

Pädagogisch-psychologisches Wissen von Lehrkräften
der Erwachsenen- und Weiterbildung:
Konzeptualisierung, Entwicklung und Validierung
eines fach- und bildungsbereichsübergreifend einsetzbaren Tests

Dissertation
zur Erlangung des Doktorgrades
der wirtschafts- und sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Christian Marx
aus Laupheim

Tübingen

März 2018

Tag der mündlichen Prüfung:

08.10.2018

Dekan:

Professor Dr. rer. soc. Josef Schmid

1. Gutachter:

Professor Dr. phil. Josef Schrader

2. Gutachter:

Professor Dr. phil. Bernhard Schmidt-Hertha

Danksagung

Diese Dissertationsschrift ist anschließend an das Projekt Think (Using Digital Media to Assess Generic Aspects of Teachers' Professional Knowledge in Different Educational Contexts) entstanden, einem Projekt des WissenschaftsCampus Tübingen, das vom Deutschen Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e.V. in Bonn (DIE), dem Hector-Institut für Empirische Bildungsforschung in Tübingen (HIB), der Universität Freiburg und der Universität Mannheim bearbeitet wurde. Durch diese Rahmenbedingungen eröffnete sich mir die Möglichkeit, exzellente Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Disziplinen kennen zu lernen, von denen ich einigen namentlich danken möchte.

Mein erster Dank gilt Herrn Professor Dr. Josef Schrader, der mir die Möglichkeit gegeben hat, im Projekt Think und damit im WissenschaftsCampus zu arbeiten, einem sehr an- und aufregenden Arbeitskontext. Danken möchte ich Herrn Schrader zudem für die Möglichkeit, meine Dissertationsschrift an das Projekt Think anzuschließen, sowie seine hilfreichen Empfehlungen bei deren Erstellung. Beeindruckt hat mich, dass er trotz vielfacher Verpflichtungen qua Amt als wissenschaftlicher Direktor des DIE ein großes Interesse an der Diskussion von Befunden aus dem Projekt Think und meiner Dissertationsschrift sowie an deren erfolgreichen Abschluss hatte. Herrn Professor Dr. Bernhard Schmidt-Hertha danke ich für hilfreiche Gespräche, sein Interesse an meiner Arbeit und für seine Bereitschaft, meine Dissertationsschrift als Gutachter zu unterstützen.

Mein besonders herzlicher Dank gilt Frau Dr. Annika Goeze, Leitung der DIE-Nachwuchsgruppe „Professionelle Kompetenzen des Weiterbildungspersonals“. Die Zusammenarbeit mit ihr in der Nachwuchsgruppe hat mich als Wissenschaftler gewiss nachhaltig positiv geprägt und eine für mich passendere Betreuung hätte ich mir nicht vorstellen können. Ihr möchte ich danken für ihr außergewöhnliches und unermüdliches Engagement bei der Betreuung meiner Arbeit im Projekt Think, der gemeinsamen Vorbereitung zahlreicher Konferenzauftritte und bei der Erstellung der vorliegenden Dissertationsschrift. Durch ihre ständige Bereitschaft, meine Manuskripte zu diskutieren, ihre wertvollen Ratschläge, ihr Vertrauen in mich und meine Arbeit hat sie einen großen Anteil am erfolgreichen Abschluss dieser Dissertationschrift.

Danken möchte ich auch allen Kolleginnen und Kollegen im Projekt Think, am DIE in Bonn und an der Universität Tübingen, die mich alle auf ihre Weise und mit ihrer Expertise unterstützt haben, insbesondere: Professor Dr. Matthias Alke, Eva Bonn, Professorin Dr. Viola Deutscher (geb. Klotz), Dr. Thomas Hallmayer, Dr. Dörthe Herbrechter, Dr. Petra Hetfleisch, Professor Dr. Augustin Kelava, Bianca Pollety, Jana Richter, Moritz Sahlender, PD Dr. Hannes Schröter, Dorett Schneider, Dr. Sabine Schöb und Professorin Dr. Tamar Voss.

Für die erholsame Ruhe außerhalb des Wissenschaftssystems und alle erdenkliche Unterstützung danke ich meiner Familie und meinen Verwandten, insbesondere: Meinen Eltern Angela und Otto, meinen Schwestern Andrea und Birgit, meiner Tante Erika Rommel, meinem Onkel Dr. Thomas Christ mit Familie, Familie Vogt und Familie Rothe. Meiner Partnerin Juliane Rothe danke ich ganz besonders herzlich, vor allem für die Rück- und Nachsicht, dass Projektarbeit und Dissertationsschrift regelmäßig gemeinsame Abende, Wochenenden und Urlaube durchkreuzen durften.

Für das friedvolle und bunte Zusammenleben über mehrere Jahre danke ich meiner Wohngemeinschaft in Bonn, Jens Bauer, Anja, Bernhard, Levke und Jonte Budnik, Maren und Yanay Kraushaar, Professor Dr. Ludwig Morenz.

Für die gemeinsamen Unternehmungen und für alle kritischen Fragen zu meiner Arbeit und zu erziehungswissenschaftlicher Forschung danke ich allen Freunden, insbesondere: Michael Brenz, Daniel Gründler mit Familie, Manuel Holstein mit Familie, Thilo Klaiber, Thomas Kohler mit Familie, Denis Kuhlmann, Michael Schmidt, Andreas Schnell, Joachim Schork, Philipp Schühle, Familie Seemüller, Dr. Sven Rieger, Konrad Willar und Familie Winter.

Zusammenfassung

Die vorliegende Dissertationsschrift beschäftigt sich mit der Konzeptualisierung des pädagogisch-psychologischen Wissens (PPK) von Lehrkräften und daran anschließend mit der Entwicklung und Validierung eines fach- und bildungsbereichsübergreifend einsetzbaren Tests. Für das PPK wird auf der Grundlage theoretisch-konzeptueller Überlegungen aus allen Bildungsbereichen und empirischen Befunden aus dem Schulkontext davon ausgegangen, dass es vermittelt über die Qualität von Lehr-Lernsituationen für den Lernerfolg der Teilnehmenden von Bedeutung ist. Die bislang entwickelten Tests zur Erfassung des PPK wurden für Schul-Lehrkräfte konzipiert, die Erfassung des PPK bei Lehrkräften aus anderen Bildungsbereichen stellt ein Desiderat dar. Der Fokus dieser Arbeit liegt auf der Erfassung dieses Wissens bei Lehrkräften der Erwachsenen- und Weiterbildung (EB/WB). Ein Test zur Erfassung des PPK kann in der pädagogischen Praxis der EB/WB z.B. im Rahmen von Aus-, Fort- und Weiterbildung der Lehrkräfte oder bei Rekrutierungsentscheidungen hilfreich sein sowie in einer empirischen Lehr-Lernforschung für die EB/WB Anwendung finden, für die es bislang nur wenige der nötigen Tests und Fragebögen gibt. Zudem ist ein solcher Test zur Erforschung der Frage nach der Allgemeinheit des PPK nötig. Diese Frage ist aus theoretisch-konzeptueller sowie für die Ausbildung von Lehrkräften auch aus praktisch-curricularer Sicht relevant. Darüber hinaus ist diese Frage aus forschungsstrategischer Sicht für die Entwicklung eines Tests zur Erfassung des PPK von Lehrkräften der EB/WB von Bedeutung. Die Frage nach der fach- und bildungsbereichsübergreifenden Allgemeinheit des PPK wurde immer wieder (implizit) gestellt, bislang wurde aber nicht begonnen, diese bildungsbereichsübergreifend empirisch zu überprüfen.

Die Dissertationsschrift verfolgt das Ziel, durch die Entwicklung eines Tests, der darauf ausgelegt ist, fach- und bildungsbereichsunabhängig das potentiell allgemeine PPK von Lehrkräften aus verschiedenen Bildungsbereichen erfassbar zu machen, Hinweise auf die Frage nach der tatsächlichen Allgemeinheit des PPK zu erlangen. Was die Frage nach der Allgemeinheit des PPK genau meint, wird im Manteltext aufgegriffen. Die Erforschung der Allgemeinheit des PPK ist ein längerfristig angelegtes Vorhaben. Erste Schritte wurden im Projekt ThinkK (*Using Digital Media to Assess Generic Aspects of Teachers' Professional Knowledge in Different Educational Contexts*) umgesetzt, einem Projekt-Cluster des WissenschaftsCampus Tübingen, das vom Deutschen Institut für Erwachsenenbildung (DIE, Josef Schrader (Leitung), Annika Goeze, Christian Marx), dem Hector-Institut für Empirische Bildungsforschung in Tübingen (HIB, Verena Gindele, geb. Hoehne), der Universität Freiburg (Thamar Voss, vormals HIB) und der Universität Mannheim (Viola Deutscher, geb. Klotz, vormals DIE) von April 2013 bis Dezember 2017 bearbeitet wurde. An dieses Projekt schließt

die vorliegende Dissertationsschrift an und in den drei Studien der Dissertationsschrift wird Datenmaterial aus dem Projekt Think verwendet.

Ziel eines Tests ist es, das interessierende Konstrukt so valide wie möglich zu erfassen. Im Manteltext der vorliegenden Arbeit stehen Validitätsüberlegungen bei der Operationalisierung des PPK sowie Möglichkeiten der Validitätsüberprüfung im Fokus, da die Darstellung dieser in den in ihrem Umfang begrenzten Studien der Dissertationsschrift nicht geleistet werden kann. Dabei sind sowohl Fragen nach dem Einfluss des Aufgabenformats und der Aufgabeninhalte auf die Validität sowie Möglichkeiten zur Überprüfung der faktoriellen und nomologischen Validität der Aufgaben Gegenstand der Darstellung.

Studie 1 stellt die Frage nach der potentiellen bzw. tatsächlichen Allgemeinheit einer Konzeptualisierung des PPK und bearbeitet diese durch eine systematische Literaturrecherche mit anschließender Inhaltsanalyse relevanter Literaturquellen. Die entwickelte Konzeptualisierung wird von Praxis- und Wissenschafts-Experten aus Schule und EB/WB als bildungsbereichsübergreifend eingeschätzt. Zwei Facetten des PPK, das Wissen über Lehr- und Lernziele sowie das Wissen über Individual- und Lernprozessdiagnostik, wurden operationalisiert und an Lehrkräften aus Schule und EB/WB erprobt. Dabei zeigte sich für beide Lehrkräftestichproben eine zweidimensionale Struktur, wobei die Erfassung des Wissens nur für die Schul-Lehrkräfte mit zufriedenstellender Reliabilität gelang. Zudem zeigten sich bei einigen Aufgaben differenzielle Itemfunktionen.

Studie 2 greift eine für die EB/WB besonders relevante Facette des PPK auf und stellt die Frage nach der Dimensionalität dieser Facette, dem Wissen über Lehr-Lernmethoden und -konzepte. Diese Frage stellt sich aus theoretisch-konzeptueller Perspektive für jede der insgesamt acht konzeptualisierten Facetten, wird jedoch an dieser für die EB/WB besonders zentralen Facette dahingehend exemplifiziert, ob und inwiefern aus theoretisch-konzeptueller und empirischer Perspektive eine Ein- oder Mehrdimensionalität plausibel erscheint. Die empirischen Befunde weisen konform zu den Annahmen auf einen Generalfaktor und Spezifitäten und damit auf Mehrdimensionalität für diese Facette hin.

Studie 3 stellt die Frage, ob und inwiefern die Vorbildung und die Erfahrungen von Lehrkräften aus der EB/WB in einem Zusammenhang stehen mit dem Wissen über Lehr-Lernmethoden und -konzepte. Hierbei zeigte sich u.a. erwartungskonform ein positiver Zusammenhang zwischen dem Umfang an einschlägiger Fort- und Weiterbildung mit dem Wissen über Lehr-Lernmethoden und -konzepte, von dem insbesondere Lehr-Novizen besonders zu profitieren scheinen. Die Bedeutung dieses Befunds für die Aus-, Fort- und Weiterbildung von Lehrkräften in der EB/WB wird diskutiert, ebenso der erwartete Unterschied, dass Lehrkräfte in der beruflich-betrieblichen WB verglichen mit Lehrkräften in der allgemeinen EB über mehr Wissen in dieser zentralen Facette verfügen.

Die Bedeutung der Befunde insgesamt für die Validierung des Tests sowie in ihrer Bedeutung für die Frage nach der potentiellen bzw. tatsächlichen Allgemeinheit des PPK ist Gegenstand der Gesamtdiskussion der Dissertationsschrift. Abschließend werden Möglichkeiten der Testweiterentwicklung und der weiteren Erforschung der Frage nach der Allgemeinheit des PPK vorgestellt.

Inhaltsverzeichnis

1 Gründe für die Entwicklung eines Tests zur Erfassung des pädagogisch-psychologischen Wissens von Lehrkräften in der Erwachsenen- und Weiterbildung...	1
1.1 Die Erforschung von Lehr-Lernsituationen der Erwachsenen- und Weiterbildung	1
1.2 Die Erforschung der Annahme der fach- und bildungsbereichsübergreifenden Allgemeinheit des pädagogisch-psychologischen Wissens	3
2 Bestehende Konzeptualisierungen des pädagogisch-psychologischen Wissens..	5
3 Testentwicklung	8
3.1 Empirische Grundlage der Aufgaben der acht Wissens-Facetten.....	10
3.2 Validitätsüberlegungen bei der Operationalisierung des (allgemeinen) pädagogisch-psychologischen Wissens	13
3.2.1 Validität und Aufgabeninhalte	16
3.2.2 Validität und Aufgabenformate.....	20
4 Testvalidierung.....	24
4.1 Faktorielle Validierung.....	24
4.1.1 Reflektive und formative Messmodelle.....	25
4.1.2 Modelle der Item Response Theory	26
4.1.3 Modellgeltungstests	29
4.1.4 Überprüfung auf differentielle Itemfunktionen.....	31
4.2 Nomologische Validierung.....	33
5 Fragestellungen der empirischen Studien	36
6 Studie 1 Pädagogisch-psychologisches Wissen von Lehrkräften aus Schule und Erwachsenenbildung: Entwicklung und Erprobung eines Testinstruments.....	38
7 Studie 2 Das Wissen über Lehr-Lernmethoden und -konzepte von Lehrkräften der Erwachsenen- und Weiterbildung – Wie lässt sich die Dimensionalität valide abbilden?	82
8 Studie 3 Lehrkräfte in der Erwachsenen- und Weiterbildung: Zusammenhänge zwischen Vorbildung und Erfahrung mit dem Wissen über Lehr-Lernmethoden und -konzepte	121

9 Gesamtdiskussion – Fragestellungen, Befunde und Perspektiven.....	148
9.1 Zusammenfassung zentraler Ergebnisse	149
9.2 Kritische Betrachtung der berichteten Befunde.....	152
9.3 Verbleibende Desiderate: Potentiale für und Ausblick auf anschlussfähige Forschungsmöglichkeiten.....	153
9.3.1 Aufgabenweiterentwicklung	153
9.3.2 Fortsetzung der Testvalidierung.....	157
9.3.3 Weitere Bearbeitung der Frage der Allgemeinheit des pädagogisch-psychologischen Wissens	158
10 Literaturverzeichnis	160

1 Gründe für die Entwicklung eines Tests zur Erfassung des pädagogisch-psychologischen Wissens von Lehrkräften in der Erwachsenen- und Weiterbildung

Die vorliegende Arbeit berichtet über die Konzeptualisierung des pädagogisch-psychologischen Wissens (PPK) und daran anschließend über die Entwicklung und Validierung eines fach- und bildungsbereichsübergreifend einsetzbaren Tests zur Erfassung dieses Wissens bei Lehrkräften, in dieser Arbeit mit dem Schwerpunkt auf Lehrkräften in der Erwachsenen- und Weiterbildung (EB/WB). Die Entwicklung eines solchen Tests ist sinnvoll, weil es in der pädagogischen Praxis der EB/WB derzeit keine validen Tests gibt, die z.B. zur Ermittlung von Fort- und Weiterbildungsbedarfen (Schöb, Rohs, Biel & Scholze, 2016; Schrader, Hohmann & Hartz, 2010) oder bei Rekrutierungsentscheidungen (Gieseke, 2018; Goeze & Schneider, 2014) über das Ausmaß an vorhandenem Wissen bei Lehrkräften aufklären könnten. Auch die wissenschaftliche Erforschung von gelingendem Lehren und Lernen in der EB/WB dürfte von einem geprüften Test profitieren. Darüber hinaus ist ein solcher Test notwendig zur Erforschung der Annahme der fach- und bildungsbereichsübergreifenden Allgemeinheit des PPK im Rahmen der Frage nach einer fach- und bildungsbereichsübergreifenden empirischen Lehr-Lernforschung und der tatsächlichen Allgemeinheit der Allgemeinen Didaktik (Terhart, 2002, 2013). Diese Anwendungsmöglichkeiten werden nachfolgend ausgeführt.

1.1 Die Erforschung von Lehr-Lernsituationen der Erwachsenen- und Weiterbildung

Die EB/WB als vierter Bereich des Bildungssystems hat nicht erst seit den Migrationsbewegungen der vergangenen Jahre für Individuen zunehmend an Wichtigkeit gewonnen, um sich privat und/oder beruflich fort- und weiterbilden zu können, sondern auch aus der Perspektive des nationalen Gesamtgesellschaftssystems, u.a. als Möglichkeit, die internationale Wettbewerbsfähigkeit zu erhalten. Aufgrund der zunehmenden Wichtigkeit dieses Bildungsbereichs wird die Frage nach der Effektivität der EB/WB seit einigen Jahren von der Politik zunehmend gestellt (Bundesministerium für Arbeit und Soziales, 2014; Europäische Kommission, 2001, 2007). Vor diesem Hintergrund gewinnt die pädagogische Praxis der EB/WB und die Erforschung der Interdependenzen zwischen 1) Ausbildung der in der EB/WB tätigen Lehrkräfte, 2) deren professioneller Kompetenz 3) dem Lernerfolg der Teilnehmer¹ und schließlich auch 4) der erfolgreichen Teilhabe der Lerner am (gesellschaftlichen und) beruflichen Leben zusätzlich an Wichtigkeit.²

¹ Um den Lesefluss zu erleichtern, wird im Folgenden nur die männliche Geschlechtsform verwendet, es sind aber beide Geschlechter gemeint.

² Offen ist allerdings, welche Bedeutung die EB/WB für das (Nicht-)Vorhandensein von (basalen) Kompetenzen hat.

Die Erforschung der genannten Wirkungskette (vgl. Terhart, 2012) ist Gegenstand der empirischen Lehr-Lernforschung. Die Frage nach Effekten im Rahmen der auf die EB/WB bezogenen empirischen Lehr-Lernforschung spielte lange Zeit eine eher untergeordnete Rolle (siehe Boeren, 2018; Daley, Martin & Roessger, 2018; Fuhr et al., 2015), was sich auch im Forschungsstand entsprechend widerspiegelt (siehe aber Goeze, 2016; Goeze, Hettfleisch & Schrader, 2013; Goeze, Zottmann, Vogel, Fischer & Schrader, 2014). Fragen nach den Ausprägungen der Grundkompetenzen der erwachsenen Bevölkerung sowie deren Zusammenhang z.B. mit dem beruflichen Erfolg sind mittlerweile durch Large Scale Assessments (LSA) wie dem „*Programme for the International Assessment of Adult Competencies*“ (PIAAC)³, seinen Vorgängerstudien „*Adult Literacy and Lifeskills Survey*“ (ALL) und „*International Adult Literacy Survey*“ (IALS) oder dem „*Nationalen Bildungspanel*“ (NEPS) vergleichsweise gut beforscht, wofür ein Grund das Vorliegen entsprechender Tests und Fragebögen in Kombination mit den nötigen Forschungsdesigns ist.

Ähnlich wie für den Schulbereich erreicht, z.B. durch die Projekte „*Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern*“ (COACTIV, Kunter, Baumert et al., 2011) oder „*Teacher Education and Development Study in Mathematics*“ (Blömeke, Kaiser & Lehmann, 2010a, 2010b), beginnt sich für die EB/WB eine die Lehrkräfteforschung einbeziehende empirische Lehr-Lernforschung erst zu entwickeln (siehe auch Goeze, 2016). Denn die Forschung (zur Professionalisierung) in der EB/WB konzentrierte sich lange Zeit auf das planend-disponierend tätige Personal (Schrader, 2010a), hingegen wurden Lehr-Lernsituationen (z.B. Herrle, 2013; Kade, Nolda, Dinkelaker & Herrle, 2014) und die Lehrkräfte in der EB/WB erst in den vergangenen Jahren verstärkt Gegenstand der Forschung (z.B. Lattke & Jütte, 2014).

Die Lehrkräfteforschung in der EB/WB kann grob in zwei Bereiche unterteilt werden: 1) Studien, die (u.a.) die berufliche und soziale Lage der Lehrkräfte erforschen (Autorengruppe wb-personalmonitor, 2016; Dobischat, Elias & Rosendahl, 2018; WSF Wirtschafts- und Sozialforschung, 2005) und 2), Studien zur professionellen Kompetenz von Lehrkräften. Zweitgenannte lassen sich untergliedern in Studien, die sich (eher) auf die Lehr-Lernsituationen vor- oder nachgelagerten Handlungen und den damit verbundenen Kompetenzaspekten beziehen (z.B. Haberzeth, 2010; Pachner, 2013; Stanik, 2016) oder (eher) auf Kompetenzaspekte, die die konkrete Lehr-Lernsituation betreffen (Arnold, Krämer-Stürzl & Siebert, 2005; Goeze, 2016; Kraft, Seitter & Kollwe, 2009; Rohs, Rott, Schmidt-Hertha & Bolten, 2017; Schrader et al., 2010; Siebert, 2012).

³ Im Unterschied zum „*Programme for International Student Assessment*“ (PISA) lässt sich PIAAC nicht als Evaluation der EB/WB interpretieren.

Dem angesprochenen Mangel an Tests und Fragebögen, v.a. zur Erfassung der professionellen Kompetenz von Lehrkräften (siehe Marx, Goeze & Schrader, 2014a), wird aktuell durch verschiedene Projekte begegnet: Die Arbeitsgruppe von Bonnes und Hochholding (Universität Konstanz) und eine ab April 2018 am DIE (Deutsches Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e.V., Bonn) forschende Nachwuchsgruppe von Herbrechter entwickeln jeweils Instrumente zur Erfassung von Überzeugungen von Lehrkräften aus der EB/WB. Im Projekt *„Medienpädagogische Kompetenz des beruflichen Weiterbildungspersonals zur Unterstützung des Einsatzes digitaler Medien in formalen, non-formalen und informellen Lernsettings“ (MEKWEP)* wird ein Test zur Erfassung der medienpädagogischen Kompetenz entwickelt (Rohs et al., 2017), im Projekt *„Grundlagen zur Entwicklung eines trägerübergreifenden Anerkennungsverfahrens für die Kompetenzen Lehrender in der Erwachsenen-/Weiterbildung“ (GRETA)* ein Portfolio für Lehrkräfte (Bosche & Strauch, angenommen; Lencer & Strauch, 2016) und im Projekt *“Using Digital Media to Assess Generic Aspects of Teachers’ Professional Knowledge in Different Educational Contexts“ (Think)* ein Test zur Erfassung des PPK.

1.2 Die Erforschung der Annahme der fach- und bildungsbereichsübergreifenden Allgemeinheit des pädagogisch-psychologischen Wissens

Die vorliegende Arbeit schließt an das Projekt Think an, einem Projekt-Cluster des WissenschaftsCampus Tübingen, das vom Deutschen Institut für Erwachsenenbildung (DIE, Josef Schrader (Leitung), Annika Goeze, Christian Marx), dem Hector-Institut für Empirische Bildungsforschung in Tübingen (HIB, Verena Gindele, geb. Hoehne), der Universität Freiburg (Thamar Voss, vormals HIB) und der Universität Mannheim (Viola Deutscher, geb. Klotz, vormals DIE) von April 2013 bis Dezember 2017 bearbeitet wurde (für eine Zusammenfassung siehe Voss et al., 2017). Das Projekt Think schließt an die Konzeptualisierung und Entwicklung eines Tests zur Erfassung des PPK von Lehrkräften aus dem Schulbereich im Rahmen des Projekts COACTIV-R an (Voss, Kunter & Baumert, 2011). Ziel des Projekts Think war es, die Grundlagen zu legen für die Untersuchung der bislang nicht empirisch überprüften, aber seit Comenius (siehe z.B. in der von Flintner herausgegeben Arbeit, Comenius, 2007) angenommenen und immer wieder aufgegriffenen Frage nach einer „Universaldidaktik“ (Aebli, 2011; Terhart, 2013), der die Frage nach der fach- und bildungsbereichsübergreifenden Allgemeinheit des PPK zuzuordnen ist. Die Entwicklung eines Tests zur prinzipiellen Erfassbarkeit dieses Wissens als potentiell allgemeines Wissen stellt einen notwendigen Schritt zur Beantwortung dieser Frage dar; und auch – dies mag an dieser Stelle zunächst wie ein Zirkelschluss erscheinen – ist die Allgemeinheit des PPK Grundvoraussetzung, um bei dem derzeitigen Forschungsstand der

empirischen Lehr-Lernforschung in der EB/WB einen „gültigen“ Test für die Erfassung des PPK bei Lehrkräften in der EB/WB anbieten zu können.

