universität Freiburg.
- Hellmich, F., & G. Görel (2014). Erklärungsfaktoren für Einstellungen von Lehrerinnen und Lehrern zum inklusiven Unterricht in der Grundschule. *Zeitschrift für Bildungsforschung* 4(3), 227-240. Doi: 10.1007/s35834-014-0102-z
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (2. Aufl.). Seelze-Velber: Klett Kallmeyer.
- Helmke, A., & Schrader, F. W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91-98. Doi: 10.1016/0742-051x(87)90010-2
- Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold, & C. Griesse (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider.
- Hemker, L., Prescher, C., & Narciss, S. (2017). Design and evaluation of a problem-based learning environment for teacher training. *Interdisciplinary Journal of Problem-Based Learning*, 11(2), 10. Doi: 10.7771/1541-5015.1676
- Henry, M., Merten, T., & Wallasch, T. M. (2008). Semantic dementia. A case report in the context of an independent medical examination. *Fortschritte der Neurologie-Psychiatrie*, 76(8), 453-464. Doi: 10.1055/s-2008-1038210
- Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: Concerns and considerations. *Computational Methods in Social Sciences*, 3(1), 5-17. Doi: 10.1037/e518392013-131
- Herppich, S., Praetorius, A. K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., ... Hetmanek, 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. Doi: 10.1016/j.tate.2017.12.001
- Hertel, G., Kerr, N. L., & Messé, L. A. (2000). Motivation gains in performance groups: Paradigmatic and theoretical developments on the Köhler effect. *Journal of Personality and Social Psychology*, 79(4), 580-601. Doi: 10.1037//0022-3514.79.4.580
- Herzberg H., & Truschkat I. (2009). Lebenslanges Lernen und Kompetenz: Chancen und Risiken der Verknüpfung zweier Diskursstränge. In P. Alheit, & H. von Felden H. (Hrsg.), *Lebenslanges Lernen und erziehungswissenschaftliche Biographieforschung* (S. 111-126). Wiesbaden: VS Verlag. Doi: 10.1007/978-3-531-91520-3_7
- Hesjedal, E., Hetland, H., & Iversen, A. C. (2015). Interprofessional collaboration: Self-reported successful collaboration by teachers and social workers in multidisciplinary teams. *Child & Family Social Work*, 20(4), 437-445. Doi: 10.1111/cfs.12093
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539-1558. Doi: 10.1002/sim.1186
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557-560. Doi: 10.1136/bmj.327.7414.557

- Hillenbrand, C., Melzer, C., & J. Sung (2014). Lehrerbildung für Inklusion in Deutschland: Stand der Diskussion und praktische Konsequenzen. *Theory and Practice of Education* 19, 2, 147-171. Doi: 10.5771/9783845239569-73
- Hirschfeld, L. A., & Gelman, S. A. (Eds.). (1994). *Mapping the mind: Domain specificity in cognition and culture*. Cambridge: Cambridge University Press.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235-266. Doi: 10.1023/b:edpr.0000034022.16470.f3
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark. *Educational Psychologist*, 42(2), 99-107. Doi: 10.1080/00461520701263368
- Hoffmann, L., & Böhme, K. (2014). Wie gut können Grundschullehrkräfte die Schwierigkeit von Deutsch- und Mathematikaufgaben beurteilen? *Psychologie in Erziehung und Unterricht*, 1, 42-55. Doi: 10.2378/peu2014.art05d
- Högborg, R. (2011). Cheating as subversive and strategic resistance: Vocational students' resistance and conformity towards academic subjects in a Swedish upper secondary school. *Ethnography and Education*, 6(3), 341-355. Doi: 10.1080/17457823.2011.610584
- Hoge, R. D. (1983). Psychometric properties of teacher-judgment measures of pupil aptitudes, classroom behaviors, and achievement levels. *Journal of Special Education*, 17(4), 401-429. Doi: 10.1177/002246698301700404
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297-313. Doi: 10.3102/00346543059003297
- Horstkemper, M. (2004). Diagnosekompetenz als Teil pädagogischer Professionalität. *Neue Sammlung*, 44, 201–214.
- Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM Mathematics Education*, 48(1-2), 41-53. Doi: 10.1007/s11858-016-0759-y
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424. Doi: 10.1037//1082-989x.3.4.424
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. Doi: 10.1080/10705519909540118
- Huang, H., Liu, Y., & Chen, Y. (2018). Preservice preschool teachers' responses to bullying scenarios: The roles of years of study and empathy. *Frontiers in Psychology*, 9, 1-13. Doi: 10.3389/fpsyg.2018.00175
- Huber, S. G., & Ahlgrimm, F. (Hrsg.). (2012). *Kooperation*. Münster: Waxmann.
- Hung, W. (2006). The 3C3R model: A conceptual framework for designing problems in PBL. *Interdisciplinary Journal of Problem-based Learning*, 1(1), 6. Doi: 10.7771/1541-5015.1006

- Hwang, S. Y., & Kim, M. J. (2006). A comparison of problem-based learning and lecture-based learning in an adult health nursing course. *Nurse Education Today*, 26(4), 315-321. Doi: 10.1016/j.nedt.2005.11.002
- Ilgen, J. S., Humbert, A. J., Kuhn, G., Hansen, M. L., Norman, G. R., Eva, K. W., ... Sherbino, J. (2012). Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. *Academic Emergency Medicine*, 19(12), 1454-1461. Doi: 10.1111/acem.12034
- Imhof, M. (2016). *Psychologie für Lehramtsstudierende*. Wiesbaden: Springer.
- Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* (6. Aufl.). Weinheim: Beltz.
- Ise, E., Dolle, K., Pixner, S., & Schulte-Körne, G. (2012). Effektive Förderung rechenschwacher Kinder. *Kindheit und Entwicklung*, 21, 181-192. Doi: 10.1026/0942-5403/a000083
- Ise, E., Engel, R. R., & Schulte-Körne, G. (2012). Was hilft bei der Lese-Rechtschreibstörung? *Kindheit und Entwicklung*, 21(2), 122-136. Doi: 10.1026/0942-5403/a000077
- Istance, D., & Paniagua, A. (2018). Teachers as Designers of Learning Environments: The Importance of Innovative Pedagogies, Educational Research and Innovation. Paris: OECD Publishing. Doi: 10.1787/9789264085374-en
- Jacobsen, A. (2001). *Die gesellschaftliche Wirklichkeit der Polizei: Eine empirische Untersuchung zur Rationalität polizeilichen Handelns*. Dissertation. Universität Bielefeld.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35(1), 21-35.
- Jäger, R. S. (2012). Pädagogische Diagnostik und Förderung: Vom Erkennen zum Handeln. *bwp@Berufs- und Wirtschaftspädagogik*, 22, 1-13.
- Jäger, R. S., Frey, A., Jäger-Flor, D., Lissmann, U. & Riebel, J. (2010), Pädagogische Diagnostik. In R. S. Jäger, P. Nenniger, H. Petillon, B. Schwarz, & B. Wolf (Hrsg.), *Empirische Pädagogik 1990–2010. Eine Bestandsaufnahme der Forschung in der Bundesrepublik Deutschland. Band 1: Grundlegende empirische pädagogische Forschung* (S. 49-88). Landau: Verlag Empirische Pädagogik.
- Jensen, B., Sonnemann, J., Roberts-Hull, K., & Hunter, A. (2016). *Beyond PD: Teacher professional learning in high-performing systems. Teacher quality systems in top performing countries*. Washington, DC: National Center on Education and the Economy (NCEE) & Center on International Education Benchmarking (CIEB).
- Jensen-Doss, A., & Weisz, J. R. (2008). Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology*, 76(5), 711-722. Doi: 10.1037/0022-006X.76.5.711
- Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using hidden Markov models to characterize student behaviors in learning-by-teaching environments. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. P. Lajjoie (Eds.), *International conference on intelligent tutoring systems* (pp. 614-625). Berlin: Springer. Doi: 10.1007/978-3-540-69132-7_64
- Johnson, B. (2003). Teacher collaboration: Good for some, not so good for others. *Educational Studies*, 29(4), 337-350. Doi: 10.1080/0305569032000159651

- Johnson, S. M., & Finucane, P. M. (2000). The emergence of problem-based learning in medical education. *Journal of Evaluation in Clinical Practice*, 6(3), 281-291. Doi: 10.1046/j.1365-2753.2000.00267.x
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.
- Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, 45(1), 65-94. Doi: 10.1007/bf02299613
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63-85. Doi: 10.1007/bf02300500
- Jonassen, D. H., & Henning, P. (1999). Mental models: Knowledge in the head and knowledge in the world. *Educational Technology*, 39(3), 37-42.
- Jonassen, D. H., & Hung, W. (2008). All problems are not equal: Implications for pbl. *Interdisciplinary Journal of Problem-Based Learning*, 2(2), 10. Doi: 10.7771/1541-5015.1080
- Jurkowski, S., & Müller, B. (2018). Co-teaching in inclusive classes: The development of multi-professional cooperation in teaching dyads. *Teaching and Teacher Education*, 75, 224-231. Doi: 10.1016/j.tate.2018.06.017
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and social Psychology*, 57(3), 469. Doi: 10.1016/j.tate.2018.06.017
- 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. Doi: 10.1024/1010-0652/a000076
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, 28, 73-84. Doi: 10.1016/j.learninstruc.2013.06.001
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul-und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(34), 197-209. Doi: 10.1024/1010-0652.23.34.197
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Waxmann.
- Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology*, 55, 623-655. Doi: 10.1146/annurev.psych.55.090902.142009
- Kim, L. E., & Klassen, R. M. (2018). Teachers' cognitive processing of complex school-based scenarios: Differences across experience levels. *Teaching and Teacher Education*, 73, 215-226. Doi: 10.1016/j.tate.2018.04.006
- Kim, M. K. (2015). Models of learning progress in solving complex problems: Expertise development in teaching and learning. *Contemporary Educational Psychology*, 42, 1-16. Doi: 10.1016/j.cedpsych.2015.03.005
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182. Doi: 10.1037/0033-295x.95.2.163
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363. Doi: 10.1037/0033-295x.85.5.363

- Klauer, K. J. (1982). Ein kriteriumsorientiertes Zensierungsmodell. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 14(1), 65-79.
- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, 56, 10-13.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., ... Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Berlin: BMBF.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (S. 3-22). Ashland: Hogrefe & Huber. Doi: 10.1007/s10389-009-0270-9
- Klieme, E., Maag-Merki, K., & Hartig, J. (2007). Kompetenzbegriff und Bedeutung von Kompetenzen im Bildungswesen. In J. Hartig (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung* (S. 1-15). Berlin: BMBF.
- Klug, J. (2011). *Modeling and Training a New Concept of Teachers' Diagnostic Competence*. Unpublished Dissertation. Technical University Darmstadt. Retrieved September 6, 2019, from <https://tuprints.ulb.tu-darmstadt.de/id/eprint/2838>
- Klug, J. (2017). Ein Prozessmodell zur Diagnostik und Förderung von selbstreguliertem Lernen. In A. Südkamp, & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 223-230). Münster: Waxmann.
- 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: Theory and Practice*, 22(4), 461-484. Doi: 10.1080/13540602.2015.1082729
- 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. Doi: 10.1016/j.tate.2012.10.004
- Klug, J., Bruder, S., Keller, S., & Schmitz, B. (2012). Hängen Diagnostische Kompetenz und Beratungskompetenz von Lehrkräften zusammen? Eine korrelative Untersuchung. *Psychologische Rundschau*, 63, 3-10. Doi: 10.1026/0033-3042/a000104
- Klug, J., Schultes, M. T., & Spiel, C. (2018). Assessment at school – Teachers' diary-supported implementation of a training program. *Teaching and Teacher Education*, 76, 298-308. Doi: 10.1016/j.tate.2017.10.014
- Koch, S., & Gräsel, C. (2004): Schulreformen und Neue Steuerung – erziehungs- und verwaltungswissenschaftliche Perspektiven. In S. Koch, & R. Fisch (Hrsg.), *Schulen für die Zukunft. Neue Steuerung im Bildungswesen* (S. 3-24). Baltmannsweiler: Schneider Hohengehren.
- Koenen, A. K., Dochy, F., & Berghmans, I. (2015). A phenomenographic analysis of the implementation of competence-based education in higher education. *Teaching and Teacher Education*, 50, 1-12. Doi: 10.1016/j.tate.2015.04.001
- Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie*, 216(2), 61-73. Doi: 10.1027/0044-3409.216.2.61
- Kokemohr, R. (2007). Bildung als Welt- und Selbstentwurf im Fremden. Annäherungen an eine Bildungsprozesstheorie. In H. C. Koller, W. Marotzki, & O. Sander (Hrsg.), *Bildungsprozesse und*