Denn Grundlage von Wissenstests sind zumeist empirische Befunde zu einem Gegenstandsbereich (Almond, Mislevy, Steinberg, Yan & Williamson, 2015) – hier dem Lehren und Lernen in verschiedenen Bildungsbereichen. Für die EB/WB gibt es solche Befunde bislang nur in einem sehr geringen Ausmaß (vgl. Kuper & Schrader, in Druck, 2013). Aufgrund des Mangels an entsprechenden Befunden ist die Frage nach der fach- und bildungsbereichsübergreifenden Allgemeinheit des PPK nicht nur aus einer theoretisch-konzeptuellen Perspektive, sondern für die EB/WB auch aus forschungsstrategischer Sicht von großer Wichtigkeit. Denn finden sich Hinweise auf die angenommene Allgemeinheit des PPK, dann können ggf. Befunde aus dem Schulkontext, für den es eine große Zahl an empirischen Befunden zum Lehren und Lernen gibt, auf die EB/WB übertragen werden und z.B. im Rahmen eines Tests zur Erfassung des PPK verwendet werden. Bzw.: wird ein Test zur Erfassung des PPK von Lehrkräften verschiedener Bildungsbereiche auf der Grundlage von Befunden aus dem Schulkontext entwickelt, so ist die Überprüfung auf Allgemeinheit des PPK ein nötiger Validierungsschritt, um die „Gültigkeit“ des Tests auch jenseits seines „empirischen Ursprungs“ zu gewährleisten.

Die vorliegende Dissertationsschrift legt erste Befunde für dieses längerfristig angelegte Forschungsprojekt zur Entwicklung eines Tests zur Erfassung des PPK und zur Erforschung der Allgemeinheit des PPK vor, wobei in dieser Arbeit der Fokus auf den Lehrkräften in der EB/WB liegt. Die Fragen, die in dieser Arbeit bearbeitet werden, sind ausgewählte und grundlegende Fragen, die notwendig positiv beantwortet werden müssen, bevor das im Projekt ThinK verfolgte Forschungsziel erreicht werden kann. Im folgenden Kapitel wird auf die bislang entwickelten Konzeptualisierungen des PPK eingegangen und verdeutlicht, was mit der Frage nach der Allgemeinheit des PPK genau gemeint ist.

2 Bestehende Konzeptualisierungen des pädagogisch-psychologischen Wissens

Das PPK (Lenske, Thillmann, Wirth, Dicke & Leutner, 2015; Voss et al., 2011), auch pädagogisches Wissen (Großschedl, Harms, Kleickmann & Glowinski, 2015; Voss, Kunina-Habenicht, Hoehne & Kunter, 2015) oder allgemeines pädagogisches Wissen (general pedagogical knowledge, GPK, König & Blömeke, 2012), ist definiert als Wissen, das nötig ist zur Gestaltung und Optimierung von Lehr-Lernsituationen, anwendbar ist in verschiedenen Fächern und Bildungsbereichen und das deklaratives und prozedurales Wissen⁴ beinhaltet (Voss et al., 2011). Für eine Übersicht (nicht nur) zu Definitionen des PPK, siehe Voss et al. (2015).

Das PPK ist Gegenstand verschiedener Disziplinen, z.B. der Allgemeinen Didaktik als Theorie und Praxis des Lehrens und Lernens für alle Bildungsbereiche (z.B. Kron, 2014) und der empirischen Lehr-Lernforschung (siehe z.B. Terhart, 2002). PPK ist neben dem PCK (pedagogical content knowledge, fachdidaktisches Wissen) und dem Fachwissen (content knowledge, CK) ein Bereich des professionellen Wissens von Lehrkräften, von dem auf theoretisch-konzeptueller Perspektive einer allgemeinen Didaktik (Aebli, 2011; Arnold, Blömeke, Messner & Schlörkemper, 2009; Kron, 2014) und Grundlage der empirischen Lehr-Lernforschung (Lenske et al., 2016; Voss, Kunter, Seiz, Hoehne & Baumert, 2014) angenommen wird, dass es – vermittelt über die Handlungen der Lehrkraft (z.B. Goeze, 2016) – die Unterrichtsqualität und den Lernerfolg der Lernenden beeinflusst. PPK wird zumeist implizit oder explizit als fach- und bildungsbereichsübergreifendes, also allgemeines Wissen definiert (vgl. Studie 1; Voss et al., 2015).⁵

Um zu verdeutlichen, was die im Projekt Think gestellte Frage nach der Allgemeinheit des PPK meint, bietet sich der Vergleich zwischen dem PPK und dem PCK an, die beide als Bereiche des professionellen Wissens von Lehrkräften angesehen werden; das professionelle Wissen wiederum ist ein Aspekt der professionellen Kompetenz, neben den Überzeugungen, den motivationalen Orientierungen sowie der Selbstregulation der Lehrkräfte (Baumert & Kunter, 2011), vergleiche Abb. 1.

⁴ Wissen wird im Rahmen der vorliegenden Arbeit aus kognitionspsychologischer Sicht definiert. Im Beitrag von Marx, Goeze und Schrader (2014b), der nicht Gegenstand dieser Dissertationsschrift ist, wird eine Verbindung der kognitionspsychologischen Definition von Wissen mit Definitionen von Wissen aus der Erwachsenenbildungsforschung hergestellt, z.B. zu Dewe (2010) und Hof (2001). Hintergrund der Verwendung der kognitionspsychologischen Definition von Wissen ist, dass Kerngegenstand der Kognitionspsychologie kognitive Prozesse sind, also z.B. Prozesse der Wissensspeicherung, die von dieser Disziplin seit vielen Jahren empirisch untersucht werden (zusammenfassend z.B. Rösler, 2011). Hingegen sind kognitive Prozesse nicht Gegenstand *empirischer* Forschung der EB/WB. Zur Expertiseforschung insgesamt und zur Art der Wissensspeicherung und -abrufbarkeit von Lehrkräften in der EB/WB siehe Goeze (2016).

⁵ Auf die Erforschung des PPK in verschiedenen Disziplinen (Allgemeine Didaktik, empirische Lehr-Lernforschung und EB/WB) wird in Studie 1 näher eingegangen, ebenso auf Befunde zur Wirksamkeit des PPK auf die Unterrichtsqualität und den Lernerfolg der Lernenden.

PCK ist fachspezifisches Wissen zur Gestaltung von Lehr-Lernsituationen und beinhaltet u.a. Wissen über Erklärungsmöglichkeiten fachspezifischer Sachverhalte, Wissen über typische Fehlerquellen bei Verständnisprozessen dieser Sachverhalte und damit verbunden die impliziten Wissensvoraussetzungen und kognitiven Anforderungen von Aufgaben (Baumert & Kunter, 2011). Z.B. ist das Wissen (einer Lehrkraft) über typische Fehler von Lernenden beim Lösen linearer Funktionen und deren Vermeidung für das Fach Mathematik „gültig“ – kann also zum Lernerfolg der Lernenden beitragen – nicht aber z.B. für das Fach Deutsch. Für das PPK wird verglichen mit dem PCK angenommen, dass es für verschiedene Fächer und Bildungsbereiche „gilt“ und insofern allgemeines Wissen ist (siehe z.B. Voss et al., 2015); ein Beispiel: das Wissen darüber, welche Art von Aufgaben bei Gruppenarbeiten die Lernmotivation erhöhen und dadurch schließlich den Lernerfolg wahrscheinlicher machen, kann für verschiedene Fächer und Bildungsbereiche „gültig“ sein.



Abb. 1 Das Modell der professionellen Kompetenz von Lehrkräften nach Baumert und Kunter (2011)

Die Annahme der Allgemeinheit des PPK ist eine Schlussfolgerung aus der Annahme, dass sich die Grundbedingungen des (Lehrens und) Lernens nicht zwischen den Lernenden verschiedener Bildungsbereiche unterscheiden (vgl. Tietgens, 1967); „Die Lernprozesse sind aber auf allen Stufen die gleichen. Wir kennen bis heute keine qualitativen Unterschiede zwischen dem Lernen des Kindes und demjenigen des Erwachsenen“ (Aebli, 2011, S. 16). Die Möglichkeit der Existenz einer *allgemeinen* Psychologie, die sich mit den psychologischen Funktionen beschäftigt, die allen Menschen gemein sind (z.B. Müsseler & Rieger, 2017), geht einher mit der Möglichkeit der Existenz einer *allgemeinen* (fach- und bildungsbereichsübergreifenden) Didaktik. Denn es finden sich hinsichtlich allgemeinpsychologischer Konstrukte der Lernenden wie z.B. der Motivation, Emotion und Kognition für Lernende aller Bildungsbereiche für Lehr-Lernprozesse günstige als auch ungünstige

Ausprägungen, die Ausgangspunkt des Handelns einer Lehrkraft sind, aus denen wiederum didaktische Gemeinsamkeiten verschiedener Bildungsbereiche resultieren können (Marx, Goeze & Schrader, 2014b). Die Annahme der Allgemeinheit des PPK bzw. einer Allgemeinen Didaktik schließt allerdings die Möglichkeit bildungsbereichsspezifischen bzw. altersgruppenspezifischen Wissens zur Gestaltung von Lehr-Lernsituationen nicht aus (Siebert, 2012), die durch entwicklungspsychologische „Besonderheiten“ der Lernenden begründet sein können, wie etwa (durchschnittliche) Unterschiede in der Konzentrationsfähigkeit von Individuen z.B. in den ersten Jahren des Primarbereichs oder am Ende der Lebensspanne verglichen mit Individuen im jungen bis mittleren Erwachsenenalter (zusammenfassend z.B. Berk, 2011).

Die Frage der fach- und bildungsbereichsübergreifenden Allgemeinheit des PPK wurde gestellt (z.B. Terhart, 2013) oder angenommen (Aebli, 2011), bislang allerdings nicht empirisch überprüft (vgl. Studie 1).⁶ Sie ist aus theoretisch-konzeptueller (zurückführend z.B. bis zu Comenius und der Kunst, alle alles zu lehren) und praktisch-curricularer (Ausbildung von Lehrkräften verschiedener Fach- und Bildungsbereiche) Perspektive als empirische Legitimationsgrundlage einer Allgemeinen Didaktik interessant (vgl. Studie 1). Die Aussicht – bei positiver Beantwortung der Frage nach der Allgemeinheit des PPK – auf eine etwaige „gültige“ Übertragbarkeit von Befunden oder Hypothesen von einem auf einen anderen Bildungsbereich, z.B. vom Schulbereich auf die EB/WB, begründet aus forschungsstrategischer Sicht die fundamentale Wichtigkeit der Frage nach der Allgemeinheit des PPK für die EB/WB.

⁶ Für weitere bildungsbereichsübergreifende Fragestellungen siehe z.B. Nittel, Schütz, Fuchs und Tippelt (2011) oder Nittel, Schütz und Tippelt (2014).

3 Testentwicklung

In diesem Kapitel werden die empirische Grundlage der Aufgaben (Kapitel 3.1) und Validitätsüberlegungen bei der Aufgabenentwicklung (Kapitel 3.2) detaillierter dargestellt, als dies in den hinsichtlich ihres Umfangs begrenzten drei Studien der Dissertationsschrift geleistet werden kann.

Ein erster Schritt zur Beantwortung der Frage nach der Allgemeinheit des PPK ist die Frage, ob das PPK fach- und bildungsbereichsübergreifend konzeptualisierbar ist; davon darf dann ausgegangen werden, wenn in existierender Literatur und in aktuellen Expertenurteilen für verschiedene Bildungsbereiche die gleichen Wissensfacetten als „gültig“ bzw. als Bestandteil des PPK angesehen werden. In Studie 1 wird eine fach- und bildungsbereichsübergreifende Konzeptualisierung des PPK vorgestellt, die auf einer systematischen Literaturrecherche und Inhaltsanalyse von Quellen aus der bzw. für die EB/WB und den Schulbereich aufbaut. Diese Konzeptualisierung wurde im Rahmen einer Expertenbefragung als bildungsbereichsübergreifend relevant eingeschätzt (Studie 1; Marx et al., 2014b), was die Entwicklung und Validierung eines Tests „legitimiert“ bzw. die Arbeitshypothese, dass das hier konzeptualisierte PPK allgemeines Wissen ist, zumindest nicht falsifiziert und daher weiterverfolgenswert erscheinen lassen.

Die in Studie 1 entwickelte Konzeptualisierung konstituiert sich wie viele der bildungsbereichsspezifischen Konzeptualisierungen des PPK durch mehrere inhaltliche Facetten⁷ und Subfacetten (König & Blömeke, 2009; Vinepac-Project, 2008; Voss et al., 2011). Die Facetten der Konzeptualisierung sind als Wissensdimensionen gedacht, auf denen individuelle Unterschiede abgebildet und folglich operationalisiert werden sollen. Ob die Facetten der Konzeptualisierung allerdings spezifisch genug definiert sind, um ein *operational definiertes Konstrukt* darzustellen, ist fraglich, denn mit einer operationalen Definition ist der folgende Anspruch verbunden: „With an operational definition, we have no surplus meaning or confusion about the construct. We can be very precise in the measurement of an operationally defined construct.“ (Haladyna & Rodriguez, 2013, S. 5). Die Konzeptualisierung des PPK (wenn man diese als operationale Definition verstehen will) sagt möglicherweise aber eher etwas über die (grobe) Strukturierung bzw. (Zu)Ordnung von Aufgaben des Tests aus, als über die Dimensionalität des Wissens, das bei der Beantwortung der Aufgaben einer Facette erforderlich ist. Die Frage, ob die Aufgaben einer Facette – wie durch die Konzeptualisierung angenommen – eindimensional sind, ist Gegenstand von Studie 2 und wird dort am Beispiel der Facette „Wissen über Lehr-Lernmethoden und -konzepte“ exemplifiziert.

⁷ PPK wird in verschiedenen Arbeiten auch nach Wissensarten untergliedert, z.B. König et al. (2014); Lenske, Thillmann, Wirth, Dicke und Leutner (2015).

Wie ein Konstrukt konzeptualisiert ist, hat eine Auswirkung darauf, wie bei der Operationalisierung und damit bei der konkreten Aufgaben- und Testentwicklung vorgegangen wird. An dieser Stelle sei erwähnt, dass der Einfluss der Konzeptualisierung und Definition eines Konstrukts (und damit verbunden die entsprechende Modellüberprüfung mit empirischen Befunden) sich auf lange Sicht möglicherweise auch auf Curricula und damit – in diesem Fall – die Ausbildung von Lehrkräften auswirken kann (vgl. für die Dimensionalitätsannahmen in LSA und etwaige Auswirkungen auf Curricula Borsboom & Wijsen, 2017); nicht nur deswegen ist angemessene Sorgfalt bei der Testentwicklung und -validierung und angemessene Zurückhaltung bei der Interpretation der Befunde geboten. Denn im Rahmen einer Wirkungsforschung ist weniger die Frage, *ob* PPK wirkt von Bedeutung, vielmehr die Frage *wie* und konkret *welches* PPK wirkt (vgl. Terhart, 2012).

Das methodische Vorgehen bei der *Konzeptualisierung* des PPK wird in Studie 1 ausführlich beschrieben. Studie 1 thematisiert allerdings nicht das Vorgehen bei der *Aufgabenentwicklung*. Unterschiede bestehen darin, dass sich die Konzeptualisierung auf einer theoretisch-konzeptuellen Ebene mit dem PPK von Lehrkräften beschäftigt und damit die Frage stellt, welche Wissensbereiche auf der Ebene von „Themen“ Bestandteil des PPK sind bzw. nicht sind. Die Konzeptualisierung konkretisiert aber nicht – und hierin liegt der Unterschied zur Aufgabenentwicklung – auf der Auflösungsebene von Fragen, wie z.B. „Für welche Lernziele hat sich der Lehrervortrag als wirksam erwiesen?“, sondern abstrahiert von solchen Fragestellungen und fasst diese zu insgesamt acht Wissensfacetten zusammen, siehe Abb. 2, die in insgesamt 30 Subfacetten untergliedert werden können (vgl. Studie 1).



Abb. 2 Konzeptualisierung des pädagogisch-psychologischen Wissens im Projekt ThinkK

Für die Aufgabenentwicklung genügt es nicht, „nur“ abstrahierte Facetten und damit Themen zu benennen, sondern es sind (bestenfalls) empirische Befunde mit einem höheren Auflösungsgrad nötig. Z.B. Befunde, die gelingende Wirkungsbeziehungen zwischen dem Handeln der Lehrkraft und dem Lernerfolg der Teilnehmer thematisieren. Diese Befunde sind für Lehrkräfte wiederum relevantes Wissen zur Gestaltung von Lehr-Lernsituationen. Hinweise auf die Allgemeinheit der Konzeptualisierung (vgl. Studie 1) sind noch keine Hinweise auf die Allgemeinheit der empirischen Befunde und damit die Allgemeinheit der konkreten Wirkungsbeziehungen. Das bedeutet, dass die Inhalte des Tests einer Plausibilitätsprüfung hinsichtlich ihrer Allgemeinheit unterzogen werden mussten.

Stellt man nun, wie im Rahmen dieser Arbeit bzw. dem Projekt Think, die Frage, über welches potentiell *allgemeine* PPK Lehrkräfte verfügen (sollten, um gelingende Lehr-Lernsituationen herstellen zu können), so ist man auf Befunde angewiesen, die z.B. den Lernerfolg oder die Lernmotivation von Teilnehmenden durch das didaktische und methodische Vorgehen der Lehrkräfte und damit durch Merkmale der Lehr-Lernsituation erklären oder von denen entsprechende Wirkungszusammenhänge abgeleitet werden können. Sodann stellt sich die Frage, welches empirisch fundierte Wissen es für bzw. zu Lehr-Lernsituationen in Schule und in der EB/WB gibt, das Gegenstand eines Tests werden kann. Für einen Test zur Erfassung des PPK relevante „Inhalte“ bietet die empirische Lehr-Lernforschung, die nach Terhart (2002) alle Forschungen umfasst, die sich auf die Mikroebene der Bildungsbereiche beziehen.

Nachfolgend wird eine Zusammenfassung des Forschungsstands der auf die EB/WB und den Schulbereich bezogenen empirischen Lehr-Lernforschung gegeben. Beschäftigt man sich mit Lehr-Lernsituationen, dann ist zunächst eine nähere Betrachtung des Lernens erforderlich, an dem sich die Gestaltung von Lehr-Lernprozessen ausrichtet (Terhart, 2009). Lernprozesse werden im Projekt Think in der als „Wissen über Lernprozesse von Lernenden“ bezeichneten PPK-Facette thematisiert und die Gestaltung von Lehr-Lernsituationen richtet sich zudem an weiteren Merkmalen der Lernenden aus, die in der Facette „Wissen über die Heterogenität der Lernenden“ zusammengefasst sind (vgl. auch Studie 1).

3.1 Empirische Grundlage der Aufgaben der acht Wissens-Facetten

Kognition, Motivation und Emotion, hier als Subfacetten der Facette 1) „Wissen über Lernprozesse von Lernenden“, sind Gegenstand der psychologischen Forschung sowohl in den Grundlagen- (z.B. in der allgemeinen Psychologie (z.B. Müsseler & Rieger, 2017)) als auch in den Anwendungsfächern (z.B. in der pädagogischen Psychologie (z.B. Seidel & Krapp, 2014)). In den Grundlagenfächern finden sich zumeist Forschungen zu allen Altersgruppen. Zudem bietet die EB/WB-Forschung eine große Zahl an Literatur, die auf

theoretisch-konzeptueller Ebene bzw. empirische Befunde rezipierend, zumeist aber nicht eigenständig empirisch forschend und damit für die Aufgabenentwicklung ungeeignet, sich mit dem Lernen Erwachsener aus einer konstruktivistischen Perspektive beschäftigt (Arnold, 2012; Siebert, 2012). Einen Überblick zur und eine Analyse der Rezeption des Konstruktivismus in der EB/WB bietet z.B. Berzbach (2005). Darüber hinaus finden sich auch in Fachzeitschriften der EB/WB-Forschung Studien zu Emotion (z.B. Jameson & Fusco, 2014; Tsai, Li & Cheng, 2017) und Motivation (z.B. Gorges, 2015; Gorges, Koch, Maehler & Offerhaus, 2017; Gorges, Maehler, Koch & Offerhaus, 2016; Rothes, Lemos & Gonçalves, 2017) von erwachsenen Lernenden in (Lehr-)Lernprozessen.

Die Facette 2) „Wissen über die Heterogenität der Lernenden“ besteht aus sieben Subfacetten. Zu a) *psychischen Auffälligkeiten*, b) *der Heterogenität hinsichtlich des Vorwissens und der Intelligenz*, d) *Unterschiede zwischen Geschlechtern*, f) *dem entwicklungspsychologischen Wissen* und g) *dem besonderen Förderbedarf* verfügt die oben genannte psychologische Forschung über Befunde. Zu Subfacette c) *verschiedene Lernstile und Lerntypen* finden sich ebenfalls für verschiedene Altersgruppen Befunde, z.B. bei Kolb (1985) oder bei Schrader (2008). Befunde finden sich auch für die Subfacette e) *kulturelle Besonderheiten, religiöse Orientierungen, Einstellungen von Lernenden sowie Migrationshintergrund und sozioökonomischer Status* sowohl für den Schulbereich (Maaz, Baumert, Gresch & McElvany, 2010) als auch für die EB/WB-Forschung, z.B. in den als klassische Leitstudien der deutschen Erwachsenenbildung bezeichneten Studien (vgl. z.B. Siebert, 2004), der Hildesheimer-, der Göttinger und der Oldenburg-Studie (Schulenberg, 1957; Schulenberg, Loeber, Loeber-Pautsch & Pühler, 1978; Strzelewicz, Raapke & Schulenberg, 1966). Gegenstand dieser Studien waren z.B. Zusammenhänge zwischen Schichtzugehörigkeit, Bildungsvorstellungen und Weiterbildungsbeteiligung (Strzelewicz et al., 1966) oder die Veränderungen des Bildungsverständnisses und der Weiterbildungsmotivation (Schulenberg et al., 1978), die Gegenstand von für die EB/WB spezifischen Aufgaben z.B. hinsichtlich der Motivlagen von Teilnehmenden sein können. Die bildungssoziologische Perspektive der Teilnehmer- und Weiterbildungsbeteiligungsforschung in der EB/WB blieb erhalten, z.B. in den Studien, die auf der Datengrundlage des „*Adult Education Survey*“ (AES) (Bilger & Strauß, 2015), des „*Sozioökonomisches Panel*“ (SOEP) oder des Mikrozensus durchgeführt wurden (vgl. Kuper, in Druck).

Zusammenfassend lässt sich für diese zwei Facetten des PPK eine große Zahl an Befunden für Personen aller Altersgruppen finden. Die genannten Studien verbinden allerdings nicht das vierte Glied (Lernerfolg von Lernenden bzw. deren Lernprozesse) mit dem dritten (Merkmale von qualitätsvollen Lehr-Lernprozessen) bzw. zweiten Glied (professionelle Kompetenz) der von Terhart (2012) beschriebenen Wirkungskette. Das Wissen über diese Wirkungsbeziehungen, die in den weiteren sechs Facetten des PPK

thematisiert werden, ist jedoch von großer Wichtigkeit für die Entwicklung eines Tests zur Erfassung des PPK: „Wenn man alles über Lernen weiß, weiß man noch nicht alles über die konkrete Erzeugung von erfolgreichem Unterricht“ (Terhart, 2009, S. 39).

Prozesse innerhalb von Lehr-Lernsituationen in der EB/WB wurden in den vergangenen Jahren häufig mittels des videographischen Zugangs Gegenstand der Erwachsenenbildungsforschung (z.B. Herrle, 2013; Kade et al., 2014), ebenso die Wirksamkeit von lehr-lernmethodischen Konzepten (z.B. Goeze, 2016; Goeze et al., 2013; Goeze et al., 2014; Hettfleisch, Goeze & Schrader, 2014). Diese Studien stellen allerdings nur einen geringen Teil der gesamten Erwachsenenbildungsforschung dar. Insgesamt ist das (allgemeine und fachbezogene) Lehren und Lernen in der EB/WB aus Mangel an Befunden nur eine Randerscheinung in Lehrbüchern, die sich qua Definition mit dem Erleben und Verhalten von Akteuren in Lehr-Lernsituationen *aller* Bildungsbereiche beschäftigen (siehe z.B. in Hasselhorn & Gold, 2013; Mandl & Zumbach, 2008; Mietzel, 2007; Seidel & Krapp, 2014; Wild & Möller, 2015). Es gibt nur wenige und ältere Lehrbücher, die sich explizit aus einer pädagogisch-psychologischen Perspektive auf das Geschehen in konkreten Lehr-Lernsituationen in der EB/WB beziehen (Fricke & Sarges, 1986; Weinert & Mandl, 1997). Auch in eigentlich nicht bildungsbereichsgebundenen Zeitschriften, die sich dem Lehren und Lernen widmen (z.B. Learning and Instruction, Instructional Science, Journal of Educational Psychology, Zeitschrift für Pädagogische Psychologie, Zeitschrift für Erziehungswissenschaft, Zeitschrift für Bildungsforschung) fokussieren die Beiträge zumeist den Schul- oder Hochschulbereich. Ein Grund hierfür könnte sein, dass für diese Bereiche – verglichen mit der EB/WB – relativ einfach auf sogenannte „anfallende Stichproben“ zurückgegriffen werden kann.

Zu einer sehr ähnlichen und aktuellen Einschätzung des Forschungsstands für den Bereich der EB/WB kommen Kuper und Schrader (in Druck). Auch die Arbeiten von Boeren (2018) und Daley et al. (2018), die eine Übersicht zu den verwendeten Forschungsmethoden in der EB/WB-Forschung geben, verweisen auf eine sehr geringe Zahl der Studien, die für den Bericht von Effekten nötige (quasi-)experimentelle Designs und/oder die entsprechenden Analysemethoden verwenden (siehe auch Fuhr et al., 2015)

An dieser Stelle soll nicht unerwähnt bleiben, dass ein großer Teil der EB/WB-Literatur sich auf einer theoretisch-konzeptuellen Ebene im Rahmen des Professionalisierungsdiskurses u.a. mit der professionellen Kompetenz von Lehrkräften in der EB/WB beschäftigt (z.B. Egetenmeyer & Käßlinger, 2011; Egetenmeyer & Schüssler, 2012; Fuhr, 1991; Gieseke, Tietgens, Schäffter, Venth & Müller, 1988; Hof, 2001; Lattke & Jütte, 2014; Nittel, 2000; Pätzold, 2011; Schöb et al., 2016; Sgier & Lattke, 2012; Ziep, 1990), die teilweise Gegenstand der Konzeptualisierung des PPK sind. Eine wachsende Zahl an Studien

beschäftigt sich jenseits des professionellen Wissens mit der professionellen Kompetenz von Lehrkräften (siehe Studie 2, Kapitel 2).

Zusammenfassend ist festzuhalten, dass die für die Entwicklung von Aufgaben des PPK nötigen Befunde empirischer Lehr-Lernforschung für die EB/WB bislang nur in sehr geringem Ausmaß zu finden sind. Dies bedeutet für die Entwicklung eines Tests zur Erfassung des PPK, dass die für die Entscheidung nötigen Befunde, welche Antworten auf die Test-Aufgaben die richtigen bzw. falschen sind, aus dem Schulbereich verwendet und damit zunächst auf die EB/WB „übertragen“ werden. Die Befunde aus dem Schulbereich wurden einer Plausibilitätsprüfung hinsichtlich ihrer Allgemeinheit und damit Übertragbarkeit auf die EB/WB unterzogen (vgl. die in Studie 1 und 2 abgebildeten Aufgaben), bevor sie (vorläufig bzw. solange nicht empirische Befunde und/oder theoretisch-konzeptuelle Argumente dagegensprechen) Bestandteil des Tests wurden (vgl. hierzu auch die Diskussion und den Ausblick der vorliegenden Arbeit).

Besonders interessant für die Entwicklung eines Tests, der relevantes Wissen zur Gestaltung von Lehr-Lernsituationen prüft, sind die Wirkungsbeziehungen zwischen dem Lehrkräftehandeln und dem Lernerfolg der Teilnehmer/Schüler, die besonders effektiv sind. Einen Überblick über Studien und eine Aussage über die Effektivität bestimmter (lehr-lernmethodischer) Handlungen gibt z.B. die Arbeit von Hattie (2009), die ein hilfreicher, wenn auch nicht alleiniger Ausgangspunkt bei der Aufgabenentwicklung war. Auf die Auswahl der konkreten Aufgabeninhalte und das Aufgabenformat wird im nachfolgenden Kapitel eingegangen.

3.2 Validitätsüberlegungen bei der Operationalisierung des (allgemeinen) pädagogisch-psychologischen Wissens

Die konkreten Aufgabeninhalte sowie das Aufgabenformat stehen in Zusammenhang mit bzw. haben einen Einfluss auf die Validität von Testwerten. Die Auswahl der Aufgabeninhalte und des Aufgabenformats ist vor dem Hintergrund von Validitätsüberlegungen zu begründen. Kapitel 3.2.1 legt den Fokus auf den Zusammenhang zwischen Aufgabeninhalten und Validität, Kapitel 3.2.2 thematisiert die Rolle von Aufgabenformaten hinsichtlich der Validität.

Validität ist neben der Reliabilität, der Objektivität und der Skalierbarkeit eines von zehn Gütekriterien psychologischer Tests (z.B. Kubinger, 2003), dessen Wichtigkeit unumstritten ist (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014) und an verschiedenen Stellen hervorgehoben wird: „If scores derived from a test do not inform us about the desired construct, then it makes little difference how reliable, how precise, or how free from biases the scores may be“ (Markus & Borsboom, 2013, S. xiii).

Die Wichtigkeit der Validität der Testwerte begründet den Fokus der vorliegenden Arbeit auf Validitätsüberlegungen bei der Operationalisierung sowie der Validierung der Testwerte und damit vereinfachend gesagt auf die Frage, ob der Test misst, was er zu messen vorgibt (Hartig, Frey & Jude, 2012). Validität kann formal(-theoretisch) als Korrelation der Testwerte mit dem entsprechenden Verhalten (bzw. hier Wissen) außerhalb der Testsituation beschrieben werden (Moosbrugger & Kelava, 2012). Bei der Validierung von Testwerten ist es von großer Bedeutung, welches Verständnis von Validität zugrunde liegt. Denn von der Position, die man hierzu einnimmt, hängt ab, was als Hinweis auf die Validität der Testwerte angesehen wird und was nicht (Markus & Borsboom, 2013). Die vorliegende Arbeit folgt dem Verständnis von Validität nach Messick (1989), das auch hinsichtlich potenzieller Konsequenzen der Testnutzung (vgl. Borsboom & Wijsen, 2017) zur Anlage der vorliegenden Arbeit passend erscheint: „Validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores. As such, validity is an inductive summary of both the existing evidence and the potential consequences of test interpretation and use” (Messick, 1989, S.13).

Folgt man dem Verständnis von Validität nach Messick (1989), so ist das Verhältnis zwischen Reliabilität und Validität ein anderes wie es z.B. im Rahmen der klassischen Testtheorie (KTT) formuliert wird, wo Validität als Korrelation der Testwerte mit einem Kriterium verstanden wird. Da im mündlichen Diskurs innerhalb der Erwachsenenbildungsforschung gelegentlich auf die KTT zurückgehende Beschreibungen des Verhältnisses von Validität und Reliabilität rekurriert wird, werden diese Beschreibungen aufgegriffen und aktuellere Interpretationsmöglichkeiten aufgezeigt. Die häufig formulierte Aussage „Validität kann nicht größer sein als die Reliabilität“ ist formal aus der KTT ableitbar und geht auf die Verdünnungsformel (engl. attenuation formula) zurück (siehe z.B. Rost, 2004). Diese Aussage gilt dem Validitätsverständnis von Messick (1989) folgend nur bedingt bzw. würde nur dann gelten, wenn man Validität ausschließlich als Kriteriumsvalidität (wie in der KTT) verstehen würde (Markus & Borsboom, 2013). Auch das Reliabilitäts-Validitäts-Dilemma (z.B. Rost, 2014) ist formal aus der KTT ableitbar (Validität des Gesamttests ist gleich Item-Validität geteilt durch die Item-Reliabilität) und basiert auf dem Verständnis von Validität aus der Perspektive der KTT. D.h., folgt man dem Verständnis von Validität nach Messick (1989), dann gilt die in der KTT formal beschriebene Beziehung, dass die Validität eines Tests sinkt, wenn man die Reliabilität steigert, nicht zwingend. Vielmehr wird angenommen, dass vorliegende Objektivität und Reliabilität das Erreichen einer hohen Validität begünstigen, da ein geringer Messfehler wenig konstruktirrelevante Varianz bedeutet und es ein Ziel der Validierung ist, konstruktirrelevante Varianz bei einer Messung möglichst klein zu halten (Markus & Borsboom, 2013; Moosbrugger & Kelava, 2012).

Allerdings ist – unabhängig von den formalisierten Beziehungen zwischen Reliabilität und Validität – die Annahme plausibel, dass durch eine inhaltliche Homogenisierung der Aufgaben ein breites Konstrukt ggf. nicht mehr adäquat abgebildet wird (Rost, 2004), was aber durch die Aufteilung eines Konstrukts in mehrere Untertests verhindert werden kann (Rost, 2014).

Das konkrete Vorgehen bei der Validierung von Testwerten hängt davon ab, wofür ein Test letztlich verwendet werden soll. Der im Projekt Think entwickelte Test soll das PPK mit seinen Facetten im Rahmen von wissenschaftlichen Studien ausreichend zuverlässig erfassen können und bestenfalls auch bei diagnostischen Entscheidungen in der pädagogischen Praxis eingesetzt werden, sofern sich das Wissen als aussagekräftig für das Handeln erweisen sollte. Es soll *kein* Instrument entwickelt werden, mit dem im Sinne der Datenmodi nach Cattell (1957) durch subjektive Einschätzungen der Testnehmer gewonnene Q(uestionnaire)-Daten erzeugt werden, sondern ein Leistungstest. Leistungstests können nach Rost (2004) als halb-objektiv bezeichnet werden. Die von Rost (2004) in Anlehnung an Cattell (1957) angesprochene Objektivität meint nicht die Durchführungs- oder Erhebungsobjektivität, Auswertungsobjektivität und Interpretationsobjektivität, die zum Gütekriterium der Objektivität zusammengefasst werden (z.B. Lienert & Raatz, 1998). Hier ist die Erfassungsobjektivität gemeint, d.h. Testnehmer beurteilen sich nicht selbst (für das PPK siehe z.B. Vinepac-Project, 2008). Als *halb* objektiv werden Leistungstests bezeichnet, weil sie nur in eine Richtung verfälscht werden können (faking bad).⁸ Der im Projekt Think entwickelte Test wurde als power-test konzipiert, d.h. es zählt, ob die Aufgaben richtig oder falsch gelöst wurden, nicht wie viel Zeit dafür benötigt wurde (Rost, 2004).

Aufgrund des Verwendungsziels des Tests bietet sich eine möglichst umfassende *Konstruktvalidierung* an. Derzeit stellt die Konstruktvalidität den zentralsten Aspekt der Validitätstheorie dar (für eine Zusammenfassung der Geschichte der Validitätstheorie und damit der Entwicklung und Hierarchisierung/Ordnung verschiedenen Validitätsarten siehe z.B. Hartig et al. (2012), ausführlicher Markus und Borsboom (2013)). Die Konstruktvalidierung, in ihren Grundlagen zurückgehend auf Cronbach und Meehl (1955), hat zum Ziel, das nomologische Netzwerk einer Theorie bzw. eines Modells zu überprüfen – hier das nomologische Netzwerk des PPK. Das Vorgehen bei der Konstruktvalidierung kann in zwei Schritte unterteilt werden: die Überprüfung der Dimensionalität des interessierenden Konstrukts, auch als faktorielle Validierung bezeichnet (Hartig et al., 2012), und die Überprüfung der im nomologischen Netzwerk angenommenen Beziehungen des interessierenden Konstrukts mit anderen Konstrukten. Die Überprüfung der angenommenen

⁸ Zusätzlich betonte Cattell (1957), dass die Testnehmer nicht oder nur schwer erraten können sollten, welches Merkmal Gegenstand der Testung ist, was beim vorliegenden Test, wie bei den meisten Leistungstests, *nicht* der Fall sein soll.

Dimensionalität (und deren Bestätigung) kann als notwendige Voraussetzung für den zweiten Schritt angesehen werden (vgl. z.B. Hartig et al., 2012).

Die Dimensionalitätsannahmen für das PPK wurden im Rahmen der vorliegenden Arbeit zunächst durch die Konzeptualisierung des PPK in Studie 1 definiert. Annahmen zur Dimensionalität der Facette „Wissen über Lehr-Lernmethoden und -konzepte“ sind in Studie 2 ausformuliert. Ob bzw. inwiefern es gelingt, Hinweise auf die angenommene Dimensionalität des Wissens zu bekommen, ist nicht nur von der „Existenz“ der Dimensionalität abhängig, sondern auch von der Operationalisierung (Reckase, 2009). Auf das Vorgehen bei der Operationalisierung wird im folgenden Kapitel eingegangen.

3.2.1 Validität und Aufgabeninhalte

In Studie 1 wird der Rahmen des „Itemuniversums“ (Rost, 2004) des PPK durch die Facetten und Subfacetten definiert. Das Konstrukt wird allerdings nur bis auf die Ebene der Subfacetten, nicht bis auf der Ebene der Items bzw. Aufgaben definiert. Insofern kann bei dem in Studie 1 vorgenommenen Schritt nicht von einer Inhaltsvalidierung auf der Ebene der Items/Aufgaben gesprochen werden. Denn im Rahmen einer solchen Inhaltsvalidierung stellt sich die Frage, inwiefern eine verallgemeinernde Interpretation der auf einer das Itemuniversum repräsentierenden Auswahl an Items/Aufgaben zulässig ist (Cronbach, Gleser, Nanda & Rajaratnam, 1972), was z.B. in der Behavior Domain Theory thematisiert wird (Markus & Borsboom, 2013; McDonald, 1999, 2003). Im Rahmen der BDT sind die zentralen Validierungsschritte das interessierende Konstrukt und seinen „inhaltlichen“ Rahmen festzulegen und sicherzustellen, dass eine angemessene, das Konstrukt repräsentierende Auswahl an Aufgaben getroffen wird, d.h. die Inhaltsvalidierung steht im Rahmen der BDT im Vordergrund (Markus & Borsboom, 2013).

Für die Validierung der Testwerte steht die vorliegende Arbeit vor der Herausforderung, dass der Umfang der Personenstichprobe für die Pilotstudie (Studie 1) und Hauptstudie (Studie 2 und Studie 3) durch die zur Verfügung stehenden Projektmittel in seiner Größe und Repräsentativität eingeschränkt war. Bei beiden Stichproben handelt es sich um ad-hoc-Stichproben von EB/WB-Lehrkräften sowohl aus verschiedenen Fachbereichen und Reproduktionskontexten als auch aus verschiedenen Bundesländern und damit um nicht repräsentative Stichproben, was die z.B. von Kline (2016) benannten Konsequenzen u.a. hinsichtlich der Generalisierbarkeit der Befunde nach sich zieht.⁹

⁹ Aus forschungsstrategischer Sicht ist bzw. war die Beschränkung der Projektmittel und daraus folgend eine zunächst eher explorative Untersuchung anzustellen sinnvoll, da für den Bereich der EB/WB es erst seit kurzem möglich ist, eine repräsentative Stichprobe zu ziehen, da die Charakteristika sowie der Umfang der Population erst seit kurzem relativ – abgesehen von dem Bereich der betrieblichen Weiterbildung – gut beschrieben werden können (Autorengruppe wb-personalmonitor, 2016). Auf zukünftige Möglichkeiten der Stichprobenziehung wird auch im Ausblick dieser Arbeit näher eingegangen.

Aus der Stichprobengröße resultiert, dass nicht alle Facetten und die insgesamt ungefähr 70 im Projekt Think entwickelten Aufgaben mit jeweils mehreren Antwortoptionen Gegenstand der Analysen beider Beiträge sein können, sondern lediglich eine Auswahl an Aufgaben, da ansonsten ein für die Modellparameterschätzung ungünstiges Verhältnis zwischen der Anzahl an zu schätzenden Parametern und der Stichprobengröße entsteht. Richtgröße bei der Auswahl der Anzahl an Aufgaben war die Daumenregel von Bentler und Chou (1987), die fünf bis zehn Personen pro Parameter vorschlägt; für eine exakte Bestimmung der Teststärke sind allerdings Simulationsstudien das Mittel der Wahl, wie sie mit Softwarepaketen wie Mplus (Muthén & Muthén, 1998-2015) oder R (R Development Core Team, 2017) umsetzbar sind (siehe z.B. Carsey & Harden, 2014).

In Studie 1 wurden zwei Facetten¹⁰ mit insgesamt 9 Aufgaben ausgewählt, was bei $N = 66$ Schul-Lehrkräften und $N = 147$ EB/WB-Lehrkräften die von Bentler und Chou (1987) aufgestellte Daumenregel verletzt; da es sich hierbei allerdings um eine Pilotstudie handelte, auf deren Grundlage zunächst „nur“ erste Erkenntnisse zu einem Test zur Erfassung des allgemeinen PPK erreicht werden sollten, wurde dies in Kauf genommen. Eine die Facetten angemessen repräsentierende Auswahl an Aufgaben anzubieten, war für die Pilotstudie auch aus Mangel an Auswahlmöglichkeiten kaum möglich.

Für die Hauptstudie bedeutet die Daumenregel von Bentler und Chou (1987), dass bei einer zur Verfügung stehenden Stichprobe von $N = 212$ Lehrkräften aus der EB/WB maximal ca. 40 Modellparameter geschätzt werden sollten, was bei Aufgaben mit bis zu vier Antwortoptionen die Schätzung der Parameter von fünf Aufgaben möglich macht, je nach dem, welches Messmodell verwendet wird bzw. passend erscheint (die Anzahl der frei zu schätzenden Parameter wird in Studie 2, Kapitel 6 berichtet). Für Studie 2 stellte sich dann die Frage, welche Aufgaben Gegenstand der Analysen und damit Gegenstand des Tests sein sollen, die die Facette möglichst repräsentativ abbilden.

In Studie 2 steht die Facette „Wissen über Lehr-Lernmethoden und -konzepte“ im Fokus, die für die Gestaltung von Lehr-Lernprozessen in der EB/WB allgemein als zentral betrachtet wird. Vor dem Hintergrund der Vielzahl an Kombinationen aus Arbeits- und Sozialformen als Lehr-Lernmethoden (z.B. Peterßen, 2009) erscheint eine repräsentative Abbildung der Facette mit fünf Aufgaben schwierig und Validität, aus der Perspektive der BDT betrachtet als das Ausmaß, in dem die Testwerte auf das „Aufgabenuniversum“ bzw. den „universe score“ generalisiert werden, augenscheinlich eingeschränkt zu sein. Um dennoch ein möglichst inhaltsvalides Abbild des Aufgabenuniversums zu schaffen, erfolgt die Auswahl der Aufgaben wie nachfolgend beschrieben.

¹⁰ Die Auswahl der Facetten erfolgte vor dem Hintergrund, dass aufgrund der Pilotstudienresultate für die Facette „Wissen über Lernziele und deren Umsetzung/Erreichung“ von Allgemeinheit ausgegangen werden durfte, für die Facette „Wissen über Individual- und Lernprozessdiagnostik“ diese aber zu hinterfragen war (vgl. Studie 1).

Die EB/WB ist in ein Bildungsbereich, der insgesamt thematisch sehr breit gefächerte Veranstaltungen anbietet, die jedoch im Einzelfall auch sehr spezifisch sein können (vgl. Schrader, in Druck). Vor diesem Hintergrund besteht innerhalb der EB/WB-Forschung Einigkeit auf einer theoretisch-konzeptuellen Ebene, dass die Teilnehmerorientierung das zentrale Leitprinzip des didaktischen Handelns für Lehr-Lernveranstaltungen in der EB/WB ist, deren Ziel eine Passung zwischen den Voraussetzungen der Teilnehmer und den Lernanforderungen ist (z.B. Siebert, 2012). Um Teilnehmerorientierung herzustellen, muss die Lehrkraft einerseits während der Planung die Voraussetzungen der Teilnehmer berücksichtigen und andererseits eine Partizipation der Teilnehmer während Lehr-Lernveranstaltungen erreichen (Breloer, Dauber & Tietgens, 1980). Lehr-Lernveranstaltungen können grob in die Verlaufsformen *Einstieg*, *Erarbeitung* und *Ergebnissicherung* untergliedert werden (Arnold et al., 2005).

Konsequenz dieses Leitprinzips, das häufig auf Grundlage des Konstruktivismus begründet wird (z.B. Berzbach, 2005; Siebert, 2012), ist allerdings nicht, dass die Lehrkraft mit ihrer professionellen Kompetenz unbedeutend wird. Seidel und Reiss (2014) identifizieren die Fehlinterpretation der konstruktivistischen Position, dass konstruktivistische Lernumgebungen ohne strukturierende Maßnahmen und Vermittlungsaufgaben durch die Lehrkraft umgesetzt werden können. Ebenso plädiert z.B. Geißler (2016) aus einer konstruktivistischen Perspektive für eine Strukturierung von Lehr-Lernsituationen; denn gerade die Orientierung an biographisch unterschiedlichen (Lern-)Erfahrungen, Vorwissensbeständen und Nutzenerwartungen der Teilnehmenden und deren Einbezug in Lehr-Lernsituationen sind ein anspruchsvolles Unterfangen, dessen kompetente Bewältigung ein breites Wissen über Lehr-Lernmethoden und -konzepte voraussetzt (ähnlich Pietsch, 2010).

Die dem Einstieg folgende Erarbeitungsphase kann z.B. durch einen Lehrervortrag vorstrukturiert werden. Ein Lehrervortrag bietet sich dazu an, in ein Thema ein- bzw. wieder- oder weiter einzuführen oder auch dann, wenn Informationen an die Teilnehmenden weitergegeben werden oder ein Überblick über ein Thema geschaffen werden soll. Weniger geeignet ist die Vortragsmethode, wenn der Stoff für die Lernenden komplex ist oder eine Menge Details enthält (Terhart, 2005). Gerade mit Blick auf den Konstruktivismus, auf dessen Grundlage häufig in der EB/WB argumentiert wird und der damit häufig verbundenen negativen Konnotation des Lehrervortrags ist es wichtig, um dessen Funktionen zu wissen, weswegen eine Aufgabe zur Erfassung des Wissens über den Lehrervortrag in den Test einbezogen wurde.

Ein weiterer Schritt um teilnehmerorientierte Lehr-Lernsituationen herzustellen, ist es, an das Vorwissen bzw. die Vorerfahrungen der Teilnehmenden anzuschließen, das zunächst aktiviert werden muss (z.B. Helmke, 2014; Klieme, Schümer & Knoll, 2001), wozu

beispielsweise eine Concept oder Mind Map hilfreich sein können. Weniger geeignet um bereits vorhandenes Vorwissen zu aktivieren und mit neuem Wissen zu verknüpfen, Beziehungen innerhalb eines neuen Wissensbereichs zu erarbeiten und zu sortieren sind Wiederholungsstrategien. Diese dienen eher dazu, zu erlernendes Wissen im Gedächtnis aktiv zu halten (Friedrich & Mandl, 1997). Eine Aufgabe des Tests fragt danach, wie Vorwissen (nicht) aktiviert werden kann.

Gruppenarbeiten wird nicht nur aus einer konstruktivistischen Perspektive ein großes Potenzial hinsichtlich des Wissens- bzw. Kompetenzerwerbs der Teilnehmenden zugeschrieben (Hattie, 2009). Für die Gruppenarbeit gilt, wie für alle anderen Sichtstruktur-Elemente (Oser & Baeriswyl, 2001), dass deren Vorliegen alleine kein Qualitätsmerkmal ist, denn auch hier kommt es darauf an, wie die Interaktionen der Lernenden ablaufen (können). Die Lehrkraft muss dabei u.a. das Engagement bzw. die Lernmotivation aller Gruppenteilnehmer aufrechterhalten, die unter bestimmten Bedingungen günstige oder ungünstige Ausprägungen annehmen können, die sich z.B. im sogenannten „Trittbrettfahrerphänomen“ äußern können. Dabei ist es wichtig zu wissen, a) welche Bedingungen dies sind, z.B. Ankündigung der Bewertung der Gesamtgruppenleistung oder individueller Leistungen und b) durch welche Art von Aufgaben diese hergestellt werden können (zusammenfassend z.B. Wecker & Fischer, 2014). Die unter a) und b) genannten Aspekte waren Gegenstand jeweils einer Aufgabe.

Feedback nimmt sowohl in formativen als auch summativen Evaluationen und damit auch in der Phase der Ergebnissicherung eine zentrale Rolle ein. Sie ist eine der am häufigsten eingesetzten Methoden um Lernprozesse und Verhaltensänderungen zu unterstützen (vgl. Strijbos & Müller, 2014). Der Einfluss personaler Faktoren (Attributionen, Selbstwirksamkeit und Zielorientierung) auf die Verarbeitung von Feedback spielt eine besondere Rolle. Individuelle Ursachenzuschreibungen eines erhaltenen Feedbacks werden als wesentlicher Einflussfaktor für die Rezeption und Nutzung einer Rückmeldung angesehen (vgl. Strijbos & Müller, 2014). Die Attribution einer Rückmeldung ist bedeutsam, da mit den angenommenen Ursachen für ein Ereignis motivationale und emotionale Auswirkungen sowie Überzeugungen bezüglich möglicher individueller Handlungsspielräume verbunden sind. Zentral ist hier die Frage, ob ein bestimmtes Handlungsergebnis durch eine Person aktuell und zukünftig als beeinflussbar bewertet wird – dies entspricht vor allem einer internalen, variablen und kontrollierbaren Ursachenzuschreibung wie sie im Klassifikationsschema der Gründe für Handlungsergebnisse nach Weiner (1985) konzeptualisiert werden. Weiner (1985) unterscheidet in zeitlich *interne bzw. externe, stabile bzw. variable* und *kontrollierbare bzw. unkontrollierbare* Gründe für Handlungsergebnisse. Strijbos und Müller (2014) fassen die Befunde für die Übermittlung von Rückmeldungen zusammen und konstatieren, dass die Herausforderung darin besteht, beim

Feedbackempfänger eine interne Attribuierung anzuregen, um Veränderungsmotivation zu generieren, ohne hierbei Prozesse der Selbst-Verteidigung zu aktivieren. Eine Aufgabe zum Feedback ist Teil der Analysen in Studie 2.

Die Aufgaben zusammenfassend kann gesagt werden, dass diese weniger thematisieren, *ob* bestimmte Sozial- oder Arbeitsformen eingesetzt werden, sondern vielmehr, *wie* diese ein- bzw. umgesetzt werden, um bestimmte genannte Ziele wahrscheinlicher werden zu lassen; zu betonen ist an dieser Stelle, dass das Vorliegen bestimmter Sichtstrukturen zunächst keine Rückschlüsse auf die Qualität eines Lehr-Lernprozesses gibt, da innerhalb der gleichen Sichtstruktur die Interaktion zwischen Lehrenden und Lernenden völlig unterschiedlich gestaltet werden (Oser & Baeriswyl, 2001).