- Fremdheitserfahrungen. Beiträge zu einer Theorie transformatorischer Bildungsprozesse* (S. 13-69). Bielefeld: transcript.
- Koller, H. C. (2018) (2., aktual. Aufl.). *Bildung anders denken: Einführung in die Theorie transformatorischer Bildungsprozesse*. Stuttgart: Kohlhammer.
- Koller, H. C., Marotzki, W., & Sanders, O. (Hrsg.) (2015). *Bildungsprozesse und Fremdheitserfahrung. Beiträge zu einer Theorie transformatorischer Bildungsprozesse*. Bielefeld: transcript.
- König, J., & Klemenz, S. (2015). Der Erwerb von pädagogischem Wissen bei angehenden Lehrkräften in unterschiedlichen Ausbildungskontexten: Zur Wirksamkeit der Lehrerausbildung in Deutschland und Österreich. *Zeitschrift für Erziehungswissenschaft*, 18(2), 247-277. Doi: 10.1007/s11618-015-0623-9
- König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A. M., ... Kaiser, G. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? *Zeitschrift für Erziehungswissenschaft*, 21(3), 1-38. Doi: 10.1007/s11618-017-0765-z
- König, J., Gerhard, K., Kaspar, K., & Melzer, C. (2019). Professionelles Wissen von Lehrkräften zur Inklusion: Überlegungen zur Modellierung und Erfassung mithilfe standardisierter Testinstrumente. *Pädagogische Rundschau*, 73(1), 43-64.
- König, J., Kaiser, G., & Felbrich, A. (2012). Spiegelt sich pädagogisches Wissen in den Kompetenzselbsteinschätzungen angehender Lehrkräfte? Zum Zusammenhang von Wissen und Überzeugungen am Ende der Lehrerausbildung. *Zeitschrift für Pädagogik*, 58(4), 476-491.
- Konstantopoulus, S. (2006). Fixed and mixed effects models in meta-analysis. *IZA Discussion Paper No. 2198*. Retrieved September 6, 2019, from <http://ftp.iza.org/dp2198.pdf>
- Kontopantelis, E. (2018). A comparison of one-stage vs two-stage individual patient data meta-analysis methods: A simulation study. *Research Synthesis Methods*, 9(3), 417-430. Doi: 10.1002/jrsm.1303
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547-588. Doi: 10.3102/0034654318759268
- Kramer, C., König, J., Kaiser, G., Ligtvoet, R., & Blömeke, S. (2017). Der Einsatz von Unterrichtsvideos in der universitären Ausbildung: Zur Wirksamkeit video- und transkriptgestützter Seminare zur Klassenführung auf pädagogisches Wissen und situationsspezifische Fähigkeiten angehender Lehrkräfte. *Erziehungswissenschaft*, 20(1), 137-164. Doi: 10.1007/s11618-017-0732-8
- Krammer, K. (2014). Fallbasiertes Lernen mit Unterrichtsvideos in der Lehrerinnen- und Lehrerbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32(2), 164-175. URN: urn:nbn:de:0111-pedocs-138638
- Krampen, G., Viebig, J., & Walter, W. (1982). Entwicklung einer Skala zur Erfassung dreier Aspekte von sozialem Vertrauen. *Diagnostica*, 28(3), 242-247.
- Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, 1428, 13-23. Doi: 10.1026/0012-1924.51.1.29
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2012). Leistungsbeurteilungen von Schulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44, 111-122. Doi: 10.1026/0049-8637/a000062

- Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, 42(2), 109-113. Doi: 10.1080/00461520701263376
- Kultusministerkonferenz (KMK). (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der KMK vom 16.12.2004*. Bonn: KMK. Zugriff am 10.10.2019 unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Kultusministerkonferenz (KMK). (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Bonn: KMK. Zugriff am 10.10.2019 unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf
- Kultusministerkonferenz (KMK) (2011). *Inklusive Bildung von Kindern und Jugendlichen mit Behinderungen in Schulen*. Beschluss der Kultusministerkonferenz vom 20.10.2011. Bonn.
- Kultusministerkonferenz (KMK), & Hochschulrektorenkonferenz (HRK). (2015). *Lehrerbildung für eine Schule der Vielfalt. Gemeinsame Empfehlung von Hochschulrektorenkonferenz und Kultusministerkonferenz*. Bonn: KMK. Zugriff am 10.10.2019 unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_03_12-Schule-der-Vielfalt.pdf
- Kumschick, I. R., Piwowar, V., Ophardt, D., Barth, V., Krysmanski, K., & Thiel, F. (2017). Optimierung einer videobasierten Lerngelegenheit im Problem Based Learning Format durch Cognitive Tools. Eine Interventionsstudie mit Lehramtsstudierenden. *Zeitschrift für Erziehungswissenschaft*, 20(1), 93-113. Doi: 10.1007/978-3-658-15739-5_5
- Kunter, M., Baumert, J., & Blum, W. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*. New York: Springer.
- Kunter, M., Klusmann, U., & Baumert, J. (2009). Professionelle Kompetenz von Mathematiklehrkräften: Das COACTIV-Modell. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus, & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 153-165). Weinheim: Beltz.
- Kunter, M., Kunina-Habenicht, O., Baumert, J., Dicke, T., Holzberger, D., Lohse-Bosse, H. ... Terhart, E. (2017). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung. In C. Gräsel, & Trempler, K. (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals. Entwicklung von Professionalität pädagogischen Personals* (S. 37-54). Wiesbaden: Springer VS. Doi: 10.1007/978-3-658-07274-2_3
- Kurniawati, F., de Boer, A. A., Minnaert, A., & Mangunsong, F. (2017). Evaluating the effect of a teacher training programme on the primary teachers' attitudes, knowledge and teaching strategies regarding special educational needs. *Educational Psychology*, 37(3), 287-297. Doi: 10.1080/01443410.2016.1176125

- Kuvac, M., & Koc, I. (2018). The effect of problem-based learning on the metacognitive awareness of pre-service science teachers. *Educational Studies* 21(1), 1-21. Doi: 10.1080/03055698.2018.1509783
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science*, 44(3), 197-203. Doi: 10.1007/s11251-016-9376-y
- Lajoie, S. P., & Gube, M. (2018). Adaptive expertise in medical education: Accelerating learning trajectories by fostering self-regulated learning. *Medical Teacher*, 40(8), 809–812. doi:10.1080/0142159x.2018.1485886
- Lambe, K. A., O'Reilly, G., Kelly, B. D., & Curristan, S. (2016). Dual-process cognitive interventions to enhance diagnostic reasoning: A systematic review. *BMJ Quality & Safety*, 25(10), 808-820. Doi: 10.1136/bmjqs-2015-004417
- Landis, R., Edwards, B. D., & Cortina, J. (2009). Correlated residuals among items in the estimation of measurement models. In C. E. Lance, & R. J. Vandenberg (Eds), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 195-214). New York: Routledge.
- Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, 10(3), 215. Doi: 10.1037/h0026570
- Lavine, H., & Snyder, M. (1996). Cognitive processing and the functional matching effect in persuasion: The mediating role of subjective perceptions of message quality. *Journal of Experimental Social Psychology*, 32(6), 580-604. Doi: 10.1006/jesp.1996.0026
- Leipziger, E., Tretter, T., & Gebhardt, M. (2012). Inklusion an oberfränkischen Grundschulen. *Zeitschrift für Heilpädagogik*, 10, 433–439.
- Lempp, Theresa (2011). Zivildienst als Sozialisationsfeld junger Männer – Biografische Lernprozesse und Männlichkeitskonstruktionen. Dissertation. Technische Universität Dresden.
- Leuders, T., & Prediger, S. (2016). *Scriptor Praxis: Flexibel differenzieren und fokussiert fördern im Mathematikunterricht*. Berlin: Cornelsen Scriptor.
- Leuders, T., Dörfler, T., Leuders, J., & Philipp, K. (2018). Diagnostic Competence of Mathematics Teachers: Unpacking a Complex Construct. In T. Leuders, K. Philipp, J. Leuders (Hrsg.), *Diagnostic Competence of Mathematics Teachers* (S. 3-31). Cham: Springer.
- Leuders, T., Nückles, M., Mikelskis-Seifert, S., & Philipp, K. (2019). Fachbezogene Pädagogische Kompetenzen und Wissenschaftsverständnis – Pädagogische Professionalität in Mathematik und Naturwissenschaften. In T. Leuders, M. Nückles, S. Mikelskis-Seifert, & K. Philipp (Hrsg.), *Pädagogische Professionalität in Mathematik und Naturwissenschaften* (S. 3-38). Wiesbaden: Springer. Doi: 10.1007/978-3-658-08644-2_1
- Levin, J. R. (1994). Crafting educational intervention research that's both credible and creditable. *Educational Psychology Review*, 6(3), 231-243. Doi: 10.1007/bf02213185
- Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social psychology*, 78(2), 223-246. Doi: 10.1037/0022-3514.78.2.223

- Lifshitz, H., Glaubman, R., & Issawi, R. (2004). Attitudes towards inclusion: The case of Israeli and Palestinian regular and special education teachers. *European Journal of Special Needs Education, 19*(2), 171-190. Doi: 10.1080/08856250410001678478
- Lindberg, S., Hasselhorn, M., & Lonnemann, J. (2018). Förderrelevante Diagnostik bei Lernstörungen. *Lernen und Lernstörungen, 7*, 197-201. Doi: 10.1024/2235-0977/a000233
- Linderkamp, F., & Grünke, M. (2007). Lern- und Verhaltensstörungen: Klassifikation, Prävalenz & Prognostik. In F. Linderkamp, & M. Grünke (Hrsg.), *Lern- und Verhaltensstörungen. Genese-Diagnostik-Intervention* (S. 14-28). Weinheim: Beltz.
- Linkersdörfer, J., Lonnemann, J., Lindberg, S., Hasselhorn, M., & Fiebach, C. J. (2012). Grey matter alterations co-localize with functional abnormalities in developmental dyslexia: An ALE meta-analysis. *PLOS One, 7*(8), e43122. Doi: 10.1371/journal.pone.0043122
- Liu, Y., Li, J., Zheng, Q., Zaroff, C. M., Hall, B. J., Li, X., & Hao, Y. (2016). Knowledge, attitudes, and perceptions of autism spectrum disorder in a stratified sampling of preschool teachers in China. *BMC Psychiatry, 16*(1), 142. Doi: 10.1186/s12888-016-0845-2
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267-286. Doi: 10.1037//0033-2909.127.2.267
- Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLOS One, 2*(9), 1-11. Doi: 10.1371/journal.pone.0000883
- Lortie, D. C. (1972). Team Teaching. Versuch der Beschreibung einer zukünftigen Schule. Team Teaching in der Schule. In H. W. Dechert (Hrsg.), *Team Teaching in der Schule* (S. 37-76). München: Piper.
- Machts, N., Kaiser, J., Schmidt, F. T., & Moeller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85-103. Doi: 10.1016/j.edurev.2016.06.003
- Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science, 25*(8), 1608-1618. Doi: 10.1177/0956797614535810
- Macnamara, B. N., Moreau, D., & Hambrick, D. Z. (2016). The relationship between deliberate practice and performance in sports: A meta-analysis. *Perspectives on Psychological Science, 11*(3), 333-350. Doi: 10.1177/1745691616635591
- Main, K. (2007). *A year long study of the formation and development of middle school teaching teams*. Doctoral dissertation. Griffiths University, Australia. Retrieved September 6, 2019, from <http://hdl.handle.net/10072/366879>
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*(6), 304-316. Doi: 10.3102/0013189x14545513
- Malecki, A. (2014). Sonderpädagogischer Förderbedarf – eine differenzierte Analyse. *WISTA – Wirtschaft und Statistik, 10*, 591-601.
- Marksteiner, T., Reinhard, M. A., Dickhäuser, O., & Sporer, S. L. (2012). How do teachers perceive cheating students? Beliefs about cues to deception and detection accuracy in the educational field. *European Journal of Psychology of Education, 27*(3), 329-350. Doi: 10.1007/s10212-011-0074-5

- Marksteiner, T., Reinhard, M. A., Lettau, F., & Dickhäuser, O. (2013). Bullying, cheating, deceiving: Teachers' perception of deceitful situations at school. *International Journal of Educational Psychology*, 2(2), 193-220. Doi: 10.1007/s10212-011-0074-5
- Marotzki, W. (1990). *Entwurf Einer Strukturalen Bildungstheorie. Biographietheoretische Auslegung von Bildungsprozessen in Hochkomplexen Gesellschaften*. Weinheim: Deutscher Studienverlag.
- Martinez-Murcia, F. J., Górriz, J. M., Ramírez, J., & Ortiz, A. (2016). A structural parametrization of the brain using hidden Markov models-based paths in Alzheimer's disease. *International Journal of Neural Systems*, 26(07), 1650024. Doi: 10.1142/S0129065716500246
- Marx, C., Goeze, A., Voss, T., Hoehne, V., Klotz, V., & Schrader, J. (2017). Pädagogisch-psychologisches Wissen von Lehrkräften aus Schule und Erwachsenenbildung: Entwicklung und Erprobung eines Testinstruments. *Zeitschrift für Erziehungswissenschaft*, 20(1), 165-200. Doi: 10.1007/978-3-658-15739-5_8
- Matarazzo, J. D. (1983). The reliability of psychiatric and psychological diagnosis. *Clinical Psychology Review*, 3(1), 103-145. Doi: 10.1016/0272-7358(83)90008-9
- McCaughan, K. (2013). Barrows' integration of cognitive and clinical psychology in PBL tutor guidelines. *Interdisciplinary Journal of Problem-Based learning*, 7(1), 4. Doi: 10.7771/1541-5015.1318
- McCoach, D. B., Black, A. C., & O'Connell, A. A. (2007). Errors of inference in structural equation modeling. *Psychology in the Schools*, 44(5), 461-470. Doi: 10.1002/pits.20238
- McDonald, K. M., Matesic, B., Contopoulos-Ioannidis, D. G., Lonhart, J., Schmidt, E., Pineda, N., & Ioannidis, J. P. (2013). Patient safety strategies targeted at diagnostic errors: A systematic review. *Annals of Internal Medicine*, 158, 381-389. Doi: 10.7326/0003-4819-158-5-201303051-00004
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., ... & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften: bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für pädagogische Psychologie*, 23(34), 223-235. Doi: 10.1024/1010-0652.23.34.223
- McNeill, N. J., Douglas, E. P., Koro-Ljungberg, M., Therriault, D. J., & Krause, I. (2016). Undergraduate students' beliefs about engineering problem solving. *Journal of Engineering Education*, 105(4), 560-584. Doi: 10.1002/jee.20150
- Merk, S., Cramer, C., Dai, N., Bohl, T., & Syring, M. (2018). Faktorielle Validität der Einstellungen von Lehrkräften zu heterogenen Lerngruppen. *Journal for Educational Research Online*, 10(2), 34-53. URN: urn:nbn:de:0111-pedocs-161349
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco, CA: Jossey-Bass.
- Moeyaert, M., Ferron, J. M., Beretvas, S. N., & van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, 52(2), 191-211. Doi: 10.1016/j.jsp.2013.11.003
- Moeyaert, M., Ugille, M., Ferron, J. M., Onghena, P., Heyvaert, M., Beretvas, S. N., & van den Noortgate, W. (2015). Estimating intervention effects across different types of single-subject experimental designs: Empirical illustration. *School Psychology Quarterly*, 30(1), 50. Doi: 10.1037/spq0000068