Dass diese fünf Aufgaben die Facette „Wissen über Lehr-Lernmethoden und -konzepte“ aus theoretisch-konzeptueller Perspektive befriedigend repräsentieren, ist unwahrscheinlich, allerdings stellen sie eine quasi „stichprobenartige“ Auswahl an relevanten Wissensinhalten dieser Facette dar. Eine weitere Frage ist, welche Dimensionalität den vorgestellten Aufgabeninhalten zugrunde liegt. Mit Blick auf die Unterschiedlichkeit der Aufgabeninhalte erscheint die Eindimensionalität der Facette „Wissen über Lehr-Lernmethoden und -konzepte“ unwahrscheinlich. Die Frage nach der Dimensionalität dieser Facette ist Gegenstand von Studie 2.

Das nachfolgende Kapitel begründet die Auswahl der Aufgabenformate im Projekt Think.

3.2.2 Validität und Aufgabenformate

Lienert und Raatz (1998) definieren einen Test als „wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (S. 1). Ein Test besteht zumeist aus mehreren Aufgaben (McDonald, 1999) und die einzelnen Aufgaben setzen sich aus zwei Teilen zusammen: Der Aufgabenstellung selbst, auch als Aufgabenstamm bezeichnet, und dem Antwortformat (Rost, 2004). Das Aufgabenformat als Kombination aus Aufgabenstamm und Antwortformat spielt für die Validität der Testwerte eine wesentliche Rolle (Haladyna & Rodriguez, 2013). Die Vor- und Nachteile verschiedener Aufgabenformate werden nachfolgend vor dem Hintergrund der Validität der Testwerte angesprochen. Der Fokus dieses Kapitels liegt dabei auf der *Aufgabenentwicklung* (Haladyna & Rodriguez, 2013), weniger auf dem Prozess der gesamten *Testentwicklung* (Downing & Haladyna, 2006).

Der *Aufgabenstimulus*, an den sich der Aufgabenstamm anschließt, kann unterschieden werden hinsichtlich seiner „fidelity, which is the closeness of any test task to a criterion behavior in the target domain“ (Haladyna & Rodriguez, 2013, p. 43). Arbeitsproben oder

Assessment Center wird eine „high fidelity“ zugeschrieben, weil Personen bezogen auf eine Aufgabenstellung „wirkliches“ Verhalten in situ zeigen müssen. Videobasierten Aufgaben wird eine „medium fidelity“ zugeschrieben, da der Aufgabenkontext bzw. Ausschnitte dessen in seiner/ihrer Komplexität relativ authentisch dargestellt werden kann/können. Textbasierten Aufgaben wird eine „low fidelity“ zugeschrieben, da eine Situation zumeist in ihrer Komplexität reduziert beschrieben werden muss, wodurch die „Echtheit“ der Situation verloren geht (Lievens, Peeters & Schollaert, 2008). Videosequenzen scheinen sich als Aufgabenstimulus für einen Test anzubieten, der sich auf Lehr-Lernsituationen bezieht, da die zumeist sehr vielschichtig sind (vgl. auch Studie 1).

Aufgaben können mit einem *freien oder gebundenen Antwortformat* konstruiert werden (Haladyna & Rodriguez, 2013; Jonkisz, Moosbrugger & Brandt, 2012). Vor dem Hintergrund der Validität eines Testwerts ist der Vorteil von freien Antwortformaten gegenüber gebundenen Antwortformaten, dass die Antwort auf eine Aufgabe selbst erzeugt werden muss, wodurch die Reaktionsmöglichkeiten der Personen nicht durch vorgegebene Antworten eingeschränkt sind (Rost, 2004). Zudem ist die Auswahl einer zufällig richtigen Antwort ausgeschlossen, wie sie bei verschiedenen Formen gebundener Antwortformate (z.B. Richtig-Falsch-Aufgaben, Single-Choice, Multiple-Choice) in unterschiedlichem Ausmaß möglich ist und damit die Validität eines Testwerts ggf. einschränkt (z.B. Jonkisz et al., 2012). Die Auswertungsobjektivität freier Antworten auf das Niveau gebundener Antworten zu heben, gelingt allerdings zumeist nur unter dem Einsatz großer zeitlicher (und damit auch finanzieller) Ressourcen, weswegen in der Regel die Validität freier Antworten durch Kodierfehler und damit eine geringere Auswertungsobjektivität eingeschränkt ist (Haladyna & Rodriguez, 2013; Rost, 2004). Die Auswertung freier Antworten mit Hilfe von Text-Mining, wie es z.B. mit der R-Software (R Development Core Team, 2017) möglich ist, kann diesem Nachteil zukünftig möglicherweise in Teilen Abhilfe verschaffen. Freie Antwortformate könnten „aufholen“ gegenüber gebundenen Antwortformaten hinsichtlich der Auswertungsobjektivität sowie der Auswertungsökonomie (v.a. im Rahmen von LSA), die berechtigterweise häufig ein (starkes) Argument für gebundene Antwortformate sind (z.B. Bühner, 2011; Rost, 2004) und den nicht unerheblichen Anteil von gebundenen Antwortformaten in LSA begründen, der allerdings über verschiedene LSA variiert (Gierl, Bulut, Guo & Zhang, 2017). Mit der Auswertungsökonomie verbunden sind zudem (v.a. in LSA große) finanzielle Einsparungen (Lau, Lau, Hong & Usop, 2011). Vorteil von gebundenen Antwortformaten ist zudem, dass – verglichen mit freiem Antwortformat – mehr Aufgaben im gleichen Zeitraum bearbeitet werden können, was bedeutet, dass eine reliablere Messung gelingen kann (Haladyna & Rodriguez, 2013). Zusammenfassend kann gesagt werden, dass vor dem Hintergrund einer möglichst hohen Validität der Testwerte videobasierte Aufgabenstämme mit offenem Antwortformat als am günstigsten eingeschätzt

werden können, wenn die Auswertungsobjektivität auf einem hohen Niveau sichergestellt werden kann.

Hinsichtlich des Zeitaufwands bei der Entwicklung zeigen sich Vorteile von offenen Antwortformaten gegenüber gebundenen, da – nicht nur für das PPK – sich die Entwicklung plausibler Distraktoren (möglicher falscher Antworten) als zeitaufwendig und schwierig erwiesen hat (Gierl et al., 2017), ebenso die Entwicklung von Aufgaben mittlerer bis hoher Schwierigkeit, weswegen für das PPK häufig auch ein freies Antwortformat verwendet wird (König & Blömeke, 2009).

Aufgrund der eingeschränkten zeitlichen Ressourcen bei der Entwicklung sowie der Auswertung der Aufgabenantworten bzw. der Entwicklung möglicher softwarebasierter Auswertungen im Projekt Think und nachteilig erscheinenden Auswertungsobjektivität offener Antworten, wurden zu einem großen Teil textbasierte Aufgaben mit gebundenem Antwortformat, zumeist Richtig-Falsch-Aufgaben entwickelt, von denen bis zu fünf einem gemeinsamen Aufgabenstamm zugeordnet wurden. Nachteil von Richtig-Falsch-Antworten gegenüber Single-Choice oder Multiple-Choice-Aufgaben ist die hohe Wahrscheinlichkeit, die richtige Antwort auf eine Antwortalternative zu erraten (50%), die bei Single-Choice-Aufgaben bei eins geteilt durch die Anzahl der Antwortalternativen liegt. Zur Berechnung der Ratewahrscheinlichkeit bei Multiple-Choice-Aufgaben siehe z.B. bei Bühner (2011) oder Rost (2004), wobei die Ratewahrscheinlichkeit mit zunehmender Anzahl an Antwortalternativen, und wenn mehrere Antwortalternativen richtig, sind steigt. Vorteil von Richtig-Falsch-Antworten ist, dass als falsch zu bewertende Antwortalternativen eines Aufgabenstamms keine Distraktoren wie in Single- oder Multiple-Choice Aufgaben darstellen, da jede Antwortalternative vom Testnehmer *tendenziell* neu bewertet wird, wohingegen bei Single- oder Multiple-Choice-Aufgaben die Antwortalternativen eher auch mittels Ausschlussverfahren ausgewählt bzw. ausgeschlossen werden (wobei auch bei mehreren Richtig-Falsch-Antworten pro Aufgabenstamm ggf. mittels Ausschlussverfahren vorgegangen wird). Vor- und Nachteile sowie Vorschläge für die zukünftige Aufgaben(weiter)entwicklung werden im Ausblick dieser Dissertationsschrift gegeben.

Aufgaben mit dem gleichen Aufgabenstamm, wie sie im Projekt Think entwickelt wurden, werden als Testlet bezeichnet (Wainer & Kiely, 1987). McDonald (1999) gibt eine Übersicht zu verschiedenen Aufgabentypen bei Leistungstests und wie diese üblicherweise bepunktet werden. Bei den in den Studien 1 - 3 verwendeten Aufgaben haben mehrere Richtig-Falsch-Antworten den gleichen Aufgabenstamm, und es gibt zwei Möglichkeiten, Antworten zu bepunkten: Entweder wird ein Summenscore gebildet (Beispiel: bei vier Antwortalternativen pro Aufgabenstamm können null bis vier Punkte erreicht werden) oder jede Antwortalternative einer Aufgabe wird separat als richtig oder falsch bewertet. Möglichkeiten, wie mit Testlets im Rahmen der Messmodellierung umgegangen werden kann

und wie diese Messmodelle hinsichtlich ihrer Anpassungsgüte auf die Daten bewertet werden können, sind Gegenstand von Kapitel 4.

4 Testvalidierung

Anschließend an den Gedanken der Konstruktvalidierung nach Cronbach und Meehl (1955), deren Ziel die Überprüfung des nomologischen Netzwerkes des interessierenden Konstrukts ist, wird in diesem Kapitel zunächst das Vorgehen bei der Überprüfung der Dimensionalitätsannahmen von Facetten des PPK vorgestellt (Kapitel 4.1), was auch als faktorielle Validierung bezeichnet wird (Hartig et al., 2012). Im Rahmen der faktoriellen Validierung spielt die Itemhomogenität eine zentrale Rolle. Die Frage, wie überprüft werden kann, ob und inwiefern die angenommene Dimensionalität über verschiedene Gruppen hinweg gilt und sich Hinweise auf die Messinvarianz des Tests bzw. Personenhomogenität zeigen, ist ebenfalls Gegenstand dieses Kapitels. Die Vorstellung des nomologischen Netzwerkes über die angenommenen Dimensionalitätsannahmen hinaus und das Vorgehen bei der nomologischen Validierung ist Gegenstand von Kapitel 4.2.

4.1 Faktorielle Validierung

Faktorielle Validität liegt vor, wenn die Dimensionalitätsannahmen empirisch gestützt werden können (Hartig et al., 2012). Eine Verbindung zwischen den durch einen Test gewonnenen Testwerten und dem interessierenden Konstrukt stellen psychometrische Modelle her. Durch die Analyse der Beziehung zwischen Konstrukt und Testwerten können Hinweise auf die faktorielle Validität eines Testwerts und damit letztlich des gesamten Tests gewonnen werden. Die Item Response Theory (IRT, z.B. Rasch, 1960) – auch als probabilistische Testtheorie bezeichnet (z.B. Rost, 2004) –, die KTT (grundlegend Lord & Novick, 1968) und die Behavior Domain Theory (Cronbach et al., 1972; Markus & Borsboom, 2013; McDonald, 1999) sind die derzeit am häufigsten herangezogenen Theorieansätze (Markus & Borsboom, 2013). Die daraus abgeleiteten Modelle stellen weniger Alternativen, sondern vielmehr Ergänzungen zueinander dar (Markus & Borsboom, 2013). Beispielsweise wird das Verhältnis zwischen der IRT und der KTT an prominenter Stelle folgendermaßen beschrieben: „die eine fängt dort an (klassische Testtheorie), wo die andere aufhört (die probabilistische Testtheorie), nämlich bei den Messwerten“ (Rost, 2004, S.12). Die KTT wird häufig auch als Messfehler-Theorie bezeichnet (Moosbrugger, 2012b), die sich im Kern auf die Zerlegung der beobachteten Messwerte in einen wahren Wert und einen zufälligen Messfehler bezieht, anhand derer das Ausmaß des Messfehlers bzw. die Reliabilität der Messung bestimmt werden kann. Zur Überprüfung der Validität kann die KTT, abgesehen von der Kriteriumsvalidität, keinen Beitrag leisten und die Frage kann nicht beantwortet werden, ob es gerechtfertigt ist, einzelne Testwerte eines Tests zu einem Gesamtwert aufzuaddieren (Hartig et al., 2012; Markus & Borsboom, 2013; Moosbrugger, 2012a). Daher wird im Rahmen dieser Arbeit für die Konstruktvalidierung auf das von der BDT und der IRT

thematisierte *reflektive Messmodell* zurückgegriffen, das vom *formativen Messmodell* unterschieden wird.

4.1.1 Reflektive und formative Messmodelle

Reflektive Messmodelle nehmen einen von der latenten Variablen auf die manifeste Variable gerichteten Zusammenhang an, wie z.B. im common factor model (grundlegend Thurstone, 1947) oder in Modellen der IRT. Eine graphische Veranschaulichung verschiedener reflektiver Modelle findet sich in Studie 2, Abb. 4. Reflektive Messmodelle sind von formativen Messmodellen zu unterscheiden, die einen von der manifesten Variablen auf die latente Variable gerichteten Zusammenhang annehmen (Edwards & Bagozzi, 2000). Ein wesentliches Unterscheidungsmerkmal der beiden Messmodelle ist, dass – falls ein reflektives Konstrukt abbildender Aufgabenpool gegeben ist –, verschiedene zufällig gezogene, aber repräsentierende Stichproben an Aufgaben aus diesem Aufgabenpool das reflektive Konstrukt an sich *nicht* verändern. Hingegen verändert die Zusammensetzung der Aufgaben oder Items die Bedeutung eines formativen Konstrukts, da dieses durch die Aufgaben bzw. Items konstituiert wird. Ein häufig genanntes Beispiel (Brown, 2006; Edwards & Bagozzi, 2000; Markus & Borsboom, 2013; Weiber & Mühlhaus, 2014) für ein formatives Konstrukt ist der sozioökonomische Status (SES). Würde man bei der Erfassung des SES beispielsweise das Einkommen nicht berücksichtigen, so würde sich die Bedeutung bzw. das Konstrukt an sich verändern, denn die latente Variable im formativen Modell wird durch die gewichtete Summe der Testwerte gebildet (Edwards & Bagozzi, 2000). Eine Veränderung der latenten Variablen kann also bei einem solchen formativen Modell schon durch die Veränderung nur einer Aufgabe bzw. eines Items bewirkt werden und die Inhaltsvalidität spielt bei der Auswahl der Aufgaben eine größere Rolle als bei einem reflektiven Messmodell – immer vorausgesetzt, dass ein Konstrukt repräsentierender Aufgabenpool für ein reflektives Messmodell gegeben ist. Um zu überprüfen, ob sich der angenommene Zusammenhang zwischen der latenten und den manifesten Variablen zeigt, werden für reflektive Messmodelle explorative (EFA) und/oder konfirmatorische Faktorenanalysen (CFA) verwendet. Würde ein formatives Messmodell angenommen werden, so wäre z.B. eine Hauptkomponentenanalyse Mittel der Wahl (Markus & Borsboom, 2013).

Die verschiedenen reflektiven Modelle unterscheiden sich unter anderem hinsichtlich des Skalenniveaus der manifesten und der latenten Variablen. Modelle mit kategorialen manifesten Variablen (z.B. Aufgabe gelöst/nicht gelöst) wie sie in der vorliegenden Arbeit verwendet werden, werden der IRT zugeordnet (Hartig & Höhler, 2010). Für Modelle mit kategorialen manifesten Variablen können, wie auch für Modelle mit kontinuierlichen manifesten Variablen, explorative und konfirmatorische Faktorenanalysen durchgeführt werden, unter der Verwendung von full information oder limited information Algorithmen, z.B.

marginal maximum likelihood Schätzungen (MML) via full information Algorithmen oder weighted least squares (WLS) Schätzungen via limited information Algorithmen (Chalmers, 2012; Muthén & Muthén, 1998-2015). Zur Schätzung der Modellparameter wird entweder eine Logit-Link-Funktion (bei maximum likelihood Schätzungen) oder eine Probit-Link-Funktion (z.B. bei WLS Schätzungen) verwendet (Muthén & Muthén, 1998-2015). In den Beiträgen der vorliegenden Arbeit wurden bei der Schätzung der Modellparameter mit ConQuest (Wu, Adam, Wilson & Haldane, 2007), dem R-Package TAM (Robitzsch, Kiefer & Wu, 2017) und in Mplus (Muthén & Muthén, 1998-2015) durchweg eine Logit-Link-Funktion (maximum likelihood Schätzungen) via full information Algorithmen verwendet.

4.1.2 Modelle der Item Response Theory

Ein zentrales IRT-Modell ist das dichotome Rasch Modell (Steigungen der Item Characteristic Curve (ICC) sind gleich eins) oder Einparameter-Logistisches Modell (1PL-Modell; parallel verlaufende Steigungen der ICC), das mehrere Annahmen formuliert und damit mehrere Eigenschaften hat, die es aus der Perspektive der Entwicklung eines Leistungstests attraktiv machen (z.B. Kubinger, 2005; Moosbrugger, 2012a). Dazu gehört die Annahme der lokalen stochastischen Unabhängigkeit der manifesten Variablen. Lokale stochastische Unabhängigkeit bedeutet, dass die Konstanzhaltung der latenten Variablen auf einem bestimmten Wert dazu führt, dass die Korrelationen der Personen für die manifesten Variablen verschwinden (Moosbrugger, 2012a). Diese ist eine Bedingung für die Itemhomogenität eines Tests, unter der verstanden wird, dass alle Aufgaben dieselbe (eine) latente Fähigkeit ansprechen. Mit der Eindimensionalität der manifesten Variablen ist eine weitere Annahme des Rasch Modells angesprochen, d.h., dass die Personenfähigkeiten- und die Aufgabenschwierigkeitsparameter auf einer gemeinsamen latenten Dimension liegen (z.B. Strobl, 2012). Eine weitere Annahme ist die spezifische Objektivität. Diese liegt dann vor, wenn die Schätzung der Personenfähigkeit nicht davon abhängt, anhand welcher Aufgaben die Personen verglichen werden (z.B. Strobl, 2012). Eine leistungsfähigere Person muss also unabhängig von der Aufgabe einen besseren Testwert erhalten als eine weniger leistungsfähige Person. Allerdings muss die Gültigkeit der spezifischen Objektivität für jede neue Gruppe von Personen, an der der Test angewendet werden soll, neu überprüft werden (z.B. Strobl, 2012). „Passt“ das (eindimensionale) Rasch Modell auf die Daten, so kann die Anzahl gelöster Aufgaben als eine suffiziente Statistik angesehen werden – eine weitere Annahme bzw. Eigenschaft des Rasch Modells –, und es wäre gerechtfertigt, den Summenscore zur Abbildung der Personenfähigkeit zu verwenden, da dieser dann die gesamte Information über eine Person ausschöpft (z.B. Bühner, 2011; Kubinger, 2005; Rost, 2004). Aufgabenparameter werden im Rahmen der IRT als fixed variables und Personenparameter als random variables beschrieben und für die Personenparameter wird

zumeist eine Normalverteilung angenommen (Chalmers, 2012; Wilson, Boeck & Carstensen, 2008).

Die Verwendung des Rasch Modells bzw. anderer IRT-Modelle hat mehrere Vorteile, die auch für die hier vorliegende Dissertationsschrift von Relevanz sind. Bei Gültigkeit des Rasch Modells für eine Population kann im Rahmen von Studien, bei denen nur eine begrenzte Zahl an Aufgaben verwendet werden kann, für verschiedene Gruppen von Personen nur eine Auswahl an Aufgaben eingesetzt werden und trotzdem kann eine Bestimmung der Testwerte auf einer gemeinsamen Skala vorgenommen werden. Dasselbe ist möglich, wenn im Rahmen von Längsschnittstudien Erinnerungseffekte ausgeschlossen werden sollen oder parallele Testformen verwendet werden, die jeweils unterschiedliche Aufgaben derselben Skala enthalten (Rauch & Hartig, 2012). Auch computerbasierte adaptive Tests, deren Ziel es ist, durch eine Anpassung der Aufgabenschwierigkeiten an die Personenfähigkeiten möglichst schnell und genau die Personenfähigkeit zu erfassen, verwenden Modelle der IRT (Frey, 2012).

In Modellen der IRT werden Personenfähigkeiten und Aufgabenschwierigkeiten auf einer gemeinsamen Skala (joint scale) abgebildet (z.B. Strobl, 2012), siehe Abb. 3 in Studie 1. Bei Verwendung von Logit-Link-Funktionen wird ein Logit-Wert gebildet. Diese Werte können, wie z.B. in der PISA-Studie normiert werden, z.B. auf einen Mittelwert von 500 und Standardabweichung von 100 Punkten (OECD, 2001). Dadurch werden die Testwerte anschaulicher und die Testnehmer können ggf. verschiedenen Kompetenzstufen zugeordnet werden.

Möglichkeiten zum Umgang mit Testlets

Dieser Abschnitt stellt Möglichkeiten vor, wie mit Testlets umgegangen werden kann. Mit einem gemeinsamen Aufgabenstamm gehen mehr oder weniger enge Beziehungen zwischen den Aufgaben einher, die sich ggf. nicht allein durch den Einfluss der allen Aufgaben eines Tests bzw. einer Facette – also über verschiedene Testlets hinweg – gemeinsamen latenten Variablen erklären lassen. Gibt es enge Beziehungen zwischen den Aufgaben und dem Aufgabenstamm z.B. durch das Aufgabenformat und/oder durch einen inhaltlichen Schwerpunkt, den die Aufgabe innerhalb eines Tests thematisiert, kann eine lokale Abhängigkeit der Antwortoptionen („local item dependence“, LID) auftreten. Damit liegt nicht Ein-, sondern Mehrdimensionalität vor, für die es verschiedene Möglichkeiten gibt, wie mit ihr umgegangen werden kann. Drei Ansätze können unterschieden werden (Wang & Wilson, 2005b), die alle Fälle des multidimensional random coefficients multinomial logit model (MRCMLM) nach Adams, Wilson und Wang (1997) darstellen:

1) Es wird ein (multidimensionales) dichotomes Rasch Modell auf die Daten angepasst (König & Blömeke, 2009), d.h., alle Aufgaben werden ungeachtet ihrer Zugehörigkeit zu Testlets als unabhängige Antwortoptionen analysiert und etwaige lokale Abhängigkeiten

werden nicht mitmodelliert, was zu einer Verzerrung bei der Parameterschätzung führen kann (Wang & Wilson, 2005b).

2) Im scorebasierten Ansatz bilden Summenscores für die jeweiligen Testlets den Ausgangspunkt der Analyse, berechnet als Summe der einzelnen Itemwerte innerhalb eines Testlets (vgl. Studie 1). Alle Muster von Antworten, die innerhalb des Testlets zum selben Score führen, fallen bei diesem Vorgehen in dieselbe Kategorie. Die Analyse der Testlets erfolgt dann mittels polytomen (multidimensionalen) Rasch Modellen, z.B. dem Partial Credit Modell (PCM, Masters, 1982); für eine Übersicht zu verschiedenen polytomen Rasch Modellen siehe Strobl (2012) oder Wang und Wilson (2005b). Diese Modelle umgehen bzw. missachten das Problem lokaler Abhängigkeiten von Items und sollten nur dann verwendet werden, wenn die lokale Abhängigkeit von Items innerhalb eines Testlets moderat ist und der Test einen relativ hohen Anteil an unabhängigen Items hat (Wang & Wilson, 2005b). Die Verwendung polytomer Rasch Modelle hat mehrere Nachteile, die im nachfolgenden Abschnitt aufgegriffen werden und durch die begründet wird, warum in Studie 2 auf die in Studie 1 verwendeten polytomen Rasch Modelle verzichtet wird und auf die in den letzten Jahren immer stärker genutzten item-basierten Methoden zurückgegriffen wird (Blömeke & Jenßen, 2017; Brunner et al., 2010; Brunner, Lüdtke & Trautwein, 2008; Masur, Teutsch & Trepte, 2017).

3) Bei item-basierten Methoden bilden die Antwortalternativen eines Testlets die gemeinsame Ausgangsbasis. Das bedeutet, dass die im Antwortmuster eines Testlets enthaltenen Informationen nicht verloren gehen und die aus den Antwortalternativen gewonnenen separaten Messwerte Ausgangspunkt der Messmodellierung sind und nicht die Summenscores der jeweiligen Testlets. Lokale Abhängigkeit von Items innerhalb eines Testlets wird durch zusätzliche latente Variablen explizit berücksichtigt, siehe z.B. das bifactor Model (Reise, 2012) oder das Rasch Testlet Model (Wang & Wilson, 2005b), wobei ersteres ein 2PL-Model ist (also zwei Modellparameter einführt, einen für die Personenfähigkeit bzw. die Aufgabenschwierigkeit und einen für die Steigung der item characteristic curve (ICC)), zweiteres ein 1PL-Modell ist, das nur den Parameter für die Personenfähigkeit bzw. die Aufgabenschwierigkeit einführt. Mit Hilfe dieser Modelle kann die Varianz der Antworten in generalfaktorspezifische Varianz und testletspezifische Varianz zerlegt werden. Dies stellt insofern einen Vorteil gegenüber dem Partial Credit Modell dar, als dass eine Aussage darüber getroffen werden kann, wie viel Varianz der Generalfaktor, häufig das interessierende Konstrukt, verglichen mit der testletspezifischen Varianz zwischen den Personen aufklärt. Generalfaktor und Testlet-Faktoren können im Rahmen eines weiteren Konstruktvalidierungsschritts separat Gegenstand der Konstruktvalidierung werden und entsprechend des Ergebnisses können dann die entsprechenden Testwerte berichtet werden, ggf. um die testletspezifische Varianz bereinigt.

Für die in den Punkten 1 bis 3 genannten Modelle gibt es, wie bereits angesprochen, jeweils die Möglichkeit nur einen Parameter für die Personenfähigkeit bzw. die Aufgabenschwierigkeit (Rasch, 1960) oder zusätzlich einen zweiten (2PL, Itemdiskriminationsparameter), und dritten Parameter (3PL, Rateparameter) in die Modellgleichung aufzunehmen (vgl. Birnbaum, 1968).

Auf der Grundlage von IRT-Modellen können zur Schätzung der Personenfähigkeiten verschiedene Personenschätzer genutzt werden, z.B. Weighted Likelihood Estimates (WLE), Expected A Posteriori Measures (EAP) oder Plausible Values (PV) (z.B. Davier, Gonzalez & Mislevy, 2009; Rost, 2004). In Studie 1 und Studie 3 wurden die aufgestellten Messmodelle gemeinsam mit dem aufgestellten Strukturmodell berechnet, wodurch etwaige Herausforderungen bei der Schätzung von Personenparametern umgangen werden, da durch die Maximum Likelihood Schätzungen unverzerrte Personenschätzer erhalten werden.

4.1.3 Modellgeltungstests

Ob bzw. inwiefern das angenommene Messmodell auf die Daten passt, kann hinsichtlich des gesamten Modells durch globale Modellgeltungstests und auf Itemebene überprüft werden. Um verschiedene IRT-Modelle (z.B. ein ein- vs. mehrdimensionales Modell oder ein second-order vs. bifactor Modell), hinsichtlich ihrer Passung auf die Daten gegeneinander zu überprüfen, gibt es verschiedene Möglichkeiten, von denen die nachfolgend vorgestellten alle auf der Likelihood eines Modells beruhen: „Die Likelihoodfunktion beschreibt die Wahrscheinlichkeit der Daten in Abhängigkeit von den Modellparametern unter der Annahme, dass das Modell gilt“ (Rost, 2004, S. 112). Dabei gilt: Je höher der Wert der Likelihoodfunktion ist, desto besser passt das Modell (vgl. Rost, 2004).

Eine Möglichkeit, auf der Basis der Likelihood unterschiedliche Modelle gegeneinander zu testen, ist die Verwendung eines Index aus der Likelihood und der Parameterzahl, der durch einen Chi-quadrat-Differenzen Test auf seine Signifikanz hin überprüft werden kann (Wu et al., 2007). Ein Vorteil dieses Vorgehens ist die inferenzstatistische Absicherung bedeutsamer Unterschiede, für die allerdings häufig die Voraussetzungen nicht erfüllt sind (vgl. Rost, 2004). Eine dieser Voraussetzungen ist, dass die Modelle ineinander geschachtelt sein müssen (nested models), d.h. eines der beiden Modelle stellt gegenüber dem anderen ein Obermodell (parental model) dar, aus dem das andere durch Restriktionen abgeleitet werden kann (Eid, Gollwitzer & Schmitt, 2010; Rost, 2004; Wu et al., 2007). Das restriktivere Modell weist dabei weniger frei zu schätzende Parameter auf (siehe z.B. Tab. 5 in Studie 1). Eine Schwierigkeit ist allerdings, dass das restriktivere Modell nicht durch das Null-Setzen einzelner Parameter aus dem Obermodell ableitbar sein darf, da ansonsten die Verteilung der Prüfstatistik unbekannt ist bzw. nicht einer Chi-quadrat-Verteilung folgt (vgl. Rost, 2004), weswegen in Studie 2 auf einen solchen Vergleich verzichtet wurde. Die im Rahmen des

Structural Equation Modeling (SEM) verwendeten Fit-Indizes (z.B. die inkrementellen Fit-Indizes wie der Comparative Fit Index (CFI) oder der Tucker-Lewis Index (TLI), z.B. Kline, 2016) können im IRT-Framework nicht zur Beurteilung der Modellgüte angewendet werden (würde man mittels eines der oben genannten WLS-Schätzer für manifeste kategoriale Variablen die Modelle schätzen, könnten diese Fit-Indizes berechnet und berichtet werden).

Eine weitere Möglichkeit für den Modellvergleich sind informationstheoretische Maße (zusammenfassend z.B. Eid et al., 2010; Rost, 2004). Auch informationstheoretische Maße basieren auf der Likelihood eines Modells und berücksichtigen die Anzahl der Parameter der Modelle. Für die Berechnung der informationstheoretischen Maße müssen die Voraussetzungen für inferenzstatistische Schlüsse nicht gegeben sein, da auf einen inferenzstatistischen Schluss verzichtet wird. Zudem müssen die Modelle nicht in der oben beschriebenen hierarchischen Beziehung stehen und können nach beliebigen Restriktionen voneinander abgeleitet werden (Rost, 2004). Informationstheoretische Maße „belohnen“ Modelle mit weniger Parametern und geringerer Komplexität gegenüber Modellen mit vielen Parametern und höherer Komplexität (zusammenfassend z.B. Eid et al., 2010). Mit vielen Parametern kann zwar eine bessere Anpassungsgüte eines Modells erreicht werden, allerdings sinkt dann zumeist die Generalisierbarkeit des Modells, da empirische Daten zufallsbedingt fehlerbehaftet sind; Hintergrund dieses Vorgehens ist das auf Ockham zurückgehende Sparsamkeitsprinzip (siehe z.B. Heller, 2012). Verschiedene Informationskriterien werden unterschieden (siehe z.B. Gollwitzer, 2012): AIC (Akaike Information Criterion, Akaike (1973)); der AICc stellt eine Korrektur des AIC für kleine Stichproben dar; BIC (Bayesian Information Criterion, Schwarz (1978)) und CAIC (Consistent AIC, Bozdogan (1987)). Für die Informationskriterien gilt: Je niedriger, desto besser passt das Modell auf die Daten.

Eine Möglichkeit, um das durch die IC-Funktion festgelegte IRT-Modell hinsichtlich seiner Passung auf die verschiedenen Aufgaben bzw. Items zu überprüfen, ist neben einem graphischen Test der ICC der Aufgaben die Berechnung und Interpretation des Mean Square Fit Index (MNSQ). Eine Übersicht zu verschiedenen Item-Fit-Indizes für IRT-Modelle gibt z.B. Rost (2004). Der MNSQ kann als *ungewichteter* (Outfit) oder als mit der individuellen Varianz *gewichteter* (Infit) Wert berechnet werden (Wilson, 2005). Zumeist wird der gewichtete MNSQ (Infit) empfohlen. Der MNSQ ist das mittlere quadrierte Residuum zwischen der aufgrund der Modellannahme erwarteten und der tatsächlichen Antwort einer Person auf eine Aufgabe/Item. Der MNSQ hat einen Erwartungswert von 1.0 und seine Werte können zwischen 0 und unendlich liegen. Bei einem MNSQ größer als 1.0 spricht man von underfit, bei Werten kleiner als 1.0 von overfit und Werte zwischen 0.8 und 1.2 werden als Hinweise auf einen guten Item-Fit gewertet (Wright & Linacre, 1994). Underfit (MNSQ > 1.0) deutet darauf hin, dass die Antwortmuster eher zufällig erscheinen. Für die

Testentwicklung bedeutet dies, dass v.a. Items mit einem (sehr) hohen MNSQ aus dem Test entfernt werden sollten. Overfit ($MNSQ < 1.0$) bedeutet, dass eher deterministische Antwortmuster vorliegen, wie sie im Guttman-Modell (siehe z.B. in Rost, 2004) zu beobachten sind d.h., eher ein „Sprung“ von „nicht lösen zu lösen“ im Antwortmuster zu beobachten ist. Dies kann im Rahmen eines Leistungstests bedeuten, dass gewisse Redundanzen zwischen den Aufgaben vorliegen können, also z.B. ein Lösungsmuster auf mehrere Aufgaben anwendbar ist (vgl. z.B. Bühner, 2011). Items mit einem niedrigen MNSQ können damit aufgrund ihrer Guttman-Modell-ähnlichen Eigenschaften nur wenig detaillierte zusätzliche Informationen über die Fähigkeit der Personen beitragen (Wright & Linacre, 1994).

Aufgrund seiner für einen Leistungstest vorteilhaften Modellannahmen (siehe Kapitel 4.1.2) wäre im Rahmen dieser Arbeit die Passung einer Rasch bzw. 1PL Modell-Variante der in den Punkten 1-3 genannten Modellen auf die Daten „wünschenswert“ und ist entsprechend Gegenstand der Modellprüfungen. Ein weiterer Schritt zur Überprüfung der Modellkonformität ist die im folgenden Kapitel berichtete Prüfung auf Personenhomogenität.

4.1.4 Überprüfung auf differentielle Itemfunktionen

Nachdem eine Skala zur Erfassung eines Konstrukts entwickelt wurde, stellt sich die Frage, ob bzw. inwiefern die Aufgaben in verschiedenen Personengruppen (oder zu verschiedenen Messzeitpunkten) dasselbe Konstrukt erfassen und die betrachteten Testwerte demnach für verschiedene Personengruppen auf derselben Skala liegen. Die Überprüfung dieses Phänomens bzw. Konzepts kann sowohl im Rahmen des SEM als auch der IRT vorgenommen werden (Reise, Widaman & Pugh, 1993) und wird als Überprüfung auf Messinvarianz (Bezeichnung wird häufig im Rahmen des SEM verwendet) oder differentielle Itemfunktion (DIF, häufig im Rahmen der IRT) bezeichnet (Buchholz & Hartig, 2017). Messinvarianz liegt vor, wenn die Zusammenhänge zwischen den Aufgaben bzw. Items nicht von der Gruppenzugehörigkeit (oder verschiedenen Messzeitpunkten) abhängen (z.B. Kline, 2016), verschiedene Personengruppen also dasselbe latente Konstrukt zur Beantwortung einer Aufgabe einsetzen, was auch als Personenhomogenität bezeichnet wird (Rost, 2004). Personenhomogenität und damit Messinvarianz bzw. kein DIF von Aufgaben über verschiedene Gruppen hinweg ist eine nötige Voraussetzung, um Gruppenunterschiede (z.B. Mittelwerte oder Varianzen) als Unterschiede in Konstruktausprägungen zwischen verschiedenen Gruppen oder derselben Gruppe über die Zeit interpretieren zu können (Brown, 2006). Die Überprüfung auf Personenhomogenität kann durch die Aufteilung der Stichprobe nach einem relevanten Merkmal und dem Vergleich der Zusammenhangsmuster sowie der Aufgabenparameterschätzungen zwischen den Gruppen vorgenommen werden.

Analysen der Personenhomogenität stellen letztlich auch Modellgeltungstests dar (Strobl, 2012).

Im Rahmen des SEM ist zur Überprüfung auf Messinvarianz die konfirmatorische Mehrgruppen-Faktorenanalyse (Multigroup confirmatory factor analysis, MGCF, Jöreskog (1971)) das übliche Verfahren und mindestens drei verschiedene Stufen der Messinvarianz werden unterschieden (Buchholz & Hartig, 2017; Kline, 2016): Konfigurale (gleiches Ladungsmuster), metrische (gleiche Faktorladungen) und skalare (gleiche Faktorladungen und Intercepts) Messinvarianz. Zeigt sich im Rahmen der Modellierung mittels der IRT für eine Aufgabe bzw. ein Item über verschiedene Gruppen hinweg unterschiedliche Aufgaben- bzw. Itemschwierigkeiten, liegt DIF vor (Mellenbergh, 1989). Da die vorliegende Arbeit auf der IRT basiert, wird zur Überprüfung der Personenhomogenität auf DIF überprüft. Hinweise auf dieselben Zusammenhangsmuster (i.S. der konfiguralen Messinvarianz) kann die Überprüfung des Rasch Modells in verschiedenen Gruppen geben (vgl. Studie 1). Auf dieses Vorgehen wurde in Studie 2 auf Grund der geringen Stichprobengröße der Lehrkräfte mit bildungswissenschaftlichem Studium und damit verbunden der Schwierigkeit, das Rasch Modell zuverlässig auf die Daten anpassen zu können, verzichtet; in Studie 1 war die Stichprobe der Schul-Lehrkräfte für dieses Vorgehen ebenso (mehr als) grenzwertig klein.

Bei der Überprüfung auf DIF stellt sich die Frage, ob sich für Personen mit gleichen Werten auf der latenten Variablen eine Abweichung zeigt zwischen dem erwarteten und dem beobachteten Wert für eine Aufgabe bzw. ein Item. Ist dies der Fall und es unterscheiden sich Personengruppen mit gleichen Werten auf der latenten Variablen in ihrem erwarteten und beobachteten Wert, liegt DIF vor; ein DIF-Item weist bzw. würde auf einen Gruppenunterschied hinweisen, den andere Items eines Tests nicht abbilden (z.B. Mellenbergh, 1989). Dadurch ist die Validität der Testwerte eingeschränkt und die Personenfähigkeitsparameter (verschiedener Gruppen) sind nicht mehr direkt vergleichbar, da die Annahme der spezifischen Objektivität nicht mehr gültig ist. Folglich stellt sich die Frage, ob alle Personen bei der Lösung der Aufgaben die gleiche latente Fähigkeit einsetzen (vgl. z.B. Osterlind & Everson, 2009). DIF kann sich z.B. auch dann zeigen, wenn das „falsche“ Messmodell auf die Daten angepasst wurde, z.B. eine relevante latente Variable nicht mitmodelliert wurde (z.B. Ackerman, 1992). Die Gültigkeit des Rasch Modells muss für alle (Sub-)Populationen neu überprüft werden (grundlegend Gustafsson, 1980). Bestenfalls zeigt sich die Invarianz der Werte über alle denkbaren Gruppen hinweg, dann sind die Modellparameter stichprobenunabhängig (Rost, 2004).

DIF-Analysen sind Gegenstand von Studie 1 und im Rahmen dieser Arbeit wird auf DIF dadurch überprüft, dass ein Interaktionseffekt zwischen Aufgaben bzw. Itemschwierigkeit (dem Wert auf der latenten Variablen) und Personengruppen (als Moderatorvariablen)

modelliert wird und dessen Größe beurteilt wird (Robitzsch et al., 2017; Santelices & Wilson, 2012; Wang & Wilson, 2005a; Wu et al., 2007).

Werden Aufgaben bzw. Items mit DIF identifiziert, so stellt sich die Frage, wie mit diesen umgegangen wird, was auch von der Größe des DIF-Effekts abhängt und welche Ursachen für den Effekt in Frage kommen (Cho, Suh & Lee, 2016). Ohne sich die Frage gestellt zu haben, warum sich ein DIF-Effekt zeigt, sollte ein Item keinesfalls einfach aus einem Test entfernt werden, aber auch nicht schlichtweg ignoriert werden (Osterlind & Everson, 2009). Verschiedene Interpretationsmöglichkeiten, wodurch DIF entstehen kann, bietet Zumbo (2007).

4.2 Nomologische Validierung

Nach der Überprüfung der Item- und der Personenhomogenität stellt sich die Frage, ob das durch eine latente Variable abgebildete auch das aus theoretisch-konzeptueller Sicht angenommene Konstrukt ist: „The first idea is that one should not always assume that a common factor corresponds to an existing attribute.“ (Markus & Borsboom, 2013, S. 11). Im Rahmen dieses zweiten Schritts der Konstruktvalidierung stellt sich also die Frage, ob die latente Variable, für die sich Zusammenhänge zu den Aufgaben zeigen, auch tatsächlich das abbildet, was man gerne hätte, dass diese abbildet – hier die Facetten des PPK. Ist das der Fall, dann bedeutet dies, dass eine latente Variable und das theoretisch-konzeptualisierte Wissen, z.B. das Wissen über Lehr-Lernmethoden und -konzepte, sich auf dasselbe Konstrukt beziehen. Ziel der nomologischen Validierung ist es, Hinweise auf diese Frage zu bekommen.

Ansatzpunkte zur Entwicklung eines nomologischen Netzwerkes des PPK bietet das Modell der professionellen Kompetenz von Lehrkräften nach Baumert und Kunter (2011). Einigkeit besteht darin, dass professionelle Kompetenz von Lehrkräften (Kunter, Kleickmann, Klusmann & Richter, 2011) und damit auch das PPK erlernbar sind, z.B. durch ein erziehungs- und bildungswissenschaftliches Studium (Kunter et al., 2017; Sorge, Kröger, Petersen & Neumann, 2017), die praktische Lehrerausbildung (Referendariat) oder einschlägige Fort- und Weiterbildungen (Schrader, Anders & Richter, in Druck). Über den Zusammenhang zwischen PPK und der Unterrichtsqualität (Kunter & Voss, 2011; Voss et al., 2014) besteht ebenso Einigkeit und auch darüber, dass das PPK vermittelt über die Unterrichtsqualität mit dem Lernerfolg der Schüler/Teilnehmer positiv zusammenhängt (Lenske et al., 2016). Daher wird der Fokus in Studie 3 auf die Beziehung dieser Variablen zum PPK gelegt, abgesehen von der Unterrichtsqualität, da für diese keine entsprechenden Maße miterhoben werden konnten.

Das PCK ist neben dem PPK ein Bereich des professionellen Wissens von Lehrkräften (Shulman, 1986, 1987) und von diesem wird angenommen, dass es sich vom PPK

unterscheidet, da es letztlich spezifisches Wissen für den Wissenserwerb in einem bestimmten Fach ist, aber Bausteine des PPK enthalten kann. Daher werden in der Regel moderate Zusammenhänge zwischen dem PPK und dem PCK angenommen (z.B. Voss et al., 2011). Aus Mangel an in der EB/WB anwendbaren Instrumenten wurde das PCK im Rahmen der vorliegenden Arbeit nicht miterhoben und kann daher nicht in den Prozess der Konstruktvalidierung miteinbezogen werden, was aber ganz prinzipiell möglich wäre, wenn entsprechende Instrumente vorliegen würden.

Ein häufig verwendeter Kompetenzaspekt im Rahmen der Konstruktvalidierung des PPK sind die Überzeugungen von Lehrkräften (z.B. Voss et al., 2011). Professionelles Wissen und Überzeugungen stellen die kognitiven Aspekte der professionellen Kompetenz dar (Bromme, 2008). Beide beziehen sich auf die Gestaltung von Lehr-Lernsituationen und von beiden wird angenommen, dass sie in Zusammenhang mit der konkreten Ausgestaltung von Lehr-Lernsituationen und damit dem Lernerfolg der Teilnehmenden stehen, worauf Studien aus dem Schulkontext hinweisen (Kunter, Baumert et al., 2011). Als Hauptunterscheidungspunkt zwischen Wissen und Überzeugungen wird zumeist der epistemische Status der beiden Konstrukte aufgeführt (Hof, 2001; Voss, Kleickmann, Kunter & Hachfeld, 2011) und für beide Konstrukte wird erwartet, dass ihnen jeweils (mindestens) eine latente Variable zugrunde liegt. Überzeugungen könnten zur Validierung der Erfassung des Wissens herangezogen werden, wenn entsprechende Instrumente zur Erfassung dieser vorliegen würden. Da dies bislang nicht der Fall ist, können Überzeugungen im Rahmen dieser Arbeit nicht zur nomologischen Validierung herangezogen werden.

Zwei weitere Aspekte der professionellen Kompetenz von Lehrkräften sind die Selbstregulationsfähigkeit und die motivationalen Orientierungen, die beide für die psychische Dynamik und die Überwachung und Regulation des beruflichen Handelns verantwortlich sind (Klusmann, 2011). Die emotionale Erschöpfung als ein zentrales Symptom des Burnout-Syndroms (Maslach, Jackson & Leiter, 1996) wird häufig im Zusammenhang mit der Selbstregulationsfähigkeit thematisiert. Dem professionellen Wissen, damit auch dem PPK, wird eine Schutzfunktion zugeschrieben. D.h., je höher das professionelle Wissen, desto geringer ist die emotionale Erschöpfung, denn das professionelle Wissen erleichtert das berufliche Handeln der Lehrkräfte und hilft bei der alltäglichen Bewältigung von Arbeitsanforderungen (Klusmann, Kunter, Trautwein, Lüdtke & Baumert, 2008; Lauermaun & König, 2016). Inwiefern aber das PPK auch für Lehrkräfte in der EB/WB als Schutz wirksam werden kann, ist unklar. Denn die Lehrkräfte in der EB/WB sind zumeist nicht hauptberuflich tätig (Autorengruppe wb-personalmonitor, 2016). Deswegen liegt die Annahme nahe, dass, bei wahrgenommener übermäßiger und die psychische Gesundheit beeinträchtigender neben-beruflicher Belastung bzw. geringer Zufriedenheit mit der Lehrtätigkeit, zur Entlastung die Lehrtätigkeit (kurzfristig und einfach

möglich) reduziert oder beendet wird. Daher ist die baseline belasteter Lehrkräfte ggf. zu gering, um verlässliche Zusammenhänge feststellen zu können. Aus diesem Grund wird die Selbstregulationsfähigkeit nicht zur nomologischen Validierung herangezogen

Ein häufig beachteter Kompetenzbereich der motivationalen Orientierungen sind die berufsbezogenen Selbstwirksamkeitserwartungen von Lehrkräften (Depaepe & König, 2018; Lauermaun & König, 2016). Für diese ist unklar, in welcher Beziehung sie zum PPK stehen, weswegen sie nicht in den Prozess der Konstruktvalidierung aufgenommen werden. Denn einerseits wird angenommen, dass hohe Selbstwirksamkeitserwartungen mit dem PPK positiv zusammenhängen; andererseits erscheint die Annahme plausibel, dass ein hohes Wissen über die Komplexität der Gestaltung gelingender Lehr-Lernsituationen, auch darüber, was man als Lehrkraft falsch machen kann, nicht notwendigerweise die beruflichen Selbstwirksamkeitserwartungen positiv beeinflusst (vgl. Dicke et al., 2015).

Eine grundlegende Frage bei Leistungstests ist, welche Rolle domänenunabhängige, also allgemeine kognitive Fähigkeiten spielen. Dies wird in Studie 2 durch die Cattell-Horn-Carroll Theory of Cognitive Abilities (CHC; zusammenfassend siehe Flanagan & Dixon, 2013; Schneider & McGrew, 2012) angesprochen wird. Eine explizite Verbindung der wissensbezogenen Bereiche der professionellen Kompetenz von Lehrkräften zur CHC stellt die Arbeit von Blömeke und Jenßen (2017) her; auch Voss und Kunter et al. (2011) konnten einen Zusammenhang zwischen allgemeinen kognitiven Fähigkeiten und dem PPK zeigen. Eine ausdrückliche Berücksichtigung in Messmodellen des PPK, z.B. durch einen Generalfaktor, fanden allgemeine kognitive Fähigkeiten bislang noch nicht (siehe Blömeke & Jenßen, 2017). Kognitive Fähigkeiten wurden im Projekt ThinK nicht miterhoben. Offen ist damit die Frage, zu welchem Anteil sich Lösungswahrscheinlichkeiten auf allgemeine kognitive Fähigkeiten bzw. auf das PPK zurückführen lassen, was aus einer Validierungsperspektive eine zentrale Frage ist (vgl. auch Studie 2).

Zur nomologischen Validierung des PPK von Lehrkräften in der EB/WB können die eingangs erwähnten ausbildungsbezogenen Merkmale (erziehungs- und bildungswissenschaftliches Studium, einschlägige Fort- und Weiterbildungen, praktische Lehrerausbildung) herangezogen werden (siehe Studie 3). Maße der Unterrichtsqualität, PCK, Überzeugungen und allgemeine kognitive Fähigkeiten können in zukünftigen Studien zur Validierung herangezogen werden (siehe Gesamtdiskussion der vorliegenden Dissertationsschrift).

5 Fragestellungen der empirischen Studien

In den vorangehenden Kapiteln wurden die Schritte der Konzeptualisierung, Testentwicklung und -validierung eines Tests zur Erfassung des PPK vorgestellt. Ein Schwerpunkt lag dabei auf der Darstellung der Frage nach der Allgemeinheit des PPK in ihrer Bedeutung für die Allgemeine Didaktik und eine fach- und bildungsbereichsübergreifende (und damit die EB/WB einbeziehende) empirische Lehr-Lernforschung. Ein zweiter Schwerpunkt lag auf der Darstellung von Validitätsüberlegungen im Rahmen der Testentwicklung und Möglichkeiten der Validitätsüberprüfung der aus dem Test resultierenden Testwerten. Die in Kapitel 3 und 4 der Dissertationsschrift vorgestellten Schritte der Testentwicklung und Validierung sind Grundlage bzw. Vorarbeiten und Vorüberlegungen zu den drei Studien der vorliegenden Dissertationsschrift, die in den Studien selbst aufgrund der Veröffentlichung in Form von umfangsbeschränkten Zeitschriftenbeiträgen nur zusammengefasst und verkürzt aufgegriffen werden können.