- Mohd-Ali, S., Baharun, H., Harun, H., Mat Saad, N. S., Puteh-Behak, F., Massari, N., ... Ahmad Mahir, N. (2016). Problem-based learning (PBL), case-design training and model for language practitioners. *IJOABI*, 7(7), 1-3. Doi: 10.18769/ijasos.336976
- Moldavsky, M., Pass, S., & Sayal, K. (2014). Primary school teachers' attitudes about children with attention deficit/hyperactivity disorder and the role of pharmacological treatment. *Clinical Child Psychology and Psychiatry*, 19(2), 202-216. Doi: 10.1177/1359104513485083
- Moll, K., Kunze, S., Neuhoﬀ, N., Bruder, J., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PLOS One*, 9(7), e103537. Doi: 10.1371/journal.pone.0103537
- Moll, K., Wallner, R., & Landerl, K. (2012). Kognitive Korrelate der Lese-, Leserechtschreib- und der Rechtschreibstörung. *Lernen und Lernstörungen*, 1, 7-19. Doi: 10.1024/2235-0977/a000002
- Monteiro, S. D., Sherbino, J. D., Ilgen, J. S., Dore, K. L., Wood, T. J., Young, M. E., ... & Howey, E. (2015). Disrupting diagnostic reasoning: Do interruptions, instructions, and experience affect the diagnostic accuracy and response time of residents and emergency physicians? *Academic Medicine*, 90(4), 511-517. Doi: 10.1097/acm.0000000000000614
- Monteiro, S. M., & Norman, G. (2013). Diagnostic reasoning: Where we've been, where we're going. *Teaching and Learning in Medicine*, 25(1), S26-S32. Doi: 10.1080/10401334.2013.842911
- Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology*, 53(1), 17-29. Doi: 10.1348/000711000159150
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364-386. Doi: 10.1177/1094428106291059
- Mulder, M. (2011). The concept of competence: Blessing or curse? In I. Torniainen, S. Mahlamäki-Kultanen, P. Nokelainen, & P. Iisley (Eds.), *Innovations for Competence Management* (pp. 11-24). Lahti: Lahti University of Applied Sciences.
- Mulholland, S. (2016). ADHD-specific knowledge and attitudes of teachers (ASKAT): Development and validation of a new research instrument. *International Journal of Educational Research*, 77, 109-116. Doi: 10.1016/j.ijer.2016.03.010
- Mulholland, S. M., Cumming, T. M., & Jung, J. Y. (2015). Teacher attitudes towards students who exhibit ADHD-type behaviours. *Australasian Journal of Special Education*, 39(1), 15-36. Doi: 10.1017/jse.2014.18
- Mullen, E. J. (2016). Reconsidering the 'idea' of evidence in evidence-based policy and practice. *European Journal of Social Work*, 19(3-4), 310-335. Doi: 10.1080/13691457.2015.1022716
- Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). Pros and cons of structural equation modeling. *Methods Psychological Research Online*, 8(2), 1-22.
- Naumann, J. (1974). *Medien-Märkte und Curriculumrevision in der BRD. Eine bildungsökonomische Studie zu den Entstehungsbedingungen und Verbreitungsmechanismen von Lernmitteln und Unterrichtstechnologien*. Berlin: Max-Planck Institut für Bildungsforschung.
- Neumann, M., Schnyder, I., Trautwein, U., Niggli, A., Lüdtke, O., & Cathomas, R. (2007). Schulformen als differenzielle Lernmilieus. *Zeitschrift für Erziehungswissenschaft*, 10(3), 399-420. Doi: 10.1007/s11618-007-0043-6

- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nickerson, R. S. (1999). How we know-and sometimes misjudge-what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737-759. Doi: 10.1037//0033-2909.125.6.737
- Niedermaier, G. (Hrsg.) (2012). *Kompetenzen entwickeln, messen und bewerten*. Linz: Trauner Verlag.
- Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, 10(3), 186-213. Doi: 10.1207/s15327957pspr1003_1
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139-154. Doi: 10.1016/j.jmp.2008.12.005
- Nohl, A. M. (2015). Typical phases of transformative learning: A practice-based model. *Adult Education Quarterly*, 65(1), 35-49. Doi: 10.1177/0741713614558582
- Noordzij, G., & Te Lindert, A. (2010). *The effects of goal orientation and quality of problems on students' motivation in a problem based learning environment*. Poster presented at the self-determination theory conference, Gent, Belgium.
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15(5), 625-632. Doi: 10.1007/s10459-010-9222-y
- Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The causes of errors in clinical reasoning: Cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine*, 92(1), 23-30. Doi: 10.1097/acm.0000000000001421
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. Doi: 10.1126/science.aab2374
- Nuissl, E., Schiersmann, C., & Siebert, H. (2002). *Kompetenzentwicklung statt Bildungsziele*. REPORT Literatur- und Forschungsreport Weiterbildung (Bd. 49). Bielefeld: Bertelsmann.
- Nystrom, D. T., Williams, L., Paull, D. E., Graber, M. L., & Hemphill, R. R. (2016). A theory-integrated model of medical diagnosis. *Journal of Cognitive Engineering and Decision Making*, 10(1), 14-35. Doi: 10.1177/1555343415618965
- OECD (2009). *Creating effective teaching and learning environments. First results from TALIS*. Paris: OECD.
- OECD (2014). *TALIS 2014 Results. An International Perspective on Teaching and Learning, TALIS*. Paris: OECD.
- OECD (2018). *TALIS 2018 Results: An International Perspective on Teaching and Learning, TALIS*. Paris: OECD.
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe, & W. Helsper (Hrsg.), *Pädagogische Professionalität. Untersuchungen zum Typus pädagogischen Handelns* (S. 70-182). Frankfurt am Main: Suhrkamp.
- Oevermann, U. (2008). Profession contra Organisation? Strukturtheoretische Perspektiven zum Verhältnis von Organisation und Profession in der Schule. In W. Helsper, S. Busse, M. Hummrich, & R.-T. Krammer (Hrsg.), *Pädagogische Professionalität in Organisationen* (S. 55-77). Doi: 10.1007/978-3-531-90777-2_4

- 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. Doi: 10.1037/e567342014-001
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. Doi: 10.1126/science.aac4716
- Oser, F. (1997). Standards in der Lehrerbildung. Teil 1: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 15(1), 26-37.
- Osteen, P. (2010). An introduction to using multidimensional item response theory to assess latent factor structures. *Journal of the Society for Social Work and Research*, 1(2), 66-82. Doi: 10.5243/jsswr.2010.6
- Ostermann, A. (2018). Factors influencing the judgment accuracy of teachers. In T. Leuders, K. Philipp, & J. Leuders (Eds.). *Diagnostic Competence of Mathematics Teachers – Unpacking a complex construct in teacher education and teacher practice* (pp. 95-108). Cham: Springer. Doi: 10.1007/978-3-319-66327-2_5
- Ostermann, A., Leuders, T., & Philipp, K. (2019). Fachbezogene diagnostische Kompetenzen von Lehrkräften – Von Verfahren der Erfassung zu kognitiven Modellen zur Erklärung. In T. Leuders, M. Nückles, S. Mikelskis-Seifert, & Philipp, K. (Hrsg.), *Pädagogische Professionalität in Mathematik und Naturwissenschaften* (S. 93-116). Wiesbaden: Springer Spektrum. Doi: 10.1007/978-3-658-08644-2_4
- Ou, J., Xie, L., Wang, P., Li, X., Zhu, D., Jiang, R., ... Liu, T. (2013). Modeling brain functional dynamics via hidden Markov models. In Institute for Electrical and Electronical Engineers (Eds.), *6th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 569-572). Doi: 10.1109/NER.2013.6695998
- Oudman, S., van de Pol, J., Bakker, A., Moerbeek, M., & van Gog, T. (2018). Effects of different cue types on the accuracy of primary school teachers' judgments of students' mathematical understanding. *Teaching and Teacher Education*, 76, 214-226. Doi: 10.1016/j.tate.2018.02.007
- Pagano, R. R. (2010). *Understanding statistics in the behavioral sciences* (9th ed.). Belmont, CA: Thomson Wadsworth.
- Paleczek, L., Seifert, S., & Gasteiger-Klicpera, B. (2017). Influences on teachers' judgment accuracy of reading abilities on second and third grade students: A multilevel analysis. *Psychology in the Schools*, 54(3), 228-245. Doi: 10.1002/pits.21993
- Pant, A. (2018). *Teamplayer oder Einzelkämpfer – wie Lehrkräfte in Deutschland kooperieren. Ergebnisse einer Forsa-Studie*. Vortrag auf dem Kongress der Deutschen Schulakademie am 13. und 14. September 2018 in Berlin.
- Park, W. J., & Hwang, S. D. (2013). Effects of teachers' knowledge and empathy on educational intervention for ADHD: Focused on the mediating effect of empathy. *Journal of Korean Academy of Psychiatric and Mental Health Nursing*, 22(1), 45-55. Doi: 10.12934/jkpmhn.2013.22.1.45
- Petermann, F. (2018). Umschriebene Entwicklungsstörungen. *Kindheit und Entwicklung*, 27(1), 1-4. Doi: 10.1026/0942-5403/a000239

- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123-205. Doi: 10.1016/s0065-2601(08)60214-2
- Petty, R. E., & Wegener, D. T. (1998). Matching versus mismatching attitude functions: Implications for scrutiny of persuasive messages. *Personality and Social Psychology Bulletin* 24(3), 227-240. Doi: 10.1177/0146167298243001
- Peukert, H. (2000). Reflexionen über die Zukunft von Bildung. *Zeitschrift für Pädagogik*, 46(4), 507-524.
- Phillips, W. J., Fletcher, J. M., Marks, A. D., & Hine, D. W. (2016). Thinking styles and decision making: A meta-analysis. *Psychological Bulletin*, 142(3), 260-290. Doi: 10.1037/bul0000027
- Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York: Norton & Co.
- Pogge, D. L., Wayland-Smith, D., Zaccario, M., Borgaro, S., Stokes, J., & Harvey, P. D. (2001). Diagnosis of manic episodes in adolescent inpatients: Structured diagnostic procedures compared to clinical chart diagnoses. *Psychiatry Research*, 101(1), 47-54. Doi: 10.1016/S0165-1781(00)00248-1
- Polanyi, M. (1966). *The Tacit Dimension*. Chicago: University of Chicago Press.
- Popper, K. R. (1934). *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: Springer.
- Praetorius, S., Al-Kabbani, D., Bohndick, C., Hilkenmeier, J., König, S. T., Müsche, H. S., ... Klingsieck, K. B. (2016). Spielend Lehrer/in werden: Problembasiertes Lernen mit virtuellen Schülerinnen/Schülern. *Zeitschrift für Hochschulentwicklung*, 11(3), 231-245. Doi: 10.3217/zfhe-11-03/14
- Praetorius, A. K., Greb, K., Lipowsky, F., & Gollwitzer, M. (2010). Teachers' diagnostic perception – What role does students' grade performance play in the teacher's assessment of mathematical self-concepts? *Journal for Educational Research Online/Journal für Bildungsforschung Online*, 2(1), 121-144. URN: urn:nbn:de:0111-opus-45707
- Praetorius, A. K., Karst, K., Dickhäuser, O., & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 2, 81-91. Doi: 10.2378/peu2011.art30d
- Praetorius, A. K., Koch, T., Scheunpflug, A., Zeinz, H., & Dresel, M. (2017). Identifying determinants of teachers' judgment (in) accuracy regarding students' school-related motivations using a Bayesian cross-classified multi-level model. *Learning and Instruction*, 52, 148-160. Doi: 10.1016/j.learninstruc.2017.06.003
- Praetorius, A. K., Lipowsky, F., & Karst, K. (2012). Diagnostische Kompetenz von Lehrkräften: Aktueller Forschungsstand, unterrichtspraktische Umsetzbarkeit und Bedeutung für den Unterricht. In R. Lazarides, & A. Ittel (Hrsg.), *Differenzierung im mathematisch-naturwissenschaftlichen Unterricht. Implikationen für Theorie und Praxis* (S. 115-146). Bad Heilbrunn: Klinkhardt.
- Prenzel, M., & Gräsel, M. (2015). Der Startschuss ist gefallen. Die Qualitätsoffensive Lehrerbildung läuft. *Schulverwaltung spezial*, 5, 4-7.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand M., ... Schielefe, U. (Hrsg.). (2005). *PISA 2003: Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* Münster: Waxmann.

- Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1-20. Doi: 10.1017/s0140525x02000018
- Price, G. R., Yeo, D. J., Wilkey, E. D., & Cutting, L. E. (2018). Prospective relations between resting-state connectivity of parietal subdivisions and arithmetic competence. *Developmental Cognitive Neuroscience*, 30, 280-290. Doi: 10.1016/j.dcn.2017.02.006
- Pullen, P. C. (2017). Prevalence of LD From Parental and Professional Perspectives: A Comparison of the data from the National Survey of Children's Health and the Office of Special Education Programs' Reports to Congress. *Journal of Learning Disabilities*, 50(6), 701-711. Doi: 10.1177/0022219416659447
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71-90. Doi: 10.1016/j.dr.2016.06.004
- Rajapakse, J. C., & Zhou, J. (2007). Learning effective brain connectivity with dynamic Bayesian networks. *Neuroimage*, 37(3), 749-760. Doi: 10.1016/j.neuroimage.2007.06.003
- Ramaekers, S. P. J. (2011). *On the development of competence in solving clinical problems: Can it be taught? Or can it only be learned?* Doctoral dissertation, Utrecht University, Netherlands. Retrieved September 6, 2019, from <https://dspace.library.uu.nl/handle/1874/208653>
- Rapp, D. N., & Kendeou, P. (2009). Noticing and revising discrepancies as texts unfold. *Discourse Processes*, 46(1), 1-24. Doi: 10.1080/01638530802629141
- Rasch, D., & Schott, D. (2018). *Mathematical Statistics*. Oxford: Wiley. Doi: 10.1002/9781119385295
- Raufaste, E., Eyrolle, H., & Mariné, C. (1998). Pertinence generation in radiological diagnosis: Spreading activation and the nature of expertise. *Cognitive Science*, 22(4), 517-546. Doi: 10.1016/S0364-0213(99)80048-2
- Rauner, F., & Zeymer, H. (1991). *Auto und Beruf: Technischer Wandel und Berufsbildung im Kfz-Gewerbe*. Bremen: Donat.
- Ravitz, J. (2009). Summarizing findings and looking ahead to a new generation of PBL research. *Interdisciplinary Journal of Problem-based Learning*, 3(1), 4-11. Doi: 10.7771/1541-5015.1088
- Reeves, P. M., Pun, W. H., & Chung, K. S. (2017). Influence of teacher collaboration on job satisfaction and student achievement. *Teaching and Teacher Education*, 67, 227-236. Doi: 10.1016/j.tate.2017.06.016
- Reh, S., & Schelle, C. (2010). Arbeiten an Fällen in einem „Lehr-Forschungs-Projekt“. In B. Koch-Priewe, F.-U. Kolbe & J. Wildt (Hrsg.), *Grundlagenforschung und mikrodidaktische Reformansätze zur Lehrerbildung* (S. 197-211). Bad Heilbrunn: Klinkhardt.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E. & Köller, O. (Hrsg.) (2016). *PISA 2015. Eine Studie zwischen Kontinuität und Innovation*. Münster: Waxmann.
- Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial intelligence*, 32(1), 57-95. Doi: 10.1016/0004-3702(87)90062-2
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1-37. Doi: 10.1111/cogs.12086
- Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, 31(2), 115-121. Doi: 10.1207/s15326985ep3102_3