In Studie 1 wurde der Schwerpunkt auf die *Konzeptualisierung* des allgemeinen PPK von Lehrkräften in verschiedenen Bildungsbereichen gelegt und die Frage gestellt, ob bzw. durch welche Facetten sich ein fach- und bildungsbereichsübergreifendes PPK inhaltlich beschreiben lässt. Ein zweiter Schwerpunkt von Studie 1 lag auf der *Erprobung eines Tests* zur Erfassung von zwei Facetten des PPK bei Lehrkräften aus Schule und EB/WB vor dem Hintergrund der Frage nach der Allgemeinheit dieses Wissens. Für die in dieser Studie verwendeten Aufgaben wurde der in Kapitel 4.1.2 vorgestellte scorebasierte Umgang gewählt und die Frage nach der Dimensionalität der beiden Facetten sowie nach DIF gestellt. Durch den zweiten Schwerpunkt ist Studie 1 auch der *faktoriellen Validierung* zuzuordnen.

In Studie 2 wurde die *Konzeptualisierung* des „Wissens über Lehr-Lernmethoden und -konzepte“ als Facette des PPK bei Lehrkräften in der EB/WB aufgegriffen. Vor dem Hintergrund der Vielzahl der unterschiedlichen Themen und Wissensbereiche, die innerhalb dieser Facette thematisiert werden, wird die Frage nach der Dimensionalität dieser Facette gestellt. Für die für diese Studie entwickelten Aufgaben wurde der itembasierte Umgang gewählt.

In Studie 3 lag der Schwerpunkt auf der *nomologischen Validierung* der Testwerte der Facette „Wissen über Lehr-Lernmethoden und -konzepte“. Es wurde die Frage gestellt, ob und inwiefern Elemente des nomologischen Netzwerkes des PPK, konkret der Tätigkeitsschwerpunkt der Lehrkräfte in allgemeiner Erwachsenenbildung oder beruflich-betrieblicher Weiterbildung, dem aktuellen Lehrumfang in Wochenstunden, einem erziehungs- und bildungswissenschaftlichen Studium, der Lehrerschaft in Jahren sowie

dem Umfang an absolvierten Stunden in Fort- und Weiterbildung in Zusammenhang zu diesem Wissen stehen.

6 Studie 1 Pädagogisch-psychologisches Wissen von Lehrkräften aus Schule und Erwachsenenbildung: Entwicklung und Erprobung eines Testinstruments

Marx, C., Goeze, A., Voss, T., Hoehne, V., Klotz, V. K. & Schrader, J. (2017). Pädagogisch-psychologisches Wissen von Lehrkräften aus Schule und Erwachsenenbildung. Entwicklung und Erprobung eines Testinstruments. *Zeitschrift für Erziehungswissenschaft*, 20 (S1), 165-200.

Zusammenfassung

Das pädagogisch-psychologische Wissen (engl. PPK) von Lehrkräften gilt als ein wichtiger Aspekt der professionellen Kompetenz von Lehrkräften. Die bislang entwickelten Konzepte zum PPK gehen implizit oder explizit davon aus, dass dieses allgemein ist und nicht spezifisch für verschiedene Fächer und/oder Bildungsbereiche. An dieser Annahme setzt der Beitrag an und geht empirisch der Frage nach der bildungsbereichs- und fachübergreifenden Allgemeinheit des PPK von Lehrkräften in Schule und Erwachsenenbildung (EB) nach. Aufbauend auf einer anhand von Expertenurteilen ($N = 44$) validierten Konzeptualisierung des PPK wird die Entwicklung eines digitalen Instruments zur Erfassung des PPK von Lehrkräften aus Schule und EB beschrieben, das text- und videobasierte Aufgaben integriert. Im Beitrag werden zwei Facetten des PPK exemplarisch fokussiert und an insgesamt $N = 212$ Lehrkräften aus Schule und EB auf Struktur, Reliabilität und differenzielle Item-Funktionen (DIF) hin überprüft. Die Ergebnisse weisen – weitgehend übereinstimmend mit den Befunden einer Inhaltsanalyse relevanter Literaturquellen sowie der durchgeführten Expertenbefragung – auf die bildungsbereichsübergreifende Allgemeinheit des PPK hin. Beide Facetten ließen sich reliabel messen, und es fanden sich geringe DIF-Effekte der Aufgaben. Die Befunde werden im Hinblick auf die Potenziale videobasierter Instrumente zur Kompetenzmessung bei Lehrkräften diskutiert.

Schlüsselwörter

Pädagogisch-psychologisches Wissen; Videobasiertes Testen; Schule; Erwachsenenbildung/Weiterbildung; Professionelle Kompetenz von Lehrkräften

Abstract

Teachers' pedagogical/psychological knowledge (PPK) is an important aspect of teacher competence. All conceptualisations developed so far assume that PPK is general and not specific for subjects or educational contexts. The authors take up this assumption and examine empirically whether PPK of school teachers and adult education teachers is general, i. e. generic for subjects and educational contexts. Based on a conceptualisation validated through an expert survey ($N = 44$) the development of a digital test instrument with text- and videobased items is described. The article reports on the results of the structure, reliability, and DIF analyses of two PPK-facets on a total of $N = 212$ school and adult education teachers; analyses showed that both facets can be measured reliably with only small DIF-effects. The results point to the fact that – largely in line with the results of a literature content analysis and the expert survey – PPK is general over educational contexts. Finally, results are discussed with regard to the potentials of videobased instruments to measure teachers' professional competence.

Keywords

Pedagogical-psychological knowledge; Videobased testing; School; Adult and further education; Teachers' professional competence

1 Allgemeinheit oder Spezifität pädagogisch-psychologischen Wissens: Positionen der Forschung

Zu den grundlegenden und empirisch gut belegten Annahmen der Forschung zur Expertise von Lehrkräften zählt, dass professionelles Wissen für die erfolgreiche Bewältigung beruflicher Anforderungen notwendig ist (Bromme 1992; Blömeke et al. 2010b; Baumert und Kunter 2011b). Mit Shulman (1986; 1987) hat sich die Untergliederung des professionellen Wissens in Fachwissen, fachdidaktisches und allgemeines pädagogisches Wissen, das oft auch als pädagogisch-psychologisches Wissen (pedagogical/psychological knowledge, PPK) bezeichnet wird, weitgehend durchgesetzt; diese Unterscheidung und ihre Einführung in den deutschsprachigen Raum durch Bromme (1992) sind in verschiedene Modelle der Kompetenzen von Lehrkräften eingegangen (Blömeke et al. 2010a, 2010b; Baumert und Kunter 2011a) und bilden den Ausgangspunkt einer großen Zahl von Forschungsprojekten zum Professionswissen von Lehrkräften, z. B. in den Projekten „Adaptive Lehrkompetenz“ (Beck et al. 2008), TEDS-M (Teacher Education and Development Study in Mathematics; Blömeke et al. 2010a), COACTIV (Cognitive Activation in the Classroom; Kunter et al. 2011) oder BilWiss (Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehrerausbildung; Terhart et al. 2012).

Etliche Studien zeigen für den Schulbereich, dass Teilaspekte des fachlichen und fachdidaktischen Wissens von Lehrkräften empirisch solide erfasst werden können, wenn auch bislang nur für ausgewählte Unterrichtsfächer wie etwa Mathematik (Riese und Reinhold 2012; Depaepe et al. 2013; Jüttner et al. 2013), dass es systematisch zwischen Lehrkräften mit unterschiedlichen Ausbildungswegen variiert (Krauss et al. 2008) und sich positiv insbesondere auf den Lernfortschritt der Schülerinnen und Schüler auswirkt (Baumert et al. 2010; Riese und Reinhold 2012; Depaepe et al. 2013; Jüttner et al. 2013). Für das (allgemeine) PPK, das als Wissen zur erfolgreichen Gestaltung und Optimierung von Lehr-Lernsituationen in verschiedenen Fächern definiert wird (Voss et al. 2011), ist der Forschungsstand weniger entwickelt (Voss et al. 2015). Die Bedeutung des PPK für die Gestaltung gelingender Lehr-Lernsituationen wird sowohl in der Allgemeinen Didaktik als auch in der pädagogischen Psychologie immer wieder betont (vgl. Baumert und Kunter 2011a); zudem ist dieses Wissen in pädagogischen Ausbildungsgängen schulischer und außerschulischer Bildungsbereiche seit jeher curricular ebenso fest verankert wie in Methodenratgebern (Grell und Grell 2010), die in großer Zahl genutzt werden (Hoffmann 2011). Die Entwicklung valider und reliabler Testinstrumente hat jedoch erst vergleichsweise spät begonnen. Daher liegen auch weniger Studien vor, die sich mit der Relevanz dieses Wissens für verschiedene Aspekte des Unterrichts bzw. des bereichsübergreifenden Lehrens beschäftigen. Die vorliegenden Studien weisen darauf hin, dass das PPK sich in der Unterrichtsqualität manifestiert, z. B. zusammenhängt mit der Effizienz der Klassenführung

sowie der konstruktiven Unterstützung der Lernenden (Voss et al. 2011, 2014) und darüber vermittelt Lernfortschritte indirekt unterstützt (Lenske et al. 2016). Weiterhin konnten Zusammenhänge des PPK zu Selbstwirksamkeitsüberzeugungen und beruflichem Wohlbefinden der Lehrkräfte aufgezeigt werden (Lauermaun und König 2016).

Explizit oder implizit gehen die bislang entwickelten Konzepte davon aus, dass das PPK *für verschiedene Fächer und/oder Bildungsbereiche* „gilt“ und es insofern generisch, d. h. allgemein und nicht spezifisch ist. Auffallend ist jedoch, dass *beide Annahmen* bislang noch nicht empirisch überprüft wurden. Dies gilt sowohl für die Allgemeine Didaktik als auch für die pädagogische Psychologie als zwei Teildisziplinen, die wesentlich zur Konzeptualisierung des PPK beigetragen haben.

Den Gegenstand der Allgemeinen Didaktik bilden „Fragen des Lehrens und Lernens auf *allen* Stufen des Bildungssystems und in *allen* inhaltlichen Lernbereichen“ (Terhart 2002, S. 77 – Hervorh. nicht im Orig.). Ähnlich stellt sich die Situation in der pädagogischen Psychologie dar: Nach Bromme ist mit pädagogischem Wissen „der Bereich des Wissens gemeint, der relativ unabhängig von den Fächern gültig ist“ (1997, S. 177; siehe auch Voss et al. 2015). Jedoch: „Historisch betrachtet und von seinen aktuellen Diskussionsforen her ist Allgemeine Didaktik sehr deutlich und dominierend von der Situation der Schule bzw. des schulischen Unterrichts aus konstruiert worden, und nur selten unter Rekurs auf andere (nichtschulische) Lehr- Lern-Felder“ (Terhart 2013, S. 226). Dies ist ebenso für den Bereich der pädagogischen Psychologie zu konstatieren. Daher ist es auch nicht überraschend, dass für den Schulkontext mittlerweile mehrere Instrumente zur Erfassung (von Facetten) des PPK von Lehrkräften entwickelt wurden, die das pädagogische und psychologische Wissen über Unterrichtsprozesse abbilden (z. B. König und Blömeke 2009; Voss et al. 2011). Im Projekt BilWiss (Kunina-Habenicht et al. 2012; Terhart et al. 2012; Kunter et al. 2017) wurde ein umfassendes Instrument zur Messung fachübergreifender bildungswissenschaftlicher Kenntnisse entwickelt. Der Test erfasst deklaratives und konzeptuelles Wissen über unterrichtsbezogene Inhalte sowie über das Unterrichten hinausgehende bildungsrelevante Inhalte (Kunina-Habenicht et al. 2012; Linninger et al. 2015). In BilWiss wird damit, verglichen mit den anderen Instrumenten zum PPK, eine breitere Konzeptualisierung vorgenommen, deren Grundlage Curricula verschiedener Lehramtsstudiengänge darstellten. Eine nicht nur fach- *sondern zudem auch explizit bildungsbereichsübergreifende* Konzeptualisierung, die über die Schul-Bildungsbereiche hinausgeht, sowie ein entsprechendes Instrument fehlen jedoch.

Die skizzierten Forschungsdesiderate bilden den Ausgangspunkt für das im Folgenden vorgestellte Projekt Think¹¹, das sich der Frage nach der *bildungsbereichs- und fachübergreifenden* Allgemeinheit oder Spezifität des PPK widmet. Der vorliegende Beitrag präsentiert erste Befunde aus diesem Projekt. Aufbauend auf einer im Projekt erarbeiteten und anhand von Expertenurteilen validierten fach- und bildungsbereichsübergreifenden Konzeptualisierung des PPK wird die Entwicklung eines Instruments zur Erfassung des PPK von Lehrkräften aus Schule und Erwachsenenbildung beschrieben und es werden Ergebnisse der Testentwicklung und -erprobung berichtet.

Das Instrument beinhaltet neben text- auch videobasierte Aufgaben. Durch videobasierte Aufgaben kann die Komplexität von Lehr-Lernsituationen authentisch abgebildet werden, ohne dass – anders als bei textbasierten Fallschilderungen – parallel ablaufende Geschehnisse (nicht) thematisiert und damit (nicht) hervorgehoben werden müssen. Da der Informationsgehalt von Videosequenzen der tatsächlichen Lehr-Lernsituation näherkommt als notwendigerweise selektierende Fallschilderungen, kann durch den Einsatz von Videos die ökologische Validität in der Erfassung handlungsrelevanten Wissens und Könnens gesteigert werden (Lievens et al. 2008). Daher wird auch in der Lehrerbildung erfolgreich mit authentischen Videofällen gearbeitet, z. B. verwendet der OBSERVER (Seidel et al. 2010) videobasierte Aufgaben zur Erfassung der professionellen Wahrnehmung von Unterrichtssituationen (Stürmer und Seidel 2015). Ebenso wird die Analyse- und Diagnosefähigkeit (Goeze et al. 2013) oder die Klassenführungskompetenz (Hellermann et al. 2015) mit Hilfe videobasierter Aufgaben erfasst und gefördert. Bisher entwickelte videobasierte Tests zur Erfassung des PPK zeigen ausschließlich Videos von und für Lehrkräfte aus dem Schulbereich und fokussieren einzelne Domänen, z. B. Mathematik (König et al. 2014) oder Physik (Beck et al. 2008), beziehen zum Teil aber auch verschiedene Fächer mit ein (Seidel et al. 2010).

Das hier vorgestellte Vorhaben öffnet mit dem Ansatz, PPK fach- und bildungsbereichsübergreifend erfassen zu wollen anhand von unterschiedlichen, aber stets pädagogisch möglichst generischen Videoinhalten, breitere Einsatz- und damit auch Vergleichsmöglichkeiten für und zwischen verschiedenen Fach- bzw. Bildungsbereichen – aber auch für und zwischen unterschiedlichen Formaten der Aus- und Fortbildung von Lehrkräften (von (non-)formalen bis hin zu informellen ‚learning by doing‘-Formen). Dies schafft die Grundlage für eine teildisziplinenübergreifende, gemeinsame erziehungswissenschaftliche Forschung an übergreifend relevanten Phänomenkreisen – hier dem (allgemeinen?) Wissen von Lehrkräften –, die bisher noch weitgehend Neuland ist.

¹¹ „Using Digital Media to Assess Generic Aspects of Teachers' Professional Knowledge in Different Educational Contexts“ ist ein Projekt-Cluster des WissenschaftsCampus Tübingen, das vom Deutschen Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e. V. in Bonn und dem Hector-Institut für Empirische Bildungsforschung in Tübingen bearbeitet wird.

Im nächsten Abschnitt erörtern wir aber zunächst die Frage, warum ein Vergleich des Lehrens und Lernens in Erwachsenenbildung und Schule für die Beantwortung der Frage nach der Allgemeinheit oder Spezifität des PPK hilfreich ist.

2 Lehren und Lernen in Erwachsenenbildung und Schule: Institutionelle, curriculare und didaktische Gemeinsamkeiten und Unterschiede

Das konzeptionell für wünschenswert gehaltene bzw. tatsächlich verfügbare PPK von Lehrkräften in Schule und Erwachsenenbildung (EB)¹² zu vergleichen, mag den einen naheliegend, den anderen dagegen ungewöhnlich erscheinen und bedarf daher der Begründung. Dies ist auch deshalb notwendig, weil die Schul- und Erwachsenenbildungsforschung nur selten systematisch aufeinander Bezug nehmen, zumeist ausgehend von der EB (siehe z. B. Tietgens 1967). Nur an wenigen Stellen wird auf die Vielzahl an Schnittmengen z. B. bei pädagogischen Konzepten verwiesen, die Wissens- und Erfahrungsaustausch zwischen den beiden Bildungsbereichen möglich machen (z. B. Zeuner 2003). Ein großer Teil der Erwachsenenbildungsforschung hat seine Aufgabe lange Zeit darin gesehen, die Eigenständigkeit des Lehrens und Lernens im „vierten“ Bildungsbereich dadurch zu legitimieren, dass die Differenzen zur Schule betont werden, etwa durch den Verweis auf die Freiwilligkeit der Teilnahme oder die Erfahrungs- und Handlungsorientierung des Lernens, während gleichzeitig immer schon und immer noch ein Teil ihrer nebenberuflich tätigen Lehrkräfte aus der Schule stammt(e) (Tietgens 1967). Beides hat den Blick auf *institutionelle*, *curriculare* und *didaktische* Gemeinsamkeiten und Unterschiede des Lehrens und Lernens in beiden Bildungsbereichen eher verdeckt als durch theoretisch informierte, vergleichende Analysen ausgeleuchtet.

Auf der *institutionellen* Ebene zeigen sich auf den ersten Blick erhebliche Differenzen zwischen Schule und EB: hier ein staatlich verantworteter, curricular reglementierter und für alle verpflichtender Bildungsbereich, dort ein pluralistisch von öffentlichen, korporativen, kommerziellen und betrieblichen Anbietern getragener und im Wesentlichen auf freiwilliger Teilnahme gründender Bildungsbereich (Schrader 2011). Gemeinsamkeiten zeigen sich jedoch u. a. darin, dass es – mindestens aus der Perspektive der Lehrkräfte – im Kern um *organisierte* (und nicht informelle) Lernprozesse geht und dass die Lehrkräfte in beiden Bereichen mit relativ hoher pädagogischer Autonomie agieren (können) (Fend 2009).

Die institutionellen Gemeinsamkeiten und Differenzen kommen vor allem in den Beschäftigungsbedingungen der Lehrkräfte zum Ausdruck, wie Tab. 1 zeigt. Während die EB sich inzwischen nach Teilnehmerzahlen zum größten Bildungsbereich entwickelt hat, ist es die allgemeinbildende Schule nach wie vor im Blick auf die Zahl der Unterrichtsstunden. Es

¹² Der Begriff der Erwachsenenbildung wird hier synonym mit dem der Weiterbildung verwandt, schließt also sowohl allgemeine und politische als auch berufsbezogene Angebote ein.

agieren in beiden Bildungsbereichen derzeit annähernd gleich viele Lehrkräfte.¹³ Unterschiede zeigen sich hier vor allem in den Anteilen der Geschlechter (höherer Anteil von Männern in der EB als in der Schule, insbesondere in der beruflichen und betrieblichen Weiterbildung), in der Altersstruktur (höherer Anteil von über 65-Jährigen in der EB), vor allem aber in den Beschäftigungsverhältnissen: Während die Schule mehrheitlich vollzeitbeschäftigte Lehrkräfte kennt, die als Beamte oder Angestellte an einer Schule arbeiten, wird die EB geprägt durch ehrenamtliche, neben- oder freiberufliche Vertragsverhältnisse mit einem hohen Anteil an Honorarkräften, die an meist mehreren unterschiedlichen Weiterbildungseinrichtungen lehrend tätig sind. Daraus ergeben sich u. a. erhebliche Differenzen in der Einkommenssituation. Große Differenzen zeigen sich zudem in der pädagogischen Vorbildung. Während ein pädagogisches Studium (und ein anschließendes Referendariat) die Regelvoraussetzung für eine Beschäftigung im Schulbereich darstellt, ist der staatlich bzw. tarifvertraglich gering reglementierte Bereich der EB „offen“ sowohl für ausschließlich fachlich als auch für (zusätzlich) pädagogisch qualifizierte Lehrkräfte; derzeit verfügt nur etwa ein knappes Drittel der Lehrkräfte über einen pädagogischen Haupt- oder Nebenfachabschluss.

Blickt man auf die *curriculare* Ebene und insbesondere auf die Inhalte und Themen des Lehrens und Lernens in Schule und EB, so zeigt sich auf den ersten Blick ein klar strukturierter, inhaltlich homogener Lehrplan der Schule als Teil eines „Weltcurriculums“ (Meyer 2005) und eine kaum überschaubare Vielfalt an Einzelthemen in der EB. Wählt man allerdings einen etwas höheren Auflösungsgrad und fragt nach den Modi der Weltbegegnung, die durch Lehr-Lernprozesse in Schule und EB gefördert werden, so zeigen sich eher Gemeinsamkeiten. Für die Schule hat Baumert (2002) in der Auseinandersetzung um die bildungstheoretische Fundierung international vergleichender Large-Scale-Assessments eine an die Tradition der kategorialen Bildung nach Klafki (1996) angelehnte Beschreibung des Kerncurriculums skizziert, die formale und materiale Bildung unterscheidet. Danach bereitet Schule, die eher formale Seite von Bildungsprozessen betonend, auf vier Modi der Weltbegegnung vor: eine *kognitiv-instrumentelle* Modellierung der Welt, eine *ästhetisch-expressive*, eine *normativ-evaluative* Weltbegegnung sowie eine *Auseinandersetzung mit Problemen konstitutiver Rationalität*. Für die EB rekonstruierte Schrader (2003) auf empirisch-induktivem Wege die Gesamtheit des Leistungsangebots von Weiterbildungseinrichtungen vor allem im Blick auf die Frage, welche Art von Wissen für die Weltbegegnungen Erwachsener angeboten und vermittelt wird: *Handlungswissen* zum Umgang mit der Welt der Sachen und Symbole, *Interaktionswissen* zur Verbesserung der

¹³ Da in den einschlägigen Studien der Erwachsenenbildung (zuletzt Deutsches Institut für Erwachsenenbildung et al. 2016) der Bereich der quantitativ äußerst bedeutsamen betrieblichen Weiterbildung empirisch nicht angemessen erfasst wird, dürfte die Zahl der Lehrkräfte in der EB tatsächlich größer sein als die Zahl der Lehrkräfte der Schule.

Handlungsfähigkeit gegenüber der sozialen Welt, soweit sie in (klein- und großgruppenbezogenen) Interaktionen zum Ausdruck kommt, *Identitätswissen* zum Umgang der Lernenden mit sich selbst, *Orientierungswissen* zum Umgang mit Fragen nach dem guten und richtigen Leben, nach Fundament und Sinn menschlicher Existenz, Werten und Normen.

Tab. 1 Lehrende in Schule und Erwachsenenbildung nach Geschlecht, Arbeitszeitvolumen und Abschluss

	Allgemeinbildende Schulen, insgesamt (N=664.140)		Berufliche Schulen (N=122.047)		Erwachsenenbildung (N=529.767 ^a)	
	N	%	N	%	N	%
<i>Geschlecht</i>						
Männlich	185.774	27,97	60.799	49,82	263.489	49,74
Weiblich	478.366	72,03	61.248	50,18	262.202	49,49
Ohne Angabe ^b	-	-	-	-	4 076	0,77
<i>Arbeitszeitvolumen^c</i>						
Vollzeit	415.160	62,51	84.868	69,54	95.281	17,99
Teilzeit	248.980	37,49	37.179	30,46	434.486	82,01
<i>Abschlüsse</i>						
Mit pädagogischem Abschluss ^d	632.633	95,25	98.717	80,88	137.358	25,93
Ohne pädagogischem Abschluss ^d	28.531	4,30	20.220	16,57	351.114	66,28
Mit pädagogischem Nebenfach ^d	-	-	-	-	36.550	6,90
Ohne Angabe	3 007	0,45	3 110	2,55	4 745	0,90

Anmerkung. Die Berechnungen basieren für den Bereich der Schule auf Auswertungen des Statistischen Bundesamts (2015a, 2015b) für das Schuljahr 2014/2015 und für den Bereich der Erwachsenenbildung auf Auswertungen des wb-personalmonitors (Erhebungszeitraum: Juni bis August 2014).

^a Angaben basieren auf Hochrechnungen.

^b Für das Merkmal Geschlecht werden keine fehlenden Werte ausgewiesen.

^c „Vollzeitbeschäftigte Lehrkräfte sind Personen im Beamten-, Angestellten- oder sonstigen Dienstverhältnis, die mit voller Regelpflichtstundenzahl [...] tätig sind. Teilzeitbeschäftigte Lehrkräfte sind Personen [...], deren individuelle Pflichtstundenzahl [...] bis zu 50 % der Regelpflichtstunden ermäßigt worden ist“ (Statistisches Bundesamt, 2015a, S.7). Die Berechnung für die Lehrkräfte der Erwachsenen-/Weiterbildung ist für die Arbeitszeitvolumina analog erfolgt.

^d Für die Lehrkräfte im schulischen Bereich wird hier die Angabe „mit“ bzw. „ohne Lehramtsprüfung“ gezählt. Die Gesamtzahl unter „Abschluss“ unterscheidet sich von der oben genannten, da für das Merkmal Geschlecht keine fehlenden Werte ausgewiesen werden, siehe Fußnote b.

Geht man auf die materiale Ebene der konkreten Fächer und Inhaltsbereiche, geht es nach Baumert (2002) im Anschluss an Tenorth (1994) im Schulbereich vor allem um fünf kulturelle

Basiskompetenzen: die Beherrschung der Verkehrssprache, eine mathematische Modellierungsfähigkeit, eine fremdsprachliche Kompetenz, eine IT-Kompetenz sowie – übergreifend – die Selbstregulation des Wissenserwerbs. In der EB finden sich analog neben dem weiten Feld der (fachbezogenen) beruflichen Weiterbildung Angebote der allgemeinen und politischen Weiterbildung. Hier geht es zum einen um eine kompensatorische Vermittlung schriftsprachlicher und alltagsmathematischer Basiskompetenzen in der Muttersprache und das Nachholen von Schulabschlüssen. Weitere Inhaltsbereiche zielen auf die Erweiterung von Kommunikations- und Schlüssel Fähigkeiten (Fremdsprachen, soziale Kompetenzen) und der Verhaltens- und Wissensausstattung Erwachsener für ihre besonderen Rollen und Lebensräume außerhalb von Politik und Beruf (in Angeboten der Familien-, Gesundheits-, Umwelt-, Verbraucher-, Freizeit- oder kulturellen Bildung) (Körper et al. 1995). Hinzu kommen Angebote der politischen Bildung im engeren Sinn. Während die Inhalte der Schule für *alle* verpflichtend sind, durch Noten bewertet werden und schließlich in Abschlüssen münden, handelt es sich in der EB um überwiegend freiwillig wahrgenommene Angebote, die in Abhängigkeit von Biografie und Lebenslauf genutzt werden und – abseits von abschlussbezogenen Weiterbildungen – zumeist keine Notengebung erfordern.

Auf der *didaktischen* Ebene bedeutet dies, dass die *Inhalte* der EB und Schule zwar variieren (können), wenn auch in geringerem Ausmaß als zumeist unterstellt, zugleich aber die *Ziele* in der Schule mehr von den nach Alter gestuften Laufbahnen und in der EB mehr von variierenden lebensweltlichen Anforderungen geprägt werden. Aus Sicht der Lehrkräfte sind es aber vor allem die Gemeinsamkeiten in den Strukturmerkmalen organisierter Lehr-Lernprozesse, aber auch vieler „Fächer“ bzw. Themenbereiche, die für das ihnen verfügbare PPK bedeutsam sind. Das gemeinsame pädagogische Fundament von Schule und EB ist zielgerichtetes und intentionales Lernen durch die Aneignung von Wissen, das in aller Regel in organisierten Gruppen und unter Anleitung einer Lehrkraft stattfindet (Siebert 2006). Die Gestaltung von Lehr-Lernsituationen folgt jeweils denselben grundlegenden Prinzipien und Zielen: unter Berücksichtigung individueller Merkmale der Lernenden (z. B. Teilnehmerorientierung/Schülerorientierung) sollen aufbauend auf der („Fern“-)Diagnostik des Lernstands der Lernenden bzw. dessen Einschätzung z. B. durch Klassenstufe, Sprachlevel (z. B. A2) oder andere Leistungseinteilungen (z. B. Word für Fortgeschrittene) möglichst gelingende Lernprozesse ermöglicht werden, um selbstbestimmte Lern- oder fremdbestimmte Lehrziele erreichen zu können; und nicht zuletzt findet sich für die konkrete Gestaltung von Lehr-Lernsituationen für die verschiedenen Bildungsbereiche vielfältig Literatur, die die gleichen Methoden, Konzepte und Lehr-Lernzielkategorien thematisieren, auch wenn sie selten als bildungsbereichsübergreifend ausgewiesen werden und jeweils bildungsbereichsspezifische Terminologien verwenden (Marx et al. 2014).

Insgesamt zeigt ein Blick auf die institutionellen, curricularen und didaktischen Merkmale von Schule und EB eine Fülle von Gemeinsamkeiten, so dass die Frage nach der bildungsbereichs- und fachübergreifenden Allgemeinheit des PPK von Lehrkräften in diesen beiden Bildungsbereichen viel versprechend untersucht werden kann (vgl. Abschn. 3). Diese Gemeinsamkeiten resultieren in gemeinsamen Anforderungen an die Lehrkräfte, z. B. bezogen auf das Wissen zur Diagnostik von Lernständen oder Kenntnisse über unterschiedliche Lehr- und Lernzielkategorien. Der Vergleich zeigt aber auch Unterschiede auf: Nicht alles Wissen – so lässt sich vermuten – müsste in beiden Bildungsbereichen die gleiche Relevanz und täglich geübte Anwendungspraxis aufweisen (z. B. Wissen über Bezugsnormen bei der Notengebung). Daher werden in diesem Beitrag u. a. die Wissensbereiche zur Diagnostik sowie zu Lehr- und Lernzielen einer genaueren empirischen Analyse unterzogen (Abschn. 4 sowie 6).

3 Fragestellungen

Die Frage nach der bildungsbereichs- und fachübergreifenden Allgemeinheit des PPK von Lehrkräften in Schule und EB wird systematisch in folgenden Schritten untersucht:

Fragestellung 1a: Wie fach- und bildungsbereichsübergreifend wird das PPK in der bisherigen Literatur relevanter Teildisziplinen „gedacht“, d. h. implizit oder explizit konzeptualisiert? *Fragestellung 1b:* Durch welche Facetten lässt sich ein fach- und bildungsbereichsübergreifendes PPK inhaltlich beschreiben? Diese Fragen werden auf der Basis einer umfassenden, disziplinübergreifenden Literaturstudie beantwortet. Die daraus erarbeitete fach- und bildungsbereichsübergreifende Konzeptualisierung des PPK wird in einem dritten Schritt durch Experten¹⁴ aus Wissenschaft und Praxis von Schule und EB validiert. Anhand qualitativer und quantitativer Daten wird der *Fragestellung 1c* nachgegangen: Wird, und ggf. inwieweit, dieser Konzeptualisierung von den Experten Vollständigkeit, bildungsbereichs- und fachübergreifende Relevanz und Allgemeingültigkeit zuerkannt?

Projektziel ist die Entwicklung, Erprobung und Fertigstellung eines auf Basis der Antworten dieser ersten drei Forschungsfragen konstruierten Testinstruments, das auf Vorarbeiten von Voss et al. (2011) beruht und mit dem im Projekt ThinK das PPK von Lehrkräften aus unterschiedlichen Bildungsbereichen und Fächern empirisch valide und reliabel erfasst werden soll.

Neben den obigen Fragestellungen 1a, b, c, die sich auf die Konzeptualisierung des PPK beziehen und in Abschn. 4.1, 4.2 und 4.3 beantwortet werden, soll in den darauffolgenden Abschnitten die Testentwicklung (Abschn. 5) sowie -erprobung (Abschn. 6)

¹⁴ Um den Lesefluss zu erleichtern, wird im Folgenden für Experten nur die männliche Geschlechtsform verwendet, es sind aber beide Geschlechter gemeint.

fokussiert werden, Letzteres unter der *Fragestellung 2*: Welche Unterschiede finden sich in der Struktur sowie Reliabilität des Konstrukts, und zeigen sich differenzielle Item-Funktionen für die Lehrkräfte der beiden Bildungsbereiche? Diese Frage wird, wie oben bereits erläutert, anhand von zwei Facetten des PPK („Wissen über Individual- und Lernprozessdiagnostik“ und „Wissen über Lehr- und Lernziele“) untersucht. Ab dieser Stelle beziehen sich die empirischen Daten ausschließlich auf verschiedene Bildungsbereiche, nicht Fächer.

Die Mehrstufigkeit des Vorgehens ist notwendig, um die Frage zu beantworten, ob bzw. in welchem Ausmaß das PPK allgemein oder spezifisch für unterschiedliche Bildungsbereiche und Fächer ist, also ob ggf. einzelne Facetten des PPK in einem der beiden Bildungsbereiche zusätzlich nötig, differenzierter erforderlich oder aber verzichtbar sind. Eine Antwort auf diese bislang unbearbeitete Forschungsfrage könnte nicht zuletzt die Diskussion um künftige Inhalte der Aus- und Fortbildung des Lehrpersonals in unterschiedlichen Bildungsbereichen beeinflussen.

4 Konzeptualisierungen des pädagogisch-psychologischen Wissens in Schule und Erwachsenenbildung

4.1 Wie fach- und bildungsbereichsübergreifend wird das PPK in der bisherigen Literatur relevanter Teildisziplinen implizit oder explizit konzeptualisiert?

Um vorliegende Konzeptualisierungen des PPK in Schule und EB zu erfassen, wurde zunächst eine systematische Literaturrecherche in zentralen Datenbanken (FIS-Bildung, ERIC, PsychINDEX, PsychInfo) mit insgesamt 303 datenbankspezifischen Schlagwortkombinationen durchgeführt (z. B. „Kursleiter“ und „Kompetenz“; „Lehrer“ und „Pädagogisches Handeln“; „trainers“ und „teaching skills“). Diese Recherche ergab mehr als 9000 Treffer. Die Literaturquellen wurden zunächst drei klar definierten Kategorien (einschlägig; gegebenenfalls einschlägig; nicht einschlägig) zugeordnet. Insgesamt wurden 600 Literaturquellen von zwei Urteilern doppelt kodiert. Die Urteilerübereinstimmung war mit einem gewichteten Kappa von $k_w = 0,75$ gut (Cohen 1968), so dass die übrigen Literaturangaben nur noch von je einem Codierer bewertet wurden. Insgesamt wurden 158 „einschlägige“, mehr als 800 „ggf. einschlägige“ und mehr als 8000 „nicht einschlägige“ Literaturquellen erfasst. Bei Sichtung aller „einschlägigen“ und ausgewählter Teile der „ggf. einschlägigen“ Literatur zeigte sich, dass die als relevant identifizierten Quellen auf 76 reduziert werden konnten, da etliche Literaturquellen denselben Forschergruppen entstammten bzw. sich jeweils auf das gleiche Modell von PPK bezogen wurde.

In der EB-Forschung wird die Frage, ob das PPK von Lehrkräften ähnlich oder unterschiedlich konzeptualisiert werden kann, vor allem auf die Gemeinsamkeiten und Unterschiede der Bildungsbereiche insgesamt, nicht der einzelnen „Fächer“ bezogen; die

Antworten sind nicht einheitlich. Auf der einen Seite stehen Autorinnen und Autoren, die die Gleichheit/Ähnlichkeit von Lehr-Lernsituationen in EB und Schule negieren (z. B. Ziep 1990; Brokmann-Nooren et al. 1994; Vinepac-Project 2008). Andere wiederum verweisen auf grundlegende Gemeinsamkeiten. So betont etwa Siebert mit Verweis auf Comenius, „dass es eine allgemeine Didaktik und besondere Didaktiken gibt. So unterscheidet sich eine Grundschuldidaktik von einer Didaktik der Altenbildung, gleichzeitig gelten bestimmte Aufgaben und Prinzipien für alle pädagogischen Felder“ (2006, S. 5). Tietgens konstatiert, dass „die Gesichtspunkte eines erwachsenengerechten Verfahrens die Grundbedingungen des Lehrens und Lernens nicht aufheben“ (1967, S. 9) und greift folgerichtig in seiner Einführung in das didaktische Handeln der Erwachsenenbildung auf die (schulbezogene) lehr-lerntheoretische Didaktik von Heimann, Otto und Schulz zurück (Tietgens 1981).

Für die Allgemeine Didaktik ist die Befundlage eindeutiger. Sie kennt eine Vielzahl von Theorien und Modellen (Kron 2014, Lüders 2014), für die noch kaum versucht wurde, sie in einem (Rahmen-)Modell zu integrieren (Zierer 2013). Gemeinsam ist diesen Ansätzen, dass sie sich trotz der ausgewiesenen Allgemeinheit ausschließlich auf den Schulbereich beziehen (Terhart 2013). Diesen Sachverhalt kritisiert ein prominenter Repräsentant der Schulpädagogik: „Eine Allgemeine Didaktik, die tatsächlich allgemein genug ist, um die unterschiedlichsten Felder des Lehrens und Lernens inklusive ihrer Rahmenstrukturen mit abzudecken und dabei sowohl Allgemeines als auch je Besonderes im Rahmen eines theoretischen Konzepts zu beschreiben, bleibt Aufgabe“ (Terhart 2013, S. 226).

Etwas anders stellt sich die Situation in der pädagogischen Psychologie bzw. in der empirischen Lehr-Lernforschung dar. Die Frage nach der bildungsbereichsübergreifenden Relevanz des PPK wird kaum gestellt und mit dem schulbezogenen Schwerpunkt der Forschung gleichsam implizit „beantwortet“. In den empirischen Studien bilden Lehramtsstudierende, Referendare oder ausgebildete Lehrkräfte durchweg die Stichproben, was mit Blick auf die fachübergreifende und schulbezogene Konzeptualisierung des PPK konsequent ist (Voss et al. 2015). Das Bewusstsein darüber, dass PPK ggf. *(k)ein* fachübergreifendes Wissen ist, besteht, da die verschiedenen Studien in der Regel ausdrücklich auf ihren Geltungsbereich hinweisen – z. B. Physik- (Lenske et al. 2016) oder Mathematikunterricht (Voss et al. 2011) – oder explizit einen fachübergreifenden Ansatz wählen (Kunter et al. 2017). Der Nachweis der Gültigkeit bzw. die Wirksamkeit (einzelner Facetten) des PPK als fachübergreifendes Wissen steht jedoch noch aus.

Festzuhalten ist demnach, dass empirische Prüfungen der explizit oder implizit formulierten Annahmen zur Allgemeinheit oder Besonderheit des PPK in der Allgemeinen Didaktik, der empirischen Lehr-Lernforschung sowie der EB-Forschung bislang nicht vorgenommen wurden.

4.2 Durch welche Facetten lässt sich ein fach- und bildungsbereichsübergreifendes PPK inhaltlich beschreiben?

Bisherige Konzeptualisierungen von PPK und verwandten Konstrukten beruhen in der Regel auf Untersuchungen der beruflichen, insbesondere didaktischen Anforderungen an Lehrkräfte. Die 76 Literaturquellen aus der oben beschriebenen systematischen Literaturrecherche, darunter z. B. die Ergebnisse der Delphi-Studie des BilWiss-Projekts (Kunina-Habenicht et al. 2012) oder die Konzeptualisierung aus dem Vinepac-Project (2008), wurden auf der Basis des Ablaufmodells induktiver Kategorienbildung nach Mayring (2008) zunächst explorativ inhaltsanalytisch auf ihre Facetten hin betrachtet. Als (Such-)Heuristik diente dabei die Variante des von Helmke entwickelten Angebots-Nutzungs-(Wirkungs-)Modells des Unterrichts von Seidel (2014). Sie unterteilt in Bereiche, die sich (vorwiegend) auf das Lehren (Angebot) oder vorwiegend auf das Lernen (Nutzung des Angebots, beeinflusst durch individuelle Merkmale) beziehen. Davon unterscheidet Seidel die Individual- und Lernprozessdiagnostik (welche Wirkungen hat das Angebot?); Letzteres benennt z. B. das BilWiss-Projekt – neben den beiden erstgenannten Bereichen – als die auf den „Klassenraum bezogenen Aufgaben“ (Linninger et al. 2015). Dieses Modell „passt“ grundsätzlich auf beide hier betrachteten Bildungsbereiche, wie der in Abschn. 2 durchgeführte Vergleich zeigt, der auf curriculare und didaktische Gemeinsamkeiten verweist.

Auf dieser Grundlage und der Konzeptualisierung des PPK aus dem COACTIV-Forschungsprogramm (Voss et al. 2011) wurden die erwähnten 76 Literaturquellen aus beiden Bildungsbereichen auf gemeinsame Kategorien hin gesichtet. Obwohl dieses Verfahren zu recht ähnlichen Ergebnissen führte, wurde schließlich mit dem auf diese Weise ergänzten Kategoriensystem das gesamte Material beider Bildungsbereiche erneut durchgearbeitet. Dieses Kategoriensystem stellte die vorläufige Konzeptualisierung des bildungsbereichsübergreifenden PPK im Projekt dar. Sie besteht aus folgenden acht Facetten, die sich aus 29 Subfacetten zusammensetzten (Marx et al. 2014) und in der Literatur beider Bildungsbereiche thematisiert werden.

1. Wissen über Lernprozesse von Lernenden
2. Wissen über Heterogenität der Lernenden
3. Wissen über Lehr-Lernmethoden und -konzepte
4. Wissen über Lehr- und Lernziele
5. Wissen über Führung von Lerngruppen
6. Wissen über Kommunikation und Interaktion mit Lernenden
7. Wissen über Gestaltung von Lernumgebungen
8. Wissen über Individual- und Lernprozessdiagnostik¹⁵

¹⁵ Die Facetten wurden, verglichen mit Marx et al. (2014), sprachlich geringfügig verändert.

Nach Voss et al. (2015) lassen sich diese acht Facetten auf einer übergeordneten Ebene gliedern in Wissen über Lernen und Lernende (1 und 2), Wissen über den Umgang mit der Lerngruppe als komplexem sozialen Gefüge (5 und 6), Wissen über das methodische Repertoire (3, 4 und 8) sowie Wissen über die Gestaltung von Lernumgebungen (7).

Die Konzeptualisierung inklusive ihrer Subfacetten ist in Tab. 2 und 3 exemplarisch für die beiden Facetten „Wissen über Lehr- und Lernziele“ und „Wissen über Individual- und Lernprozessdiagnostik“ abgebildet, die Gegenstand der weiteren Analysen in diesem Beitrag sind (vgl. die Subfacetten der übrigen sechs Facetten im Anhang).

Tab. 2 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Individual- und Lernprozessdiagnostik“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens</i> <i>V</i>	<i>V</i> <i>critical</i>	<i>N</i>
a Methoden zur Erfassung von Leistungen, Motivationen, Emotionen, Persönlichkeitsmerkmale und individuellen Besonderheiten	EB	0	4	7 ^a	7 ^a	3,17	0,79	0,72	0,67	18
	SK	0	1	6	18 ^a	3,68	0,56	0,89	0,64	25
b Bezugsnormen und deren Vor- und Nachteile	EB	0	4	8 ^a	6	3,11	0,76	0,70	0,67	18
	SK	1	3	6	14 ^a	3,38	0,88	0,79	0,64	24
c Wissen über Grundlagen der Diagnostik	EB	1	6 ^a	6 ^a	5	2,83	0,92	0,61	0,67	18
	SK	0	1	8	16 ^a	3,60	0,58	0,87	0,64	25

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext

^a Modus

Tab. 3 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Lehr- und Lernziele“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens V</i>	<i>V critical</i>	<i>N</i>
a Lernziel-taxonomien	EB	0	2	10 ^a	5	3,18	0,64	0,73	0,67	17
	SK	0	3	11 ^a	11 ^a	3,32	0,69	0,77	0,64	25
b Lernziel-kategorien	EB	0	4	6	7 ^a	3,18	0,81	0,73	0,67	17
	SK	0	5	10 ^a	10 ^a	3,20	0,76	0,73	0,64	25

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext

^a Modus

4.3 Inwieweit wird der entwickelten Konzeptualisierung des PPK durch Experten aus Schule und Erwachsenenbildung Allgemeinheit, Relevanz und Vollständigkeit zuerkannt?

Die auf der Basis der Forschung aus unterschiedlichen Bildungsbereichen entwickelte Konzeptualisierung des PPK wurde sodann einer Beurteilung von Experten aus Wissenschaft und Praxis von Schule und EB unterzogen, vor dem Hintergrund der Frage, ob und ggf. inwieweit dieser Konzeptualisierung Vollständigkeit, bildungsbereichs- und fachübergreifende Relevanz und Allgemeingültigkeit zuerkannt wird für die Gestaltung von Lehr-Lernsituationen (Forschungsfrage 1c). An früherer Stelle haben wir bereits dargelegt, dass die entwickelte Konzeptualisierung – abgesehen von einer Subfacette („Wissen über non-verbale Kommunikation“) – von Experten aus beiden Bereichen unterschiedslos als generisch, d. h. fach- und bildungsbereichsübergreifend gültig, eingeschätzt wird (Marx et al. 2014).

Tab. 4 Stichprobe der Expertenbefragung nach Bildungskontext sowie Zugehörigkeit zu Wissenschaft oder pädagogischer Praxis

	Experten aus der pädagogischen Praxis	Experten aus der Wissenschaft	
Experten Schulkontext	<i>n</i> =9	<i>n</i> =17	<i>n</i> =26 (59%)
Experten EB	<i>n</i> =10	<i>n</i> =8	<i>n</i> =18 (41%)
	<i>n</i> =19 (43%)	<i>n</i> =25 (57%)	<i>N</i> =44

Weitergehende Auswertungen einer offenen Ausgangsfrage an die Experten, was sie unter PPK verstünden, zeigten, dass nur ein wesentlicher Aspekt in der Konzeptualisierung unberücksichtigt war („Wissen über Kriterien der Unterrichtsqualität/der Qualität von Lehr-Lernsituationen“, die wir als weitere Subfacette aufgenommen haben). Antworten auf die Frage, *für wie relevant Experten das in den Facetten und Subfacetten thematisierte PPK für die Gestaltung von Lehr-Lernsituationen in ihrem jeweiligen Bildungsbereich halten*, wurden über ein vierstufiges Antwortformat (vgl. Tab. 2 und 3) erhoben. Alter und Geschlecht (vgl. Tab. 4) wurden nicht erfragt, um die Anonymität der Experten aus der Wissenschaft zu wahren.

Die Einschätzungen der Experten wurden mit Aikens *V* analysiert (Aiken 1985), einem Koeffizienten zur Bewertung der Inhaltsvalidität, der Werte zwischen 0 und 1 annimmt (je größer Aikens *V*, desto höher die Inhaltsvalidität). *Überschreitet* Aikens *V* den kritischen Wert ($V_{critical}$), ist dies hinsichtlich der Inhaltsvalidität *positiv* zu werten und bedeutet, dass die Experten überzufällig häufiger zustimmende als ablehnende Kategorien wählen. Schätzen die Experten *beider* Bildungsbereiche *diesselbe* Subfacette jeweils überzufällig häufig als relevant für ihren Bildungsbereich ein, weist dies auf die bildungsbereichsübergreifende Relevanz im Sinne einer Allgemeingültigkeit der Facette hin.

Die Experten aus EB und Schule bewerteten alle Subfacetten des PPK als relevant in „ihrem“ Bildungsbereich (vgl. Tab. 2 und 3 sowie im Anhang Tab. 9, 10, 11, 12, 13 und 14), lediglich eine Subfacette, „Wissen über Grundlagen der Diagnostik“ der Facette „Wissen über Individual- und Lernprozessdiagnostik“, wurde von den EB-Experten nicht überzufällig häufig als relevant für Lehr-Lernsituationen der EB eingeschätzt. Dieser Befund wird nachvollziehbar, wenn man bedenkt, dass „Diagnostik“ von den EB-Experten vermutlich mit der Vergabe von Noten assoziiert wurde, was in der EB im Gegensatz zur Schule tatsächlich in den meisten Veranstaltungen nicht vorkommt.¹⁶

¹⁶ Die Subfacette „Unterschiede zwischen den Geschlechtern“ der Facette „Wissen über Heterogenität der Lernenden“ wird von den Schul-Experten nicht überzufällig häufig als relevant für Lehr-Lernsituationen in der Schule eingeschätzt (vgl. Tab. 10 im Anhang). Eine mögliche Erklärung hierfür ist, dass die Schul-Lehrkräfte Unterschiede zwischen den Geschlechtern im Sinne einer sozialen Erwünschtheit tendenziell als nicht relevant

Die bisherigen Untersuchungsschritte zeigen, dass PPK in unterschiedlichen Forschungsdiskursen unterschiedlich stark als bildungsbereichsübergreifend konzeptualisiert wird. Auf der Basis dieser Forschungsliteratur gelingt es jedoch, PPK explizit bildungsbereichs- und fachübergreifend zu konzipieren, und zwar so, dass Experten aus beiden Bildungsbereichen kaum etwas hinzuzufügen haben: Insbesondere der Vergleich der jeweils bildungsbereichsspezifischen Experten-Einschätzungen zeigt, dass – bis auf wenige, zumeist gut interpretierbare Ausnahmen – alle 29, später 30 Subfacetten als vollständig das Phänomen erfassend, allgemein gültig und jeweils bereichsrelevant eingeschätzt werden.

5 Entwicklung von Aufgaben zur fach- und bildungsbereichsübergreifenden Erfassung des PPK

Anschließend an das Testinstrument von Voss et al. (2011) und der oben vorgestellten Konzeptualisierung des PPK mit ihren acht Facetten wurden 67 Testaufgaben entwickelt. Da das Diagnostizieren von Lern(zwischen)ständen und -prozessen, auch wenn es nicht in Noten mündet, ein zentraler Aspekt bei der Gestaltung von Lehr-Lernsituationen in allen Bildungsbereichen ist, wird diese Subfacette zunächst weiterhin als Teil des allgemeinen PPK konzeptualisiert, obwohl Experten aus der EB hier zu einer zurückhaltenden Einschätzung gekommen waren.

Die Aufgabenstämme bestanden entweder aus kurzen Texten oder Videos, an die sich entweder geschlossene oder offen gestellte Aufgaben anschlossen. Die videobasierten Aufgaben stellen ein wichtiges Element des neu entwickelten Testinstruments dar, da diese gegenüber textbasierten Aufgaben Vorteile versprechen. Neben der einleitend angesprochenen ökologischen Validität erreichen videogestützte (und interaktiv aufgebaute) Verfahren regelmäßig höhere Akzeptanzwerte – die höchste soziale Validität – bei den Teilnehmern (Kanning et al. 2006; Syring et al. 2015). Zudem bieten Videos als Aufgabenstimuli die Möglichkeit, prozedurales, stärker handlungsnahes Wissen von Lehrkräften zu erfassen und sich somit der Anwendung von Wissen in tatsächlichen Unterrichtssituationen anzunähern (Stürmer und Seidel 2015). In den bisherigen Formaten des videobasierten Testens der professionellen Kompetenz von Lehrkräften wurden entweder mehrere kurze (König und Blömeke 2009; Voss et al. 2011; Gold und Holodyski 2015) oder eine bzw. wenige längere Videosequenzen (Beck et al. 2008) eingesetzt.