- Retelsdorf, J., Bauer, J., Gebauer, S. K., Kauper, T., & Möller, J. (2014). Erfassung berufsbezogener Selbstkonzepte von angehenden Lehrkräften (ERBSE-L). *Diagnostica*, 60, 98-110. Doi: 10.1026/0012-1924/a000108
- Rey, T., Lohse-Bossenz, H., Wacker, A., & Heyl, V. (2018). Adaptive Planungskompetenz bei angehenden Lehrkräften in der zweiten Phase der Lehrerbildung. Befunde einer Pilotierungsstudie aus Baden-Württemberg. *heiEDUCATION Journal. Transdisziplinäre Studien zur Lehrerbildung*, 1-2, 127-150. URN: <http://nbn-resolving.de/urn:nbn:de:bsz:16-heied-238298>
- Rheinberg, F. (2001). Bezugsnormen und Leistungsbeurteilung. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 59-71). Weinheim: Beltz.
- Richter, D., & Pant, H. A. (2016). *Lehrerkooperation in Deutschland. Eine Studie zu kooperativen Arbeitsbeziehungen bei Lehrkräften der Sekundarstufe I*. Gütersloh: Bertelsmann.
- Ricken, G. (2017). Kompetent sein für Inklusive Schulen heißt auch Diagnostizieren lernen. In M. Gercke, S. Opalinski, & T. Thonagel (Hrsg.), *Inklusive Bildung und gesellschaftliche Exklusion* (S. 187-199). Wiesbaden: Springer VS. Doi: 10.1007/978-3-658-17084-4_12
- Riek, K., & van Ophuysen, S. (2016). Nicht immer zählt nur Leistung – schulformabhängige Prädiktoren der Übergangsempfehlung. In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler, & R. Schmidt (Hrsg.), *Facetten grundschulpädagogischer und-didaktischer Forschung* (S. 13-18). Wiesbaden: Springer VS. Doi: 10.1007/978-3-658-11944-7_1
- Rödel, S. S. (2018). *Negative Erfahrungen und Scheitern im schulischen Lernen: phänomenologische und videographische Perspektiven*. Wiesbaden: Springer.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2., vollst. überarb. u. erw. Aufl.). Bern: Huber.
- Rotgans, J. I., Schmidt, H. G., Rosby, L. V., Tan, G. J., Mamede, S., Zwaan, L., & Low-Beer, N. (2019). Evidence supporting dual-process theory of medical diagnosis: A functional near-infrared spectroscopy study. *Medical eEducation*, 53(2), 143-152. Doi: 10.1111/medu.13681
- Roth, H. (1971). *Pädagogische Anthropologie. Bd. 2, Entwicklung und Erziehung: Grundlagen einer Entwicklungspädagogik*. Hannover: Schroedel.
- Rothland, M., Biederbeck, I., Grabosch, A., & Heiligttag, N. (2018). Autonomiestreben, Paritätsdenken und die Ablehnung von Kooperation bei Lehramtsstudierenden. *Zeitschrift für Erziehungswissenschaft*, 21(3), 589-610. Doi: 10.1007/s11618-017-0786-7
- Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology*, 76(3), 429-444. Doi: 10.1348/000709905x53589
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78. Doi: 10.1037/0003-066X.55.1.68
- Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine. What it is and what it isn't. *BMJ: British Medical Journal*, 313(7050), 170-171. Doi: 10.1136/bmj.313.7050.170c
- Sainani, K. (2010). The importance of accounting for correlated observations. *PM&R*, 2(9), 858-861. Doi: 10.1016/j.pmrj.2010.07.482

- Salem, S., Weskott, T., & Holler, A. (2017). Does narrative perspective influence readers' perspective-taking? An empirical study on free indirect discourse, psycho-narration and first-person narration. *Glossa: A Journal of General Linguistics*, 2(1), 61. Doi: 10.5334/gjgl.225
- Sangestani, G., & Khatiban, M. (2013). Comparison of problem-based learning and lecture-based learning in midwifery. *Nurse Education Today*, 33(8), 791-795. Doi: 10.1016/j.nedt.2012.03.010
- Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM Mathematics Education*, 43(1), 133-145. Doi: 10.1007/s11858-010-0292-3
- Santagata, R., & Sandholtz, J. H. (2018). Preservice teachers' mathematics teaching competence: Comparing performance on two measures. *Journal of Teacher Education*, 70(5), 1-13. Doi: 10.1177/0022487117753575
- Santagata, R., & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM Mathematics Education*, 48(1-2), 153-165. Doi: 10.1007/s11858-015-0737-9
- Sari, H. (2007). The influence of an in-service teacher training (INSET) programme on attitudes towards inclusion by regular classroom teachers who teach deaf students in primary schools in Turkey. *Deafness & Education International*, 9(3), 131-146. Doi: 10.1179/146431507790559996
- Savage, R. S., & Erten, O. (2015). Teaching in inclusive classrooms: The link between teachers attitudes-practices and student outcomes. *Journal of Psychology & Psychotherapy*, 5(6), 1-7. Doi: 10.4172/2161-0487.1000219
- Savalei, V. (2010). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 241-264. doi:10.1080/10705511003659375
- Schäfer, S., & Seidel, T. (2015). Noticing and reasoning of teaching and learning components by preservice teachers. *Journal for Educational Research Online*, 7(2), 34-58. URN: urn:nbn:de:0111-pedocs-114895
- Scheerens, J., & Blömeke, S. (2016). Integrating teacher education effectiveness research into educational effectiveness models. *Educational Research Review*, 18, 70-87. Doi: 10.1016/j.edurev.2016.03.002
- Scheiner, T. (2016). Teacher noticing: enlightening or blinding? *ZDM Mathematics Education*, 48(1-2), 227-238. Doi: 10.1007/s11858-016-0771-2
- Schlüter, A. (2010). Didaktische Kompetenz und Intersektionalität. In N. Auferkorte-Michaelis, D. P. A. Ladwig, & I. Stahr (Hrsg.), *Hochschuldidaktik für die Lehrpraxis: Interaktion und Innovation für Studium und Lehre an der Hochschule* (S. 157-168). Opladen: Budrich. Doi: 10.2307/j.ctvddzj6k.14
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, 6(4), 147-151. Doi: 10.1027/1614-2241/a000016
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210-222. Doi: 10.1016/j.hrmr.2008.03.003

- Schmitt, N., Golubovich, J., & Leong, F. T. (2011). Impact of measurement invariance on construct correlations, mean differences, and relations with external correlates: An illustrative example using Big Five and RIASEC measures. *Assessment*, 18(4), 412-427. Doi: 10.1177/1073191110373223
- Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, 6, 474. Doi: 10.3389/fpsyg.2015.00474
- Scholz, A., Krems, J. F., & Jahn, G. (2017). Watching diagnoses develop: Eye movements reveal symptom processing during diagnostic reasoning. *Psychonomic Bulletin & Review*, 24(5), 1398-1412. 10.3758/s13423-017-1294-8
- Schrader, F. W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt am Main: Lang.
- Schrader, F. W. (2008). Diagnoseleistungen und diagnostische Kompetenzen von Lehrkräften. In W. Schneider, & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie. Band 10* (S. 168-177). Göttingen: Hogrefe.
- Schrader, F. W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 237-245. Doi: 10.1024/1010-0652.23.34.237
- Schrader, F. W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerinnen-und Lehrerbildung*, 31(2), 154-165.
- Schrader, F. W. (2017). Diagnostische Kompetenz von Lehrkräften – Anmerkungen zur Weiterentwicklung des Konstrukts. In A. Südkamp, & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 247-255). Münster: Waxmann.
- Schrader, F. W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27-52.
- Schrader, F. W., Helmke, A., Hosenfeld, I., Halt, A. C., & Hochweber, J. (2006). Komponenten der Diagnosegenauigkeit von Lehrkräften: Ergebnisse aus Vergleichsarbeiten in der Grundschule. In F. Eder, A. Gastager, & F. Hofmann (Hrsg.), *Qualität durch Standards? Beiträge zum Schwerpunktthema der 67. Tagung der AEPF* (S. 265-278). Münster: Waxmann.
- Schulte-Koerne, G. (2014). Specific learning disabilities-from DSM-IV to DSM-5. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 42(5), 369-374. Doi: 10.1024/1422-4917/a000312
- Schum, D. (2012). Classifying forms and combinations of evidence: Necessary in a science of evidence. In P. Dawid, W. Twining, & D. Vasilaki (Eds.), *Evidence, inference and enquiry* (pp. 11-36). New York: Oxford University Press. Doi: 10.5871/bacad/9780197264843.003.0002
- Schürer, S., van Ophuysen, S. B., & Bloh, B. (2018). Inklusion und Schulqualität – Welche Erwartungen haben Lehramtsstudierende an die Konsequenzen schulischer Inklusion? In S. Miller, B., Holler-Nowitzki, B. Kottmann, S. Lesemann, B. Letmathe-Henkel, N. Meyer, & K. Velten (Hrsg.), *Profession und Disziplin* (S. 289-295). Wiesbaden: Springer VS. Doi: 10.1007/978-3-658-13502-7_36

- Schwab, S., & Seifert, S. (2015). Einstellungen von Lehramtsstudierenden und Pädagogikstudierenden zur schulischen Inklusion – Ergebnisse einer quantitativen Untersuchung. *Zeitschrift für Bildungsforschung*, 5(1), 73-87. Doi: 10.1007/s35834-014-0107-7
- Schwab, S., Seifert, S., & Gasteiger-Klicpera, B. (2014). Improving reading in children with German as a first or second language. *International Journal of Early Years Education*, 22(2), 210-222. Doi: 10.1080/09669760.2014.909308
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45(3), 513-523. Doi: 10.1037/0022-3514.45.3.513
- Schwarzer, R., & Schmitz, G. S. (1999). Kollektive Selbstwirksamkeitserwartung von Lehrern: Eine Längsschnittstudie in zehn Bundesländern. *Zeitschrift für Sozialpsychologie*, 30(4), 262-274. Doi: 10.1024//0044-3514.30.4.262
- Seaman, S. L., Algina, J., & Olejnik, S. F. (1985). Type I error probabilities and power of the rank and parametric ANCOVA procedures. *Journal of Educational Statistics*, 10(4), 345-367. Doi: 10.3102/10769986010004345
- Sedlmeier, P. (1996). Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. *Methods of Psychological Research Online*, 1(4), 41-63.
- Sedlmeier, P., & Renkewitz, F. (2013). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler* (2., aktual. Aufl.). Halbergmoos: Pearson.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. *Zeitschrift für Pädagogik*, 56, Beiheft, 296-306. Doi: 10.1024/1010-0652/a000100
- Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, 16(3), 228-240. Doi: 10.1016/j.learninstruc.2006.03.002
- Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education*, 27(2), 259-267. Doi: 10.1016/j.tate.2010.08.009
- Seifried, S., & Heyl, V. (2016). Konstruktion und Validierung eines Einstellungsfragebogens zu Inklusion für Lehrkräfte (EFI-L). *Empirische Sonderpädagogik*, 1, 22-35. URN: urn:nbn:de:0111-pedocs-118527
- Selmer, J., Jonasson, C., & Luring, J. (2013). Group conflict and faculty engagement: Is there a moderating effect of group trust? *Journal of Higher Education Policy and Management*, 35(1), 95-109. Doi: 10.1080/1360080X.2013.748477
- Sepucha, K., & Ozanne, E. M. (2010). How to define and measure concordance between patients' preferences and medical treatments: A systematic review of approaches and recommendations for standardization. *Patient education and counseling*, 78(1), 12-23. Doi: 10.1016/j.pec.2009.05.011
- Sermier Dessemontet, R., Benoit, V., Bless, G. (2011). Schulische Integration von Kindern mit einer geistigen Behinderung. Untersuchung der Entwicklung der Schulleistungen und der adaptiven

- Fähigkeiten, der Wirkung auf die Lernentwicklung der Mitschüler sowie der Lehrereinstellungen zur Integration. *Empirische Sonderpädagogik*, 3(4), 291-307.
- Shalev, R., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Child and Adolescent Psychiatry*, 9, 58-64. Doi: 10.1007/s007870070009
- Shalizi, C. R., & Crutchfield, J. P. (2002). Information bottlenecks, causal states, and statistical relevance bases: How to represent relevant information in memoryless transduction. *Advances in Complex Systems*, 5(01), 91-95. arXiv:nlin/0006025
- Shavelson, R. J. (2013). On an approach to testing and modeling competence. *Educational Psychologist*, 48(2), 73-86. Doi: 10.1080/00461520.2013.779483
- Shen, L. (2010). On a scale of state empathy during message processing. *Western Journal of Communication*, 74(5), 504-524. Doi: 10.1080/10570314.2010.512278
- Sherbino, J., Dore, K. L., Wood, T. J., Young, M. E., Gaissmaier, W., Kreuger, S., & Norman, G. R. (2012). The relationship between response time and diagnostic accuracy. *Academic Medicine*, 87(6), 785-791. Doi: 10.1097/acm.0b013e318253acbd
- Shippen, M. E., Crites, S. A., Houchins, D. E., Ramsey, M. L., & Simon, M. (2005). Preservice teachers' perceptions of including students with disabilities. *Teacher Education and Special Education*, 28(2), 92-99. Doi: 10.1177/088840640502800202
- Shulman, J. H. (1992). Case methods in teacher education. In J. H. Shulman (Ed.), *Case methods in teacher education* (pp. 13-17). New York: Teachers College Press.
- Shulman, L. (1981). Disciplines of inquiry in education: An overview. *Educational Researcher*, 10(6), 5-23. Doi: 10.3102/0013189x010006005
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14. Doi: 10.3102/0013189X015002004
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-23. Doi: 10.17763/haer.57.1.j463w79r56455411
- Shulman, L. (1992). Toward a pedagogy of cases. In J. H. Shulman (Ed.), *Case methods in teacher education* (S. 1-33). Columbia: Teachers College Press.
- Shulman, L., & Elstein, A. S. (1975) Studies in Problem-Solving, Judgement and Decision Making: Implications for Educational Research. *Review of Educational Research*, 3, 3-42. Doi: 10.3102/0091732X003001003
- Singer, M., & Kintsch, W. (2001). Text retrieval: A theoretical exploration. *Discourse Processes*, 31(1), 27-59. Doi: 10.1207/s15326950dp3101_2
- Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1), 81-96. Doi: 10.1111/j.1749-6632.2009.04418.x
- Skagerlund, K., Bolt, T., Nomi, J. S., Skagenholt, M., Västfjäll, D., Träff, U., & Uddin, L. Q. (2019). Disentangling mathematics from executive functions by investigating unique functional connectivity patterns predictive of mathematics ability. *Journal of Cognitive Neuroscience*, 31(4), 560-573. Doi: 10.1162/jocn_a_01367

- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2), 108-131. Doi: 10.1207/s15327957pspr0402_01
- Solga, H., & Dombrowski, R. (2009). *Soziale Ungleichheiten in schulischer und außerschulischer Bildung. Stand der Forschung und Forschungsbedarf. Arbeitspapier*, 171. Verfügbar unter https://www.boeckler.de/pdf/p_arbp_171.pdf. Abgerufen am 06.09.2019.
- Somech, A. (2010). Participative decision making in schools: A mediating-moderating analytical framework for understanding school and teacher outcomes. *Educational Administration Quarterly*, 46(2), 174-209. Doi: 10.1177/1094670510361745
- Sonnenberg, F. A., & Beck, J. R. (1993). Markov models in medical decision making: a practical guide. *Medical Decision Making*, 13(4), 322-338. Doi: 10.1177/0272989X9301300409
- Sophian, C. (1997). Beyond competence: The significance of performance for conceptual development. *Cognitive Development*, 12(3), 281-303. Doi: 10.1016/S0885-2014(97)90001-0
- Spiel, C., & Strohmeier, D. (2012). Evidence-based practice and policy: When researchers, policy makers, and practitioners learn how to work together. *European Journal of Developmental Psychology*, 9(1), 150-162. Doi: 10.1080/17405629.2011.616776
- 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. Doi: 10.1024/1010-0652.19.12.85
- Spreng, R. N., & Mar, R. A. (2012). I remember you: a role for memory in social cognition and the functional neuroanatomy of their interaction. *Brain Research*, 1428, 43-50. Doi: 10.1016/j.brainres.2010.12.024
- Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: A systematic review of empirical mathematics education research. *ZDM Mathematics Education*, 48(1-2), 1-27. Doi: 10.1007/s11858-016-0775-y
- Stang, J., & Urhahne, D. (2016). Wie gut schätzen Lehrkräfte Leistung, Konzentration, Arbeits- und Sozialverhalten ihrer Schülerinnen und Schüler ein? Ein Beitrag zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 63, 204-219. Doi: 10.2378/peu2016.art18d
- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6), 1467-1478. Doi: 10.1037//0022-3514.48.6.1467
- Staub-Bernasconi, S. (1983). *Soziale Probleme-Dimensionen ihrer Artikulation: Umriss einer Theorie sozialer Probleme als Beitrag zu einem theoretischen Bezugsrahmen sozialer Arbeit*. Diessenhofen: Rüegger.
- Steinmetz, H. (2013). Analyzing observed composite differences across groups. *Methodology*, 9, 1-12. Doi: 10.1027/1614-2241/a000049
- Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher*, 24(6), 9-17. Doi: 10.3102/0013189X024006009
- Stiller, J., & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, 29(1), 93-104. Doi: 10.1016/j.socnet.2006.04.001

- Streit, C., Rüede, C., Weber, C., & Graf, B. (2019). Zur Verknüpfung von Lernstandeinschätzung und Weiterarbeit im Arithmetikunterricht: Ein kontrastiver Vergleich zur Charakterisierung diagnostischer Expertise. *Journal für Mathematik-Didaktik*, 40(1), 37-62. Doi: 10.1007/s13138-018-0136-1
- Strobel, J., & van Barneveld, A. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary Journal of Problem-based Learning*, 3(1). Doi: 10.7771/1541-5015.1046
- Stühling, J., & Köppe, T. (2016). Two approaches to defining internal, external, and zero-focalization. *Semiotica*, 2016(210), 191-207. Doi: 10.1515/sem-2016-0056
- Stürmer, K., Könings, K. D., & Seidel, T. (2013). Declarative knowledge and professional vision in teacher education: Effects of courses in teaching and learning. *British Journal of Educational Psychology*, 83(3), 467-483. Doi: 10.1111/j.2044-8279.2012.02075.x
- Südkamp, A., & Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum: Direkte und indirekte Einschätzungen von Schülerleistungen. *Zeitschrift für Pädagogische Psychologie*, 23(34), 161-174. Doi: 10.1024/1010-0652.23.34.161
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743. Doi: 10.1037/a0027627
- Südkamp, A., & Praetorius, A. K. (Hrsg.) (2017). *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen*. Münster: Waxmann.
- Sungur, S., & Tekkaya, C. (2006). Effects of problem-based learning and traditional instruction on self-regulated learning. *Journal of Educational Research*, 99(5), 307-320. Doi: 10.3200/joer.99.5.307-320
- Swanson, H. L., O'Connor, J. E., & Cooney, J. B. (1990). An information processing analysis of expert and novice teachers' problem solving. *American Educational Research Journal*, 27(3), 533-556. Doi: 10.3102/00028312027003533
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4), 295-312. Doi: 10.1016/0959-4752(94)90003-5
- Sweller, J. (2003). Evolution of human cognitive architecture. *The Psychology of Learning and Motivation*, 43, 215-266. Doi: 10.3200/JOER.99.5.307-320
- Syring, M., Bohl, T., Kleinknecht, M., Kuntze, S., Rehm, M., & Schneider, J. (2015). Videos oder Texte in der Lehrerbildung? Effekte unterschiedlicher Medien auf die kognitive Belastung und die motivational-emotionalen Prozesse beim Lernen mit Fällen. *Zeitschrift für Erziehungswissenschaften*, 18, 667-685. Doi: 10.1007/s11618-015-0631-9
- Syring, M., Tillmann, T., Weiß, S., & Kiel, E. (2018). Positive Einstellung zur Inklusion – ablehnende Haltung zur Umsetzung in der Schule: Analyse des Widerspruchs durch Überprüfung eines aus der Heterogenitätsforschung adaptierten Messinstruments für die Inklusion an Lehramtsstudierenden. *Psychologie in Erziehung und Unterricht*, 65(3), 206-220. Doi: 10.2378/peu2018.art12d

- Szczesny, M., & Watermann, R. (2011). Differenzielle Einflüsse von Familie und Schulform auf Leseleistung und soziale Kompetenzen. *Journal for Educational Research Online*, 3(1), 168-193. URN: urn:nbn:de:0111-opus-46879
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen, & J. S. Lang (Hrsg.), *Testing Structural Equation Models. A Sage Focus Edition. Book 154* (pp. 10-39). Newbury Park: Sage.
- Terhart, E. (2002). *Standards für die Lehrerbildung: eine Expertise für die Kultusministerkonferenz*. Münster: Institut für Schulpädagogik und Allgemeine Didaktik. Westfälische Wilhelms-Universität..
- Tettegah, S., & Anderson, C. J. (2007). Pre-service teachers' empathy and cognitions: Statistical analysis of text data by graphical models. *Contemporary Educational Psychology*, 32(1), 48-82. Doi: 10.1016/j.cedpsych.2006.10.010
- Thilakaratne, R., & Kvan, T. (2006). Competence-based assessment in professional education validation. *Quality in Higher Education*, 12(3), 315-327. Doi: 10.1080/13538320601051093
- Thomas, K., Schulte-Körne, G., & Hasselhorn, M. (2015). Stichwort Entwicklungsstörungen schulischer Fertigkeiten. *Zeitschrift für Erziehungswissenschaft*, 18(3), 431-451. Doi: 10.1007/s11618-015-0642-6
- Tischler, T., Daseking, M., & Petermann, F. (2013). Effekt der Schulform auf die Entwicklung der Lesegeschwindigkeit. *Diagnostica*, 59(4), 215-226. Doi: 10.1026/0012-1924/a000093
- Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, 112(4), 578. Doi: 10.1037/0021-843x.112.4.578
- Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, 1, 31-65. Doi: 10.1146/annurev.clinpsy.1.102803.144239
- Toom, A., Pietarinen, J., Soini, T., & Pyhältö, K. (2017). How does the learning environment in teacher education cultivate first year student teachers' sense of professional agency in the professional community? *Teaching and Teacher Education*, 63, 126–136. Doi: 10.1016/j.tate.2016.12.013
- Tressoldi, P. E. (2012). Replication unreliability in psychology: Elusive phenomena or "elusive" statistical power? *Frontiers in Psychology*, 3, 218. Doi: 10.3389/fpsyg.2012.00218
- Trittel, M., Gerich, M., & Schmitz, B. (2014). Training prospective teachers in educational diagnostics. In S. Krolak-Schwerdt, S. Glock, & M. Böhmer (Eds.), *Teachers' professional development. Assessment, training, and learning* (pp. 63-78), Boston: Sense. Doi: 10.1007/978-94-6209-536-6_5
- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education*, 45, 73-82. Doi: 10.1016/j.tate.2014.09.006
- Urhahne, D., Chao, S. H., Florineth, M. L., Luttenberger, S., & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student. *British Journal of Educational Psychology*, 81(1), 161-177. Doi: 10.1348/000709910x504500

- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35(2), 215-247. Doi: 10.3102/1076998609346961
- van de Pol, J., de Bruin, A. B., van Loon, M. H., & van Gog, T. (2019). Students' and teachers' monitoring and regulation of students' text comprehension: Effects of comprehension cue availability. *Contemporary Educational Psychology*, 56, 236-249. Doi: 10.1016/j.cedpsych.2019.02.001
- van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, 24(2), 244-276. Doi: 10.1016/j.tate.2006.11.005
- van Gog, T., Paas, F., van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237-244. Doi: 10.1037/1076-898X.11.4.237
- van Merriënboer, J. J. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs: Educational Technology.
- van Merriënboer, J. J., Clark, R. E., & De Croock, M. B. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, 50(2), 39-61. Doi: 10.1007/bf02504993
- van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning*. New York: Taylor & Francis. Doi: 10.4324/9781315113210-3
- van Mieghem, A., Verschueren, K., Petry, K., & Struyf, E. (2018). An analysis of research on inclusive education: A systematic search and meta review. *International Journal of Inclusive Education*, 1-15. Doi: 10.1080/13603116.2018.1482012
- van Ophuysen, S. (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnempfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38(4), 154-161. Doi: 10.1026/0049-8637.38.4.154
- van Ophuysen, S. (2010). Professionelle pädagogischdiagnostische Kompetenz. Eine theoretische und empirische Annäherung. In N. Berkemeyer, W. Bos, H. G. Holtappels, N. McElvany, & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 16, S. 203-234). Weinheim: Beltz Juventa.
- van Ophuysen, S., & Behrmann, L. (2015). Die Qualität pädagogischer Diagnostik im Lehrerberuf – Anmerkungen zum Themenheft „Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz“. *Journal für Bildungsforschung Online*, 7(2), 82-98. URN: urn:nbn:de:0111-pedocs-114911
- van Rossum, M. C., & Shippi, M. (2013). Computational modelling of memory retention from synapse to behaviour. *Journal of Statistical Mechanics: Theory and Experiment*, 3, P03007. Doi: 10.1088/1742-5468/2013/03/P03007
- Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2013). Team entitativity and teacher teams in schools: Towards a typology. *Frontline Learning Research*, 1(2), 86-98. Doi: 10.14786/flr.v1i2.23
- Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research review*, 15, 17-40. Doi: 10.1016/j.edurev.2015.04.002

- Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education*, 61, 47–59. Doi: 10.1016/j.tate.2017.06.021
- Verma, V., Gagliardi, F., & Ferretti, C. (2009). On pooling of data and measures. Università di Siena, Dipartimento di metodi quantitativi. Working Paper n. 84. Retrieved September 6, 2019, from https://www.researchgate.net/profile/Francesca_Gagliardi/publication/228991450_On_pooling_of_data_and_measures/links/00463521d0924d420c000000.pdf
- Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24, 80-91. Doi: 10.1016/j.tate.2007.01.004
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48. Doi: 10.18637/jss.v036.i03
- Villa, P. I. (2011). Symbolische Gewalt und ihr potenzielles Scheitern. Eine Annäherung zwischen Butler und Bourdieu. *Österreichische Zeitschrift für Soziologie*, 36(4), 51-69. Doi: 10.1007/s11614-011-0003-5
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477-511. Doi: 10.1007/s10648-016-9361-7
- von Aufschnaiter, C., Cappell, J., Dübbelde, G., Ennemoser, M., Mayer, J., Stiensmeier-Pelster, J., ... Wolgast, A. (2015). Diagnostische Kompetenz. Theoretische Überlegungen zu einem zentralen Konstrukt der Lehrerbildung. *Zeitschrift für Pädagogik*, 61(5), 738-758. URN: urn:nbn:de:0111-pedocs-127230
- von Aufschnaiter, C., & Rogge, C. (2010). Wie lassen sich Verläufe der Entwicklung von Kompetenz modellieren? *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 95-114.
- Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18(2), 187-223. Doi: 10.1007/s11618-015-0626-6
- Wagner, U., Handke, L., & Walter, H. (2015). The relationship between trait empathy and memory formation for social vs. non-social information. *BMC Psychology*, 3(1), 1-8. Doi: 10.1186/s40359-015-0058-3
- Walker, A., & Leary, H. (2009). A problem based learning meta analysis: Differences across problem pypes, implementation iypes, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, 3(1). doi:10.7771/1541-5015.1061
- Ward, P., Gore, J., Hutton, R., Conway, G. E., & Hoffman, R. R. (2018). Adaptive skill as the conditio sine qua non of expertise. *Journal of Applied Research in Memory and Cognition*, 7(1), 35-50. Doi: 10.1016/j.jarmac.2018.01.009
- Warwas, J., & Helm, C. (2018). Professional learning communities among vocational school teachers: Profiles and relations with instructional quality. *Teaching and Teacher Education*, 73, 43-55. Doi: 10.1016/j.tate.2018.03.012

- Weber, C., Danninger, D., & Feyerer, E. (2016). Segregation und Chancengerechtigkeit: Eine Mehrebenen-Cross-Country-Analyse. *Zeitschrift für Bildungsforschung*, 6(3), 225-243. Doi: 10.1007/s35834-016-0167-y
- Wedel, A. (2019). *POL-TF. Problemorientiertes Lernen mit Textfällen in Bereich Pädagogische Diagnostik. Manual*. Berlin: Technische Universität Berlin. DOI: 10.14279/depositonce-8514
- Wedel, A., & Müller, C. R. (2016). *Test zum pädagogischen Diagnosewissen von Lehrkräften (PDW)*. Unveröffentlichtes Manuskript. Berlin: Technische Universität Berlin.
- Wedel, A., & Müller, C. R. (2017). *Skala zur Erfassung des diagnostischen Selbstkonzepts von Lehrkräften (LSD)*. Unveröffentlichtes Manuskript. Berlin: Technische Universität Berlin.
- Wedel, A., & Müller, C. R. (2018). *Rekognitions- und Verifikationstest zum Erinnern diagnostischer Informationen aus Textdokumenten (ReMem)*. Unveröffentlichtes Manuskript. Berlin: Technische Universität Berlin.
- Wedel, A., Müller, C. R., & Greiner, F. (2019). *Welchen Effekt haben Intelligenz, Empathie, Textverstehen und Textmerkmale auf das spontane diagnostische Urteil von Lehramtsstudierenden?* Manuskript zur Begutachtung eingereicht.
- Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A. (2019a). Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen. In I. Gogolin, B. Hannover, & A. Scheunpflug (Hrsg.), *Evidenzbasierung in der Lehrkräftebildung. Edition der Zeitschrift für Erziehungswissenschaft Band 4* (S. 95-122). Wiesbaden: Springer VS.
- Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A. (2019b). Training teachers' diagnostic competence with problem-based learning: A pilot and replication study. In: *Teaching & Teacher Education*, 86, 1-14. Doi: 10.1016/j.tate.2019.102909
- Wedel, A., Pfetsch, J., & Ittel, A. (2019). Evidenzbasierung und Zusammenhangsanalysen in der Lehrkräftebildung – Diagnosewissen, Selbstkonzept und Einstellungen zu Inklusion in manifesten Wachstumsmodellen. In: *Journal für Psychologie*, 27(2). Im Erscheinen.
- Wegener, D. T., & Petty, R. E. (1998). The naive scientist revisited: Naive theories and social judgment. *Social Cognition*, 16(1), 1-7. Doi: 10.1521/soco.1998.16.1.1
- Weinert, F. E. (1999): Definition and Selection of Competencies: Theoretical and Conceptual Foundations. *Concepts of Competence*. Paris: OECD.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Ashland: Hogrefe & Huber.
- Weinert, F. E., Schrader, F. W., & Helmke, A. (1990). Educational Expertise. Closing the gap between educational research and classroom practice. *Social Psychology International*, 11, 163-180. Doi: 10.1177/0143034390113002
- Wenden, A. (1998). *Learner Strategies for Learner Autonomy*. Prentice Hall: Pearson.
- Westheimer, J. (2008). Learning among colleagues: Teacher community and the shared enterprise of education. In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (pp. 756-783). New York: Routledge.

- Weston, R., & Gore Jr, P. A. (2006). A brief guide to structural equation modeling. *The counseling psychologist*, 34(5), 719-751. Doi: 10.1177/0011000006286345
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66(5), 297. Doi: 10.1037/h0040934
- Whitford, D. K., & Emerson, A. M. (2019). Empathy intervention to reduce implicit bias in pre-service teachers. *Psychological Reports*, 122(2), 670-688. Doi: 10.1177/0033294118767435
- Wijnen, M., Loyens, S. M., Wijnia, L., Smeets, G., Kroeze, M. J., & van der Molen, H. T. (2017). Is problem-based learning associated with students' motivation? A quantitative and qualitative study. *Learning Environments Research*, 21(2), 1-21. Doi: 10.1007/s10984-017-9246-9
- Wijnia, L., Loyens, S. M., & Derous, E. (2011). Investigating effects of problem-based versus lecture-based learning environments on student motivation. *Contemporary Educational Psychology*, 36(2), 101-113. Doi: 10.1016/j.cedpsych.2010.11.003
- Wilcox, R. R. (2012). Introduction to robust estimation and hypothesis testing (3rd ed.). Amsterdam: Academic Press. Doi: /10.1016/b978-0-12-386983-8.00010-x
- Wilhelm, M., & Brovelli, D. (2009). Problembasiertes Lernen (PBL) in der Lehrpersonenbildung: Der Drei-Phasen-Ansatz der Naturwissenschaften. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 27(2), 195-203.
- Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, 4, 433. doi:10.3389/fpsyg.2013.00433
- Williams, K. D., & Karau, S. J. (1991). Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology*, 61(4), 570. Doi: 10.1037//0022-3514.61.4.570
- Winter, S., & Krämer, N. C. (2016). Who's right: The author or the audience? Effects of user comments and ratings on the perception of online science articles. *Communications*, 3, 339-360. Doi: 10.1515/commun-2016-0008
- Winter, S., Krämer, N. C., Rösner, & Neubaum, L. G. (2015). Don't keep it (too) simple: How textual representations of scientific uncertainty affect laypersons' attitudes. *Journal of Language and Social Psychology* 34(3), 251-272. Doi: 10.1177/0261927x14555872
- Winterton, J. (2009). Competence across Europe: highest common factor or lowest common denominator? *Journal of European industrial training*, 33(8/9), 681-700. Doi: 10.1108/03090590910993571
- Winterton, J., Delamare-Le Deist, F., & Stringfellow, E. (2005). *Prototype typology of knowledge, skills and competences. CEDEFOP Reference Series 64*. Thessaloniki: Office for Official Publications of the European Communities.
- Wirkala, C., & Kuhn, D. (2011). Problem-based learning in K–12 education: Is it effective and how does it achieve its effects? *American Educational Research Journal*, 48(5), 1157-1186. Doi: 10.3102/0002831211419491
- Wood, R., & Power, C. (1987). Aspects of the competence-performance distinction: Educational, psychological and measurement issues. *Journal of Curriculum Studies*, 19(5), 409-424. Doi: 10.1080/0022027870190503

- Woodland, R., Lee, M. K., & Randall, J. (2013). A validation study of the teacher collaboration assessment survey. *Educational Research and Evaluation*, 19(5), 442-460. Doi: 10.1080/13803611.2013.795118
- Yadav, A., Phillips, M., Lundeberg, M., Koehler, M., Hilden, K., & Dirkin, K. H. (2011). If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *Journal of Computing in Higher Education*, 23(1), 15-37. Doi: 10.1007/s12528-011-9042-y
- Youssef, M. H., & Youssef, F. (2015). Knowledge of and attitudes toward ADHD among teachers insights from a Caribbean Nation. *SAGE Open*, 5, 1-8. Doi: 10.1177/2158244014566761
- Yung, Y. F., & Bentler, P. M. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 47(1), 63-84. Doi: 10.1111/j.2044-8317.1994.tb01025.x
- Zhou, J., & Urhahne, D. (2013). Teacher judgment, student motivation, and the mediating effect of attributions. *European Journal of Psychology of Education*, 28(2), 275-295. Doi: 10.1007/s10212-012-0114-9
- Zito, M. F. (2011). *Is working together worth it? Examining the relationship between the quality of teacher collaboration, instruction, and student achievement* (Doctoral dissertation). Retrieved September 6, 2019, from https://scholarworks.umass.edu/open_access_dissertations/404
- Zlatkin-Troitschanskaia, O., Pant, H. A., Kuhn, C., Toepper, M., & Lautenbach, C. (2016). Assessment practices in higher education & results of the German research program modeling and measuring competencies in higher education (KoKoHs). *Research & Practice in Assessment*, 11, 46-54. Doi: 10.1007/978-3-658-15486-8_5
- Zlatkin-Troitschanskaia, O., & Seidel, J. (2011). Kompetenz und ihre Erfassung – das neue „Theorie-Empirie-Problem“ der empirischen Bildungsforschung? In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Bildungsforschung* (S. 218-233). New York: VS Verlag. Doi: 10.1007/978-3-531-94025-0_16
- Zlatkin-Troitschanskaia, O., Shavelson, R. J. & Kuhn, C. (2015). The international state of research on measurement of competency in higher education. *Studies in Higher Education*, 40(3), 393-411. Doi: 10.1080/03075079.2015.1004241
- Zumbach, J., & Mandl, H. (2007). *Pädagogische Psychologie in Theorie und Praxis: Ein fallbasiertes Lehrbuch*. Göttingen: Hogrefe.

ANHANG

VERSUCHSMATERIAL TEXTFÄLLE

Variante 1: Dyskalkulie, external fokalisiert

- (1) Die Vorbereitung auf Ihre erste volle Unterrichtsstunde im Referendariat ist in der letzten Phase.
- (2) In dreißig Minuten geht die Stunde los und Sie gehen wie jeden Morgen zuerst ins Lehrerzimmer im zweiten Stock.
- (3) Der Korridor ist sehr hell beleuchtet.
- (4) Sie laufen zügig den Korridor entlang, dann bleiben Sie stehen und atmen tief durch.
- (5) Weiter vorne im Korridor läuft Ihre Klasse schon in den Unterrichtsraum.
- (6) Sie nehmen das Notenbuch der Klasse aus Ihrer Tasche und schauen auf die Schulnoten.
- (7) Die Leistungen Ihrer Schülerinnen und Schüler ist über die letzten Monate immer besser geworden.
- (8) Bei Friis sind die Noten in sprachlich dominierten Fächern wesentlich besser als in mathematischen Fächern.
- (9) Und die Noten sind besser, wenn Rita fehlt.
- (10) Im Lehrerzimmer angekommen schauen Sie zu Frau Meier, der Klassenlehrerin Ihrer Klasse.
- (11) Mit ihr haben Sie bisher noch nicht über Ihre Klasse oder die Planung Ihrer ersten Unterrichtsstunde gesprochen.
- (12) Als Frau Meier zu Ihnen aufschaut, erröten Sie und gehen zögerlich zu ihr herüber.
- (13) Dann fragen Sie Frau Meier nach ihrem Eindruck über die Klasse.
- (14) Frau Meier sagt:
- (15) „Die Klasse ist wirklich motiviert.
- (16) Mir fällt aber immer wieder Friis auf.
- (17) Meidet Gruppenarbeiten, immer perfektionistisch beim Auswendiglernen aber leider wenig Freundschaften.
- (18) Könnte Hochbegabung sein, ich bin mir da ziemlich sicher.
- (19) Friis hat zu Schuljahresbeginn auch die ganzen Fragebögen ausgefüllt, schauen Sie mal da rein, alles in der Schulkarte.
- (20) Da sieht es so aus, als wenn Friis Schulstress überwiegend mit Situationskontrolle und internaler Stabilisierung bewältigt.
- (21) Sagen Sie dann mal, wie es heute war.“
- (22) Sie bedanken sich und holen dann Ihr Unterrichtsmaterial.
- (23) Texte und Diagramme für Lernstationen und Aufgaben für die Arbeit in Fünfergruppen.
- (24) Sie beginnen den Unterricht mit der Gruppenarbeit, dabei wippen Sie die ganze Zeit mit dem linken Fuß.
- (25) Die Klasse ist sehr laut.
- (26) Als die Klasse anfängt, sich mit dem ausgeteilten Material zu beschäftigen, lächeln Sie.
- (27) Sie gehen im Unterrichtsraum umher und beobachten die Gruppen.
- (28) Bei der Gruppe mit Rita und Friis bleiben Sie stehen und hören zu.
- (29) Rita (17 Jahre): „Du hast das schon wieder falsch aufgeschrieben, Friis.“

- (30) Es heißt dreihundertfünfzehn und nicht dreißigtausend-und-fünfzehn.
- (31) Und ich hab' dir gesagt, wir müssen zuerst das Volumen berechnen, nicht die Fläche.
- (32) Dann berechnen, wie viele Kugeln hineinpassen und dann das übrigbleibende Volumen angeben.“
- (33) Sie stehen wie eingefroren da.
- (34) Friis hat wirklich dreißigtausend-und-fünfzehn aufgeschrieben und auf dem Block ein Durcheinander aus Formen aufgemalt.
- (35) Friis errötet und antwortet mit lauter und etwas zittriger Stimme.
- (36) Friis (16 Jahre): „Ist ja gut, ich hab's ja gehört!
- (37) Wenn du das ansagst, dauert es noch länger, als wenn ich es lese.
- (38) Ich bin auch so schon durcheinander.“
- (39) Kurze Zeit später klingelt es zum Unterrichtsschluss.
- (40) Sie gehen mit gesenktem Kopf und ohne nach links und rechts zu schauen zurück ins Lehrerzimmer.
- (41) Dort nehmen Sie die Schulakte von Friis in die Hand und blättern darin.
- (42) Beim DAI-Fragebogen halten Sie an und schauen sich die Antworten an.
- (43) Die ganze Klasse zeigt eine überdurchschnittlich hohe wissensbezogene Angstausslösung.
- (44) Friis hat aber eher niedrige Werte in der emotionalen Angstmanifestation.
- (45) Friis hat zudem in der Selbsteinschätzung zu Hoffnungslosigkeit und zu besorgniserregenden Gedanken im Vergleich zu Gleichaltrigen einen Prozentrang von 68 und in der Klasse einen Prozentrang von 62.
- (46) Ihre Schultern entspannen sich und Sie atmen geräuschvoll aus.
- (47) Sie lesen weiter und sehen, dass Friis auch sehr motiviert ist.
- (48) Plötzlich öffnet sich die Tür des Lehrerzimmers und Sie zucken zusammen.
- (49) Frau Meier kommt auf Sie zu und Sie rutschen auf Ihrem Stuhl hin und her.
- (50) Was erzählen Sie ihr nur?

Variante 2: Dyskalkulie, psycho-narrativ

- (1) Die Vorbereitung auf Ihre erste volle Unterrichtsstunde im Referendariat ist in der letzten Phase.
- (2) In dreißig Minuten geht die Stunde los und Sie gehen wie jeden Morgen zuerst ins Lehrerzimmer im zweiten Stock.
- (3) Der Korridor ist unangenehm grell beleuchtet.
- (4) Sie sind ohnehin schon aufgereggt, aber das macht es noch schlimmer.
- (5) Sie atmen einmal ganz tief durch und bemerken, dass Sie mit Ihren Gedanken ganz woanders waren.
- (6) Dann sehen Sie, dass Ihre Klasse schon in den Unterrichtsraum geht.
- (7) Wenn Sie Ihre Klasse so vor sich sehen, sind Sie sehr stolz.
- (8) Denn die Leistung Ihrer Schülerinnen und Schüler ist immer besser geworden.
- (9) Außer Friis.
- (10) Friis ist in sprachlich dominierten Fächern wesentlich besser geworden als in mathematischen Fächern.
- (11) Komisch ist, dass die Noten von Friis immer besser sind, wenn Rita gerade fehlt.
- (12) Eigentlich wollten Sie darüber schon immer mal mit der Klassenlehrerin Frau Meier sprechen.
- (13) Sie schämen sich etwas, dass Sie das bislang immer aufgeschoben haben.
- (14) Zudem merken Sie jetzt, dass es für die Binnendifferenzierung in Ihrer ersten Unterrichtsstunde wirklich wichtig wäre.
- (15) Im Lehrerzimmer angekommen gehen Sie deswegen gleich zu Frau Meier und fragen sie nach ihrem Eindruck über die Klasse.
- (16) Frau Meier sagt:
- (17) „Die Klasse ist wirklich motiviert.
- (18) Mir fällt aber immer wieder Friis auf.
- (19) Meidet Gruppenarbeiten, immer perfektionistisch beim Auswendiglernen aber leider wenig Freundschaften.
- (20) Könnte Hochbegabung sein, ich bin mir da ziemlich sicher.
- (21) Friis hat zu Schuljahresbeginn auch die ganzen Fragebögen ausgefüllt, schauen Sie mal da rein, alles in der Schulkarte.
- (22) Da sieht es so aus, als wenn Friis Schulstress überwiegend mit Situationskontrolle und internaler Stabilisierung bewältigt.
- (23) Sagen Sie dann mal, wie es heute war.“
- (24) Ganz erleichtert bedanken Sie sich bei Frau Meier und holen Ihr Unterrichtsmaterial.
- (25) Texte und Diagramme für Lernstationen und Aufgaben für die Arbeit in Fünfergruppen.
- (26) Sie beginnen den Unterricht mit der Gruppenarbeit, obwohl Ihnen dabei etwas mulmig zumute ist.
- (27) Die Klasse ist so laut, dass Sie besorgt sind, ob sie die Arbeit einfach verweigert.
- (28) Umso mehr freuen Sie sich, als die Gruppen anfangen.
- (29) Sie finden, dass alle konzentriert und angeregt arbeiten.

-
- (30) Nach kurzer Zeit ertappen Sie sich dabei, dass Sie nur noch die Gruppe mit Friis beobachten.
 - (31) Sie fragen sich, warum Rita und Friis so angespannt wirken und hören deshalb genauer hin.
 - (32) Rita (17 Jahre): „Du hast das schon wieder falsch aufgeschrieben, Friis.
 - (33) Es heißt dreihundertfünfzehn und nicht dreißigtausend-und-fünfzehn.
 - (34) Und ich hab' dir gesagt, wir müssen zuerst das Volumen berechnen, nicht die Fläche.
 - (35) Dann berechnen, wie viele Kugeln hineinpassen und dann das übrigbleibende Volumen angeben.“
 - (36) Sie sind ganz erschrocken und irritiert, dass bei Friis wirklich dreißigtausend-und-fünfzehn steht und Sie bemerken auch, dass Friis ein Durcheinander aus Formen aufgemalt hat.
 - (37) Sie fühlen sich hilflos und wissen nicht, wie Sie reagieren sollen.
 - (38) Friis scheint ganz verzweifelt zu sein.
 - (39) Friis (16 Jahre): „Ist ja gut, ich hab's ja gehört!
 - (40) Wenn du das ansagst, dauert es noch länger, als wenn ich es lese.
 - (41) Ich bin auch so schon durcheinander.“
 - (42) Kurze Zeit später beenden Sie den Unterricht, die beobachtete Situation lässt sie aber nicht mehr los.
 - (43) Zurück im Lehrerzimmer fragen Sie sich bedrückt, was mit Friis wohl los ist.
 - (44) Da fällt Ihnen plötzlich die Schulakte ein.
 - (45) Sie blättern sie ganz aufgeregt durch und entdecken einen ausgefüllten DAI-Fragebogen.
 - (46) Die ganze Klasse zeigt eine überdurchschnittlich hohe wissensbezogene Angstauslösung.
 - (47) Friis hat aber eher niedrige Werte in der emotionalen Angstmanifestation.
 - (48) Ihnen fällt ein Stein vom Herzen.
 - (49) Dass Friis in der Selbsteinschätzung zu Hoffnungslosigkeit und zu besorgniserregenden Gedanken im Vergleich zu Gleichaltrigen einen Prozentrang von 68 und in der Klasse einen Prozentrang von 62 hat, erleichtert Sie noch mehr.
 - (50) Sie sind sich auch absolut sicher, dass Friis sehr motiviert ist.
 - (51) Plötzlich hören Sie, wie sich die Tür des Lehrerzimmers öffnet und schrecken hoch.
 - (52) Sie sehen Frau Meier auf sich zukommen und werden nervös.
 - (53) Was erzählen Sie ihr nur?

Variante 3: Prüfungsangst, external-fokalisiert

- (1) Die Vorbereitung auf Ihre erste volle Unterrichtsstunde im Referendariat ist in der letzten Phase.
- (2) In dreißig Minuten geht die Stunde los und Sie gehen wie jeden Morgen zuerst ins Lehrerzimmer im zweiten Stock.
- (3) Der Korridor ist sehr hell beleuchtet.
- (4) Sie laufen zügig den Korridor entlang, dann bleiben Sie stehen und atmen tief durch.
- (5) Weiter vorne im Korridor läuft Ihre Klasse schon in den Unterrichtsraum.
- (6) Sie nehmen das Notenbuch der Klasse aus Ihrer Tasche und schauen auf die Schulnoten.
- (7) Die Leistungen Ihrer Schülerinnen und Schüler ist über die letzten Monate immer besser geworden.
- (8) Bei Friis sind die Noten in sprachlich dominierten Fächern wesentlich besser als in mathematischen Fächern.
- (9) Und die Noten sind besser, wenn Rita fehlt.
- (10) Im Lehrerzimmer angekommen schauen Sie zu Frau Meier, der Klassenlehrerin Ihrer Klasse.
- (11) Mit ihr haben Sie bisher noch nicht über Ihre Klasse oder die Planung Ihrer ersten Unterrichtsstunde gesprochen.
- (12) Als Frau Meier zu Ihnen aufschaut, erröten Sie und gehen zögerlich zu ihr herüber.
- (13) Dann fragen Sie Frau Meier nach ihrem Eindruck über die Klasse.
- (14) Frau Meier sagt:
- (15) „Die Klasse ist wirklich motiviert.
- (16) Mir fällt aber immer wieder Friis auf.
- (17) Meidet Gruppenarbeiten, immer perfektionistisch beim Auswendiglernen aber leider wenig Freundschaften.
- (18) Könnte Hochbegabung sein, ich bin mir da ziemlich sicher.
- (19) Friis hat zu Schuljahresbeginn auch die ganzen Fragebögen ausgefüllt, schauen Sie mal da rein, alles in der Schulakte.
- (20) Da sieht es so aus, als wenn Friis Schulstress überwiegend mit Situationskontrolle und internaler Stabilisierung bewältigt.
- (21) Sagen Sie dann mal, wie es heute war.“
- (22) Sie bedanken sich und holen dann Ihr Unterrichtsmaterial.
- (23) Texte und Diagramme für Lernstationen und Aufgaben für die Arbeit in Fünfergruppen.
- (24) Sie beginnen den Unterricht mit der Gruppenarbeit, dabei wippen Sie die ganze Zeit mit dem linken Fuß.
- (25) Die Klasse ist sehr laut.
- (26) Als die Klasse anfängt, sich mit dem ausgeteilten Material zu beschäftigen, lächeln Sie.
- (27) Sie gehen im Unterrichtsraum umher und beobachten die Gruppen.
- (28) Bei der Gruppe mit Rita und Friis bleiben Sie stehen und hören zu.
- (29) Rita (17 Jahre): „Du hast das schon wieder falsch aufgeschrieben, Friis.
- (30) Es heißt dreihundertfünfzehn und nicht dreißigtausend-und-fünfzehn.

- (31) Und ich hab' dir gesagt, wir müssen zuerst das Volumen berechnen, nicht die Fläche.
- (32) Dann berechnen, wie viele Kugeln hineinpassen und dann das übrigbleibende Volumen angeben.“
- (33) Sie stehen wie eingefroren da und schauen zu Friis.
- (34) Friis errötet und antwortet mit lauter und etwas zittriger Stimme.
- (35) Friis (16 Jahre): „Ist ja gut, ich hab's ja gehört!
- (36) Wenn du das ansagst, dauert es noch länger, als wenn ich es lese.
- (37) Ich bin auch so schon durcheinander.“
- (38) Kurze Zeit später klingelt es zum Unterrichtsschluss.
- (39) Sie gehen mit gesenktem Kopf und ohne nach links und rechts zu schauen zurück ins Lehrerzimmer.
- (40) Dort nehmen Sie die Schulakte von Friis in die Hand und blättern darin.
- (41) Beim DAI-Fragebogen halten Sie an und schauen sich die Antworten an.
- (42) Die ganze Klasse zeigt eine überdurchschnittlich hohe wissensbezogene Angstausslösung.
- (43) Friis hat aber zusätzlich sehr hohe Werte in der emotionalen Angstmanifestation.
- (44) Friis hat zudem in der Selbsteinschätzung zu Hoffnungslosigkeit und zu besorgniserregenden Gedanken im Vergleich zu Gleichaltrigen einen Prozentrang von 92 und in der Klasse einen Prozentrang von 96.
- (45) Ihre Schultern verkrampfen sich und Sie atmen geräuschvoll aus.
- (46) Sie lesen weiter und sehen, dass Friis auch sehr motiviert ist.
- (47) Plötzlich öffnet sich die Tür des Lehrerzimmers und Sie zucken zusammen.
- (48) Frau Meier kommt auf Sie zu und Sie rutschen auf Ihrem Stuhl hin und her.
- (49) Was erzählen Sie ihr nur?

Variante 4: Prüfungsangst, psycho-narrativ

- (1) Die Vorbereitung auf Ihre erste volle Unterrichtsstunde im Referendariat ist in der letzten Phase.
- (2) In dreißig Minuten geht die Stunde los und Sie gehen wie jeden Morgen zuerst ins Lehrerzimmer im zweiten Stock.
- (3) Der Korridor ist unangenehm grell beleuchtet.
- (4) Sie sind ohnehin schon aufgereggt aber das macht es noch schlimmer.
- (5) Sie atmen einmal ganz tief durch und bemerken, dass Sie mit Ihren Gedanken ganz woanders waren.
- (6) Dann sehen Sie, dass Ihre Klasse schon in den Unterrichtsraum geht.
- (7) Wenn Sie Ihre Klasse so vor sich sehen, sind Sie sehr stolz.
- (8) Denn die Leistung Ihrer Schülerinnen und Schüler ist immer besser geworden.
- (9) Außer Friis.
- (10) Friis ist in sprachlich dominierten Fächern wesentlich besser geworden als in mathematischen Fächern.
- (11) Komisch ist, dass die Noten von Friis immer besser sind, wenn Rita gerade fehlt.
- (12) Eigentlich wollten Sie darüber schon immer mal mit der Klassenlehrerin Frau Meier sprechen.
- (13) Sie schämen sich etwas, dass Sie das bislang immer aufgeschoben haben.
- (14) Zudem merken Sie jetzt, dass es für die Binnendifferenzierung in Ihrer ersten Unterrichtsstunde wirklich wichtig wäre.
- (15) Im Lehrerzimmer angekommen gehen Sie deswegen gleich zu Frau Meier und fragen sie nach ihrem Eindruck über die Klasse.
- (16) Frau Meier sagt:
- (17) „Die Klasse ist wirklich motiviert.
- (18) Mir fällt aber immer wieder Friis auf.
- (19) Meidet Gruppenarbeiten, immer perfektionistisch beim Auswendiglernen aber leider wenig Freundschaften.
- (20) Könnte Hochbegabung sein, ich bin mir da ziemlich sicher.
- (21) Friis hat zu Schuljahresbeginn auch die ganzen Fragebögen ausgefüllt, schauen Sie mal da rein, alles in der Schulkarte.
- (22) Da sieht es so aus, als wenn Friis Schulstress überwiegend mit Situationskontrolle und internaler Stabilisierung bewältigt.
- (23) Sagen Sie dann mal, wie es heute war.“
- (24) Ganz erleichtert bedanken Sie sich bei Frau Meier und holen Ihr Unterrichtsmaterial.
- (25) Texte und Diagramme für Lernstationen und Aufgaben für die Arbeit in Fünfergruppen.
- (26) Sie beginnen den Unterricht mit der Gruppenarbeit, obwohl Ihnen dabei etwas mulmig zumute ist.
- (27) Die Klasse ist so laut, dass Sie besorgt sind, ob sie die Arbeit einfach verweigert.
- (28) Umso mehr freuen Sie sich, als die Gruppen anfangen.
- (29) Sie finden, dass alle konzentriert und angeregt arbeiten.

-
- (30) Nach kurzer Zeit ertappen Sie sich dabei, dass Sie nur noch die Gruppe mit Friis beobachten.
 - (31) Sie fragen sich, warum Rita und Friis so angespannt wirken und hören deshalb genauer hin.
 - (32) Rita (17 Jahre): „Du hast das schon wieder falsch aufgeschrieben, Friis.
 - (33) Es heißt dreihundertfünfzehn und nicht dreißigtausend-und-fünfzehn.
 - (34) Und ich hab' dir gesagt, wir müssen zuerst das Volumen berechnen, nicht die Fläche.
 - (35) Dann berechnen, wie viele Kugeln hineinpassen und dann das übrigbleibende Volumen angeben.“
 - (36) Sie sind ganz erschrocken und irritiert über Ritas Art.
 - (37) Sie fühlen sich hilflos und wissen nicht, wie Sie reagieren sollen.
 - (38) Friis scheint ganz verzweifelt zu sein.
 - (39) Friis (16 Jahre): „Ist ja gut, ich hab's ja gehört!
 - (40) Wenn du das ansagst, dauert es noch länger, als wenn ich es lese.
 - (41) Ich bin auch so schon durcheinander.“
 - (42) Kurze Zeit später beenden Sie den Unterricht, die beobachtete Situation lässt sie aber nicht mehr los.
 - (43) Zurück im Lehrerzimmer fragen Sie sich bedrückt, was mit Friis wohl los ist.
 - (44) Da fällt Ihnen plötzlich die Schulakte ein.
 - (45) Sie blättern sie ganz aufgeregt durch und entdecken einen ausgefüllten DAI-Fragebogen.
 - (46) Die ganze Klasse zeigt eine überdurchschnittlich hohe wissensbezogene Angstausslösung.
 - (47) Friis hat aber zusätzlich sehr hohe Werte in der emotionalen Angstmanifestation.
 - (48) Ihnen wird ganz mulmig.
 - (49) Als Sie sehen, dass Friis in der Selbsteinschätzung zu Hoffnungslosigkeit und zu besorgniserregenden Gedanken im Vergleich zu Gleichaltrigen einen Prozentrang von 92 und in der Klasse einen Prozentrang von 96 hat, fühlen Sie sich noch schlechter.
 - (50) Sie sind sich aber absolut sicher, dass Friis trotzdem sehr motiviert ist.
 - (51) Plötzlich hören Sie, wie sich die Tür des Lehrerzimmers öffnet und schrecken hoch.
 - (52) Sie sehen Frau Meier auf sich zukommen und bekommen ein wenig Panik.
 - (53) Was erzählen Sie ihr nur?

Tabelle 11

Logistische Regression für richtiges Urteil hoher und mindestens niedriger Spezifität

Schritt	Konstrukt	Richtiges Urteil mit hoher Spezifität			Richtiges Urteil mit mindestens niedriger Spezifität		
		<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>	<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>
1	Alter	0.50 (.20)	6.20, < .05	1.64	0.43 (.18)	5.53, < .05	1.54
	Geschlecht	-0.01 (.25)	0.00, > .05	0.99	0.25 (.20)	1.52, > .05	1.28
2	Alter	0.35 (.23)	2.36, > .05	1.42	0.40 (.20)	4.31, < .05	1.02
	Geschlecht	0.03 (.27)	0.01, > .05	1.03	0.26 (.20)	1.60, > .05	0.87
	FI	1.13 (.39)	8.56, < .01	3.09	0.19 (.20)	0.92, > .05	1.21
	AG	-0.17 (.28)	0.36, > .05	0.85	-0.03 (.20)	0.02, > .05	0.98
3	Alter	0.28 (.24)	1.40, > .05	1.32	0.34 (.20)	2.87, > .05	1.41
	Geschlecht	-0.16 (.31)	0.27, > .05	0.85	0.34 (.20)	0.06, > .05	1.06
	FI	1.12 (.39)	8.29, < .01	3.07	0.20 (.21)	0.95, > .05	1.23
	AG	-0.28 (.30)	0.86, > .05	0.76	-0.10 (.21)	0.24, > .05	0.90
	TE affektiv	-0.08 (.28)	0.07, > .05	0.93	0.01 (.23)	0.00, > .05	1.01
	TE kognitiv	0.48 (.35)	1.85, > .05	1.62	0.59 (.25)	5.53, < .05	1.80
4	Alter	0.27 (.23)	1.30, > .05	1.31	0.33 (.20)	2.70, > .05	1.40
	Geschlecht	-0.18 (.32)	0.34, > .05	0.83	0.05 (.23)	0.04, > .05	1.05
	FI	1.09 (.39)	7.75, < .01	2.97	0.17 (.21)	0.69, > .05	1.19
	AG	-0.31 (.31)	1.03, > .05	0.73	-0.12 (.21)	0.33, > .05	0.89
	TE affektiv	-0.06 (.29)	0.04, > .05	0.95	0.01 (.23)	0.00, > .05	1.01
	TE kognitiv	0.47 (.36)	1.71, > .05	1.60	0.59 (.25)	5.40, < .05	1.80
	Textstil	0.18 (.28)	0.40, > .05	1.19	0.19 (.20)	0.93, > .05	1.21
	Textinhalt	-0.21 (.28)	0.52, > .05	0.81	-0.02 (.20)	0.01, > .05	0.98

Tabelle 11 (Fortgeführt)

Logistische Regression für richtiges Urteil hoher und mindestens niedriger Spezifität (Fortsetzung)

Schritt	Konstrukt	Richtiges Urteil mit hoher Spezifität			Richtiges Urteil mit mindestens niedriger Spezifität		
		<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>	<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>
5	Alter	0.26 (.24)	1.20, > .05	1.30	0.39 (.21)	3.36, > .05	1.48
	Geschlecht	-0.19 (.32)	0.34, > .05	0.83	0.10 (.23)	0.17, > .05	1.10
	FI	1.10 (.40)	7.60, < .01	2.99	0.20 (.21)	0.86, > .05	1.22
	AG	-0.31 (.31)	0.96, > .05	0.74	-0.08 (.22)	0.14, > .05	0.92
	TE affektiv	-0.07 (.30)	0.05, > .05	0.94	0.07 (.24)	0.08, > .05	1.07
	TE kognitiv	0.40 (.38)	1.13, > .05	1.50	0.46 (.26)	3.05, > .05	1.58
	Textstil	0.16 (.30)	0.28, > .05	1.17	0.23 (.21)	1.14, > .05	1.26
	Textinhalt	-0.13 (.30)	0.18, > .05	0.88	0.07 (.24)	0.08, > .05	1.03
	<i>d'</i> _{wortwörtlich}	0.07 (.32)	0.05, > .05	1.07	-0.12 (.23)	0.29, > .05	0.88
	<i>d'</i> _{propositional}	0.28 (.44)	0.42, > .05	1.33	0.43 (.31)	1.93, > .05	1.54
	<i>d'</i> _{situational}	0.35 (.43)	0.67, > .05	1.42	0.65 (.30)	4.62, > .05	1.91

Tabelle 11 (Fortgeführt)

Logistische Regression für richtiges Urteil hoher und mindestens niedriger Spezifität (Fortsetzung)

Schritt	Konstrukt	Richtiges Urteil mit hoher Spezifität			Richtiges Urteil mit mindestens niedriger Spezifität		
		<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>	<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>
6	Alter	0.48 (.31)	2.40, > .05	1.62	0.45 (.24)	3.52, > .05	1.56
	Geschlecht	0.09 (.40)	0.05, > .05	1.10	0.28 (.26)	1.09, > .05	1.32
	FI	1.54 (.52)	8.90, < .01	4.67	0.20 (.24)	0.68, > .05	1.22
	AG	-0.15 (.37)	0.16, > .05	0.87	-0.08 (.23)	0.13, > .05	0.92
	TE affektiv	0.08 (.40)	0.04, > .05	1.08	0.25 (.28)	0.80, > .05	1.28
	TE kognitiv	-0.23 (.49)	0.22, > .05	0.80	0.25 (.29)	0.73, > .05	1.29
	Textstil	-0.07 (.35)	0.04, > .05	0.93	0.30 (.24)	1.53, > .05	1.35
	Textinhalt	-0.17 (.36)	0.23, > .05	0.84	0.09 (.24)	0.14, > .05	1.09
	<i>d'</i> _{wortwörtlich}	0.31 (.36)	0.74, > .05	1.36	-0.01 (.24)	0.00, > .05	0.99
	<i>d'</i> _{propositional}	0.94 (.56)	2.84, > .05	2.56	0.81 (.36)	4.99, < .05	2.24
	<i>d'</i> _{situational}	0.36 (.54)	0.44, > .05	1.43	0.95 (.36)	7.05, < .01	2.60
	SE affektiv	1.20 (.48)	6.36, < .05	3.31	0.87 (.33)	7.13, < .01	2.39
	SE kognitiv	-0.33 (.53)	0.38, > .05	0.72	0.15 (.34)	0.20, > .05	1.16
	KE	1.05 (.45)	5.40, < .05	2.85	0.50 (.30)	2.78, > .05	1.64
	IM	0.13 (.58)	0.05, > .05	1.14	-0.24 (.35)	0.47, > .05	0.79
	TI	0.12 (.44)	0.08, > .05	1.13	0.01 (.29)	0.00, > .05	1.00
	WAE	-0.42 (.51)	0.08, > .05	0.66	-0.91 (.38)	5.84, < .05	0.40

Tabelle 11 (Fortgeführt)

Logistische Regression für richtiges Urteil hoher und mindestens niedriger Spezifität (Fortsetzung)

Schritt	Konstrukt	Richtiges Urteil mit hoher Spezifität			Richtiges Urteil mit mindestens niedriger Spezifität		
		<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>	<i>B</i> (<i>SE</i>)	Walds χ^2 , <i>p</i>	<i>OR</i>
7	Alter	0.44 (.32)	1.95, > .05	1.56	0.45 (.25)	3.29, > .05	1.57
	Geschlecht	0.20 (.42)	0.22, > .05	1.22	0.27 (.26)	0.94, > .05	1.31
	FI	1.55 (.51)	9.19, < .01	4.72	0.20 (.24)	0.68, > .05	1.22
	AG	-0.05 (.39)	0.02, > .05	0.95	-0.07 (.24)	0.09, > .05	0.93
	TE affektiv	0.24 (.44)	0.30, > .05	1.27	0.25 (.29)	0.76, > .05	1.28
	TE kognitiv	-0.35 (.51)	0.47, > .05	0.71	0.25 (.30)	0.69, > .05	1.28
	Textstil	-0.03 (.35)	0.01, > .05	0.97	0.30 (.24)	1.51, > .05	1.35
	Textinhalt	-0.18 (.37)	0.24, > .05	0.84	0.08 (.24)	0.12, > .05	1.09
	<i>d'</i> _{wortwörtlich}	0.32 (.36)	0.77, > .05	1.37	-0.03 (.24)	0.01, > .05	0.97
	<i>d'</i> _{propositional}	1.08 (.60)	3.29, > .05	2.95	0.82 (.37)	5.10, < .05	2.28
	<i>d'</i> _{situational}	0.62 (.61)	1.05, > .05	1.37	0.96 (.38)	6.57, < .05	2.62
	SE affektiv	1.35 (.51)	6.92, < .01	3.85	0.88 (.33)	7.01, < .01	2.42
	SE kognitiv	-0.25 (.54)	0.22, > .05	0.78	0.15 (.35)	0.20, > .05	1.17
	KE	1.04 (.45)	5.29, < .05	2.83	0.50 (.30)	2.85, > .05	1.65
	IM	0.11 (.60)	0.03, > .05	1.11	-0.25 (.35)	0.49, > .05	0.78
	TI	0.15 (.45)	0.11, > .05	1.16	0.01 (.27)	0.00, > .05	1.00
	WAE	-0.62 (.55)	0.03, > .05	0.54	-0.92 (.39)	5.41, < .05	0.40
	D _{Selbst}	0.09 (.53)	0.03, > .05	1.09	0.15 (.26)	0.31, > .05	1.16
	D _{Andere}	-0.45 (.43)	1.08, > .05	0.64	0.00 (.27)	0.00, > .05	1.00

Anmerkungen. FI = Fluide Intelligenz, AG = Arbeitsgedächtnis, TE = Trait-Empathie, *d'* = *d'*-Wert der kognitiven Textrepräsentation, SE = State-Empathie, KE = Kognitive Elaboration, IM = Immersion, TI = Thematisches Interesse, WAE = Wahrgenommene Ähnlichkeit, D = Diagnose; *N* = 145.

In der Dissertation enthaltene Publikationen

1. Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A. (2020). Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen. In I. Gogolin, B. Hannover, & A. Scheunpflug (Hrsg.), *Evidenzbasierung in der Lehrkräftebildung. Edition der Zeitschrift für Erziehungswissenschaft Band 4* (S. 95-122). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-22460-8_5

Postprint Version. Reprinted by permission from Springer Nature: Springer, Zeitschrift für Erziehungswissenschaften, Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen, Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A., © 2020.

2. Wedel, A., Müller, C. R., Pfetsch, J., & Ittel, A. (2019). Training teachers' diagnostic competence with problem-based learning: A pilot and replication study. In: *Teaching & Teacher Education*, 86, 1-14. <https://doi.org/10.1016/j.tate.2019.102909>

Postprint Version. This manuscript is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

3. Wedel, A., Pfetsch, J., & Ittel, A. (2020). Evidenzbasierung und Zusammenhangsanalysen in der Lehrkräftebildung – Diagnosewissen, Selbstkonzept und Einstellungen zu Inklusion in manifesten Wachstumsmodellen. In: *Journal für Psychologie*, 27(2), 356-381. <https://doi.org/10.30820/0942-2285-2019-2-356>

Das Journal für Psychologie ist eine [Open Access](#) Zeitschrift, in der alle Beiträge als Volltexte frei zugänglich sind, entsprechend der Grundannahme, dass die freie öffentliche Verfügbarkeit von Forschung einem weltweiten Wissensaustausch zugute kommt.

Postprint Version. This manuscript is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

4. Wedel, A., Müller, C. R., Greiner, F. (In Review). Welchen Effekt haben Intelligenz, Empathie, Textverstehen und Textmerkmale auf das spontane diagnostische Urteil von Lehramtsstudierenden?

Diese Manuskript ist derzeit in anderer Form im Review und eine Open Access Publikation ist angestrebt.