In unserem Projekt wurden kurze Videosequenzen aus verschiedenen realen Lehr-Lernsituationen verwendet; pro Sequenz wurde nur eine Aufgabe formuliert, um etwaigen Aufgabenstamm-Effekten vorzubeugen. Die Lehr-Lernsituationen entstammen unterschiedlichen Fach- und Bildungsbereichen (EB, Hochschule und Schule, vgl. dazu z. B.

Schrader et al. 2010) und dokumentieren pädagogische Situationen, die in allen Bereichen auftreten können. Bei den Lernenden handelt es sich um Personen, die hinsichtlich ihres Alters sowohl Teilnehmende aus der EB als auch Schülerinnen und Schüler oder Studierende sein können. Um geeignete Videosequenzen zu identifizieren, wurden Videofälle aus einer Online-Plattform (www.videofallarbeit.de) analysiert, die zur Fortbildung von Lehrkräften aus verschiedenen Bildungsbereichen verwendet werden (Schrader 2010).

Eine videobasierte Aufgabe sowie jeweils eine textbasiert geschlossene und offene Aufgabe sind beispielhaft in Abb. 1 und 2 dargestellt.

Die Testaufgaben wurden vor ihrem Einsatz schrittweise durch eine Think-aloud-Vorstudie mit $N = 17$ Lehrkräften aus Schule ($n = 10$) und EB ($n = 7$) optimiert, um etwaige Quellen des Missverstehens zu identifizieren, die logische Konsistenz der Aufgaben und ihre sprachliche sowie inhaltliche Akzeptanz in beiden Bildungsbereichen zu überprüfen. Anschließend fand ein weiterer Pre-Test mit Studierenden (Lehramt und Erziehungswissenschaft, $N = 332$) statt, um erste Informationen über die Schwierigkeit einer Auswahl von Aufgaben zu erhalten. Für die darauffolgende Pilotstudie wurden die Aufgaben in eine bereits bestehende Lern- und Testumgebung für Lehrkräfte implementiert (www.videofallarbeit.de).

Im Rahmen der Auswertung der Pilotstudien wurden die Testaufgaben aller Subfacetten erneut überarbeitet, reduziert und/oder ergänzt, die Testaufgaben zu den Bereichen „Wissen über Individual- und Lernprozessdiagnostik“ sowie „Wissen über Lehr- und Lernziele“, die Gegenstand der nachfolgenden Analysen in Abschn. 5 sind, blieben weitestgehend unverändert bzw. werden für die weiteren geplanten Studien (zunächst) lediglich um Testaufgaben ergänzt und/oder reduziert.

Die Facette „Wissen über Individual- und Lernprozessdiagnostik“ wurde durch neun Aufgaben im Multiple-Choice-Format erfasst, die Facette „Wissen über Lehr- und Lernziele“ durch fünf Aufgaben, davon vier im Multiple-Choice-Format und eine mit offenem Antwortformat. Zur Auswertung der offenen Antworten wurde ein Kodierschema entwickelt, wobei deduktive und induktive Herangehensweisen kombiniert wurden, so dass die Kategorien theoretisch anschlussfähig sind und zugleich die Informationen in den Antworten ausgeschöpft wurden.

Für die insgesamt 14 Aufgaben wurden unterschiedliche Formate bzgl. der Anzahl der richtigen Antwortoptionen und Antwortoptionen pro Aufgabe gewählt. Unterschiedliche Formate stellen bei der Datenanalyse im Partial-Credit-Modell kein Problem dar (Masters 1982), bringen jedoch unterschiedliche Ratewahrscheinlichkeiten bzw. formatbedingte Leichtigkeit/Schwierigkeit mit sich (Lord 1977; Haladyna und Downing 1993; Rodriguez

2005; Kubinger und Gottschall 2007), die im Rahmen der Pilotierung des Instruments noch in Kauf genommen wurden.

In der vorliegenden Videosequenz sehen Sie den Anfang der ersten Sitzung eines Seminars.



Die Vorstellungssituation kann im Hinblick auf ein angemessenes Kommunikations- und Interaktionsgeschehen als problematisch angesehen werden.

Welches könnten hierfür die zwei Hauptgründe sein?

Kreuzen Sie bitte zwei Antworten an. Sie dürfen das Video gerne **mehrfach** anschauen.

Problematisch ist...

- | | |
|--|--------------------------|
| (A) ... die Instruktion der Lehrkraft ganz am Anfang. | <input type="checkbox"/> |
| (B) ... das Verhalten der Lehrkraft gegen Ende der Vorstellungsrunde. | <input type="checkbox"/> |
| (C) ... das statische Stehen der Lehrkraft am Kopf des Raumes. | <input type="checkbox"/> |
| (D) ... die Reihenfolge des Seminareinstiegs: erst Überblick über den Seminaraufbau, dann Vorstellungsrunde. | <input type="checkbox"/> |

Abb. 1 videobasierte Aufgabe der Facette „Wissen über Kommunikation und Interaktion mit Lernenden“

Ein positives Selbstbild sowie eine hohe Lernmotivation haben sich in der empirischen Lehr-Lernforschung als förderlich für Lernprozesse erwiesen.

Welche der nachfolgenden Instrumente und Verfahren würden Sie einsetzen, um Selbstbild und Lernmotivation bei Ihren Lernenden zu erfassen?

Kreuzen Sie bitte in jeder Zeile ein Kästchen an.

	Eher falsch	Eher richtig
(A) Soziogramme (Abbildung sozialer Netzwerke, z.B. über die Sitzordnung der Lernenden oder die Beobachtung von Gesprächskontakten)	<input type="checkbox"/>	<input type="checkbox"/>
(B) Schriftliche Prüfungen	<input type="checkbox"/>	<input type="checkbox"/>
(C) Selbsteinschätzungen	<input type="checkbox"/>	<input type="checkbox"/>
(D) Gruppenarbeiten	<input type="checkbox"/>	<input type="checkbox"/>

Stellen Sie sich vor, Sie behandeln ein Thema, das den Lernenden unterschiedlich stark vertraut ist. Sie möchten die Lehr-Lernziele dem Vorwissensstand der Lernenden anpassen.

(1) Welches Lehr-Lernziel formulieren Sie für Lernende, die mit dem Thema gar nicht vertraut sind?

Antworten Sie bitte in Stichpunkten.

.....

.....

(2) Welches Lehr-Lernziel formulieren Sie für Lernende, die mit dem Thema sehr vertraut sind?

Antworten Sie bitte in Stichpunkten.

.....

.....

Abb. 2 textbasierte Beispielaufgaben für die Facetten „Wissen über Individual- und Lernprozessdiagnostik“ und „Wissen über Lehr- und Lernziele“

6 Welche Unterschiede finden sich in der Struktur sowie Reliabilität des Konstrukts, und zeigen sich differenzielle Item-Funktionen für Lehrkräfte aus Schule und Erwachsenenbildung?

Im Vordergrund steht hier die Frage nach der Struktur und Reliabilität der ausgewählten Facetten des Instruments, da das Vorliegen beider Kriterien für eine präzise Messung bzw. für zulässige Gruppenvergleiche nötig ist. Darauf aufbauend werden Mittelwertsunterschiede und DIF-Analysen berichtet. Gegenstand der nachfolgenden Analysen sind dabei exemplarisch die beiden Facetten „Wissen über Individual- und Lernprozessdiagnostik“ und „Wissen über Lehr- und Lernziele“. Diese Facetten wurden ausgewählt, da sich bei ihnen die Frage nach der Allgemeinheit oder Spezifität auf der Grundlage der Expertenbefragung in einem Fall stellt, im anderen nicht (vgl. Abschn. 2 und 4.3).

6.1 Stichprobe

An der Studie nahmen insgesamt 212 Lehrkräfte aus EB ($n = 146$) und Schule ($n = 66$) teil. Dabei handelt es sich um eine Ad-hoc-Stichprobe, was unproblematisch ist, wenn nicht repräsentative Testleistungen erhoben, sondern – wie hier – die Testaufgaben (erstmalig) an (EB-)Lehrkräften erprobt werden sollen.

Die Lehrkräfte der EB im Alter von 23–83 Jahren ($M = 48,70$; $SD = 12,39$) verfügten zum Erhebungszeitpunkt über eine Lehrerfahrung zwischen 0 und 45 Jahren ($M = 14,36$; $SD = 11,17$) und lehrten pro Woche zwischen 1 und 50 h ($M = 13,53$; $SD = 10,76$). Der Anteil an Akademikern (77 %) sowie die Beschäftigung bei mehreren Bildungseinrichtungen weist die vorliegende Stichprobe als typisch für die EB aus, auch der Anteil an weiblichen Lehrkräften (68 %) ist für den Bereich der öffentlich geförderten EB (z. B. Volkshochschulen) typisch (Langemeyer und Martin 2014), nicht aber für den Bereich der EB insgesamt, in dem das Geschlechterverhältnis ausgewogen ist (vgl. Tab. 1). Die Lehrkräfte in der Stichprobe lehren zu der gesamten Bandbreite an Themen, von der (fachbezogenen) beruflichen Weiterbildung über Angebote der allgemeinen bis zur politischen Weiterbildung (vgl. Abschn. 2).

Die Stichprobe der EB-Lehrkräfte für die Pilotstudie wurde mithilfe von drei großen Volkshochschulen in Baden-Württemberg, Bayern und Nordrhein-Westfalen rekrutiert. Die Teilnahme fand individuell an einem vorbereiteten Laptop statt, war freiwillig und wurde honoriert. Schwierigkeiten bereitete die Rekrutierung der Stichprobe der Schul-Lehrkräfte, da im Beruf stehende Lehrkräfte befragt werden sollten, denen aus dienstrechtlichen Gründen kein Honorar gezahlt werden durfte. Die Mehrzahl der Schul-Lehrkräfte wurde durch die Online-Version des Tests befragt.

Die Lehrkräfte aus dem Schulkontext im Alter von 25–63 Jahren ($M = 45,41$; $SD = 10,60$) lehrten zum Erhebungszeitpunkt zwischen 1,5 und 37 Jahren ($M = 16,42$; $SD = 10,23$), unterrichteten eine große Bandbreite unterschiedlicher Fächer und hatten ein

Stundendeputat von durchschnittlich 21,61 h ($SD = 4,89$). An Gymnasien lehrten 49 Lehrkräfte, 18 an anderen allgemein- bzw. berufsbildenden Schulen. Der Anteil der weiblichen Lehrkräfte (59 %) lag über dem Anteil der männlichen (41 %).

6.2 Datenanalyse

Allen befragten Lehrkräften lagen prinzipiell alle Testaufgaben vor. Um Positionseffekten vorzubeugen, wurden die Testaufgaben in vier Versionen präsentiert. Da innerhalb der EB-Forschung bislang keine Erfahrungen in der Entwicklung von Tests zur Erfassung des PPK vorliegen, war unklar, wie viel (zusätzliche) Zeit die in der Regel pädagogisch weniger als im Schulbereich vorgebildeten Lehrkräfte bei der Bearbeitung der Testaufgaben benötigen. In den Fällen, bei denen die Testzeit der EB-Lehrkräfte weit über den vereinbarten Rahmen von ca. 90 min hinausging, wurde die Möglichkeit gegeben, restliche Testaufgaben zu überspringen (not reached missings) und nur noch die soziodemographischen Angaben auszufüllen; fehlende Werte bei diesen Testaufgaben wurden als solche behandelt. Alle anderen fehlenden Werte (omitted values) wurden entsprechend der bei Leistungstests üblichen – jedoch auch kritisch diskutierten – Praxis (Robitzsch und Rupp 2008) null-„imputiert“, unter der Annahme, dass diese fehlenden Werte entstanden, weil die Personen die Testaufgaben nicht lösen konnten und deswegen nicht bearbeiteten. Die Skalierung der Daten erfolgte auf der Grundlage eines Partial-Credit-Modells (Masters 1982) mit Conquest, Version 2.0 (Wu et al. 2007). In die Analysen gingen insgesamt 14 polytome Testaufgaben ein.

6.3 Überprüfung der Struktur und Reliabilität ausgewählter Facetten des PPK

Um die Struktur zu überprüfen, wurden in einem ersten Schritt an der Gesamtstichprobe der Lehrkräfte aus EB und Schule ($N = 212$) zwei Messmodelle miteinander verglichen: Messmodell 1 nimmt eine homogene Wissensstruktur an, Messmodell 2 unterscheidet zwischen den beiden Bereichen „Wissen über Individual- und Lernprozessdiagnostik“ und „Wissen über Lehr- und Lernziele“ – konform zur vorgeschlagenen Konzeptualisierung. Der Vergleich der Anpassung der beiden Messmodelle an den Datensatz wurde über einen Index aus Likelihood und Parameterzahl vorgenommen (Wu et al. 2007). Die Differenz der beiden Modell-Indizes zeigt eine signifikant bessere Anpassung des zweidimensionalen Modells; zusätzlich wurden die informationstheoretischen Maße AIC und BIC berechnet, die beide ebenfalls auf eine bessere Anpassung des zweidimensionalen Modells an den Datensatz hinweisen (vgl. Tab. 5). Die Testaufgaben zeigen für das zweidimensionale Modell mit *Weighted Mean Square*-Werten (WMNSQ) zwischen 0,89 und 1,07 eine gute Modellanpassung, die Streuung der Aufgabenschwierigkeiten kann in Relation zu den

Personenfähigkeiten als gut bezeichnet werden ($-1,372 \leq \text{Aufgabenschwierigkeit} \leq +2,000$; Abb. 3).

Tab. 5 Statistik zum ein- bzw. zweidimensionalen Modell für die Gesamtstichprobe

Modell	Abweichung	Parameter	Differenz			AIC	BIC
			Abw.	Par.	p		
1-dim.	5456,0	39	15,3	2	$\leq ,05$	5534,0	5562,0
2-dim.	5440,7	41				5522,7	5552,1

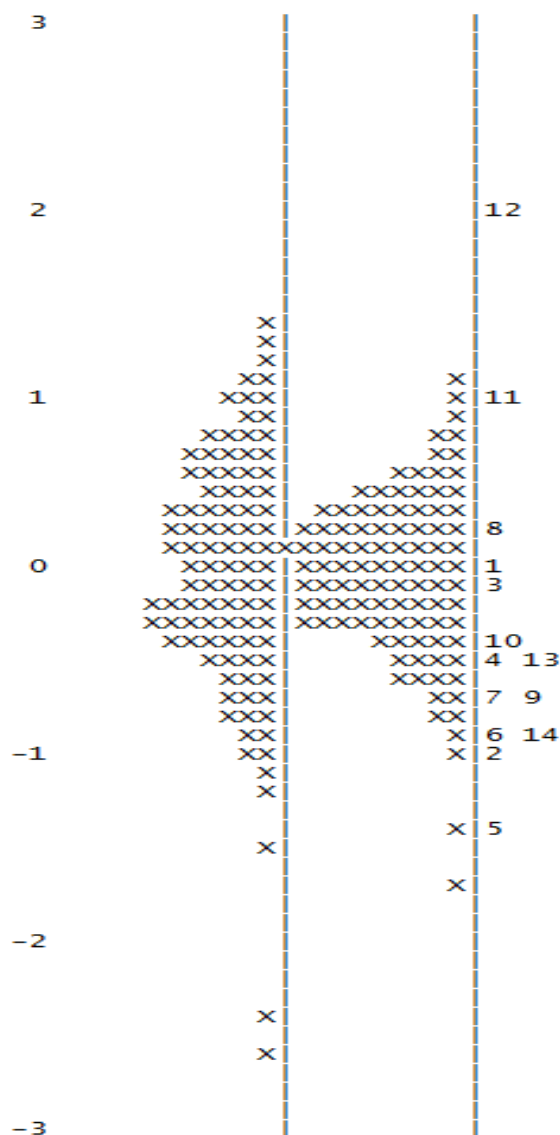
Für die Überprüfung der internen Konsistenz der beiden Subfacetten wurde die *Expected A Posteriori Estimation* (EAP) herangezogen, die eine unverzerrte Beschreibung der Population liefert und die mehrdimensionale Modellstruktur berücksichtigt (Wu et al. 2007). Für die Gesamtstichprobe zeigt das eindimensionale Modell eine akzeptable Reliabilität (0,70). Bei zweidimensionaler Modellierung finden sich Reliabilitäten von 0,70 („Wissen über Individual- und Lernprozessdiagnostik“) und 0,60 („Wissen über Lehr- und Lernziele“). Neben den Kennwerten zur Reliabilität gibt die latente Korrelation der beiden Wissensfacetten einen wichtigen Einblick in die Wissensstruktur, die mit $r = 0,89$ hoch ausfällt.

Eine zentrale Frage ist nun, inwieweit sich die für die Gesamtstichprobe gefundene Wissensstruktur auch in den beiden Teilstichproben EB-Lehrkräfte ($n = 146$) und Schul-Lehrkräfte ($n = 66$) wiederfinden lässt. Die Itemstatistiken zeigen für die Testaufgaben eine gute Modellanpassung, lediglich eine Aufgabe unterschreitet die Grenze des WMNSQ-Wertes von 0,80 bei zweidimensionaler Modellierung der Aufgaben in der Schul-Lehrkräfte-Stichprobe. Für beide Teilstichproben zeigte das zweidimensionale Modell eine bessere Anpassung an die Daten (Tab. 6 und 7). Für die Stichprobe der EB-Lehrkräfte zeigte sich für das eindimensionale Modell eine unbefriedigende Reliabilität (0,58), für die Stichprobe der Schul-Lehrkräfte hingegen eine gute Reliabilität (0,82).

Bei zweidimensionaler Modellierung wird das Wissen für die beiden Gruppen ebenfalls unterschiedlich reliabel erfasst, wobei die Messung bei Schul-Lehrkräften (0,81; 0,70) wesentlich reliabler gelingt als bei den EB-Lehrkräften (0,58; 0,44; Tab. 8).

Die Zusammenhänge zwischen den beiden Wissensbereichen fallen für die EB-Lehrkräfte mit $r = 0,83$ und die Schul-Lehrkräfte mit $r = 0,88$ ähnlich hoch aus wie für die Gesamtstichprobe.

Abb. 3 Personenfähigkeiten und Aufgabenschwierigkeiten bei zweidimensionaler Modellierung für die Gesamtstichprobe (Aufgaben 1-9: Wissen über Individual- und Lernprozessdiagnostik; Aufgaben 10-14: Wissen über Lehr- und Lernziele)



Werden Konstrukte mit Hilfe von IRT-Modellen in mehreren Gruppen skaliert, so stellt sich die Frage nach differentiellen Itemfunktionen (DIF). Dazu wurden die Aufgaben mit der in Conquest 2.0 (Wu et al. 2007) implementierten Routine untersucht. Erkennbar werden solche Disparitäten nicht an unterschiedlichen Gruppenmittelwerten, da prinzipiell denkbar ist, dass für eine Gruppe alle Aufgaben im selben Ausmaß schwerer waren, sondern an Verschiebungen der Differenzen von Aufgabenschwierigkeiten zwischen den Gruppen. DIF-Effekte liegen vor, wenn für eine Aufgabe die bedingte Lösungswahrscheinlichkeit – bezogen auf die umfassende, auf allen Aufgaben beruhende Fähigkeitsschätzung – zwischen den Gruppen variiert (Baumert et al. 2000).

Tab. 6 Statistik zu den ein- und zweidimensionalen Modellen für die Stichprobe der EB-Lehrkräfte

Modell	Abweichung	Parameter	Differenz			AIC	BIC
			Abw.	Par.	p		
1-dim.	3607,7	39	7,6	2	$\leq ,05$	3685,7	3713,7
2-dim.	3600,1	41				3682,1	3711,5

Tab. 7 Statistik zu den ein- und zweidimensionalen Modellen für die Stichprobe der Schul-Lehrkräfte

Modell	Abweichung	Parameter	Differenz			AIC	BIC
			Abw.	Par.	p		
1-dim.	1716,8	38	6	2	$\leq ,05$	1792,8	1820,1
2-dim.	1710,8	40 ^a				1790,8	1819,5

^aAufgrund der geringen Größe der Schul-Lehrkräfte-Stichprobe und den damit verbundenen fehlenden Werten für schulische Lehrkräfte wurden bei Item 12 im Partial-Credit-Model weniger Zwischenstufen geschätzt und daher auch jeweils ein freier Parameter weniger.

Tab. 8 Interne Konsistenzen des zweidimensionalen Modells für die verschiedenen Stichproben

	Wissen über Individual- und Lernprozessdiagnostik			Wissen über Lehr- und Lernziele		
	Gesamt	EB	Schule	Gesamt	EB	Schule
EAP-Reliabilität	0,70	0,58	0,81	0,60	0,44	0,70

Für die Facette „Wissen über Individual- und Lernprozessdiagnostik“ zeigte sich, dass die Schul-Lehrkräfte bei einem Standardfehler von 0,023 um 0,124 Logits signifikant besser abschnitten als die EB-Lehrkräfte, was jedoch noch kein Hinweis auf etwaige DIF-Effekte ist (Baumert et al. 2000; Wu et al. 2007). Die Chi-Quadrat-Statistik zur Prüfung der Frage, ob einzelne Aufgaben für eine der beiden Gruppen systematisch schwerer oder leichter waren, deutet eher nicht auf DIF-Effekte hin, $\chi^2(8) = 11,35$; $p = 0,18$. Für eine Aufgabe dieser Facette zeigte sich dennoch bei einem Standardfehler von 0,049, dass diese für Schul-Lehrkräfte um 0,115 Logits signifikant schwieriger war als für die EB-Lehrkräfte. Nach Santelices und Wilson (2012) handelt es sich hierbei jedoch um einen vernachlässigbaren DIF-Effekt.

Für die Facette „Wissen über Lehr- und Lernziele“ zeigte sich, dass die Schul-Lehrkräfte um 0,464 Logits signifikant besser abschnitten als die EB-Lehrkräfte. Die Chi-Quadrat-

Statistik zur Prüfung der Frage, ob einzelne Aufgaben für eine der beiden Gruppen systematisch schwerer oder leichter waren, deutet auf DIF-Effekte hin, $\chi^2(4) = 100,63$; $p = 0,00$. Dabei zeigten sich für zwei Aufgaben für die EB- Lehrkräfte signifikante DIF-Effekte bei einem Standardfehler von 0,057 und 0,297 Logits sowie bei einem Standardfehler von 0,055 und 0,297 Logits. Zwei weitere Aufgaben zeigten signifikante DIF-Effekte für die Schul-Lehrkräfte bei einem Standardfehler von 0,046 bei 0,224 Logits und bei einem Standardfehler von 0,057 bei 0,259 Logits. Diese vier DIF-Effekte sind als mittelstark einzuschätzen (Santelices und Wilson 2012) und gemeinsam mit den weiteren vorgestellten Befunden nachfolgend Gegenstand der Diskussion.

7 Zusammenfassung und Diskussion der Ergebnisse

Dieser Beitrag widmete sich der bislang konzeptionell und empirisch unbeantworteten Frage nach dem Grad der Allgemeinheit oder Spezifität des PPK *für* Lehrkräfte aus Schule und EB auf der Basis einer neu entwickelten PPK-Konzeptualisierung und einer Erprobung des Instruments zur Erfassung des PPK.

Eine Literaturstudie beantwortete die erste Forschungsfrage 1a, indem sie aufzeigte, dass in unterschiedlichen Forschungsdiskursen, die sich mit PPK befassen, dieses in unterschiedlichem Ausmaß explizit oder implizit als bildungsbereichsübergreifend konzeptualisiert wird: in der EB-Literatur geschieht dies in Teilen und auch explizit, in der Literatur zur Allgemeinen Didaktik wird – trotz ihres selbstformulierten Allgemeinheitsanspruchs – der Fokus auf schulische Lehr-Lernsituationen gelegt; ähnliches gilt für die pädagogische Psychologie und die empirische Lehr-Lernforschung. Die Frage, ob PPK fachübergreifend zu verstehen ist, wird in keinem der Diskurse angezweifelt. Des Weiteren zeigte die Literaturstudie, dass sowohl die ausschließlich auf Schule bezogenen Diskurse als auch die EB-Literatur von gleichen Anforderungen an Lehrkräfte ausgehen – sofern diese Anforderungen sich auf das unmittelbare Lehr-Lerngeschehen im Seminar- oder Klassenraum beziehen. Angesichts dieser Diskussionslage betrachten wir es als gerechtfertigt, das PPK fach- und bildungsbereichsübergreifend zu konzeptualisieren mit den in Abschn. 4 dargestellten inhaltlichen Facetten (Forschungsfrage 1b). Eine Expertenbefragung bescheinigte der auf dieser Grundlage entwickelten Konzeptualisierung bis hinunter auf die Ebene der Subfacetten weitestgehend Vollständigkeit, bildungsbereichs- und fachübergreifende Relevanz und Allgemeingültigkeit (Forschungsfrage 1c). Dieser Befund bestätigte uns darin, ein Testinstrument zur Erfassung des PPK von Lehrkräften aus EB und Schule zu entwickeln.

Den Schwerpunkt der Darstellung bilden die Befunde aus einer Pilotstudie, in deren Rahmen die Operationalisierung des gesamten PPK an einer Stichprobe aus EB- und Schul-Lehrkräften überprüft wurde. Zwei Facetten des PPK, „Wissen über Individual- und

Lernprozessdiagnostik“ und „Wissen über Lehr- und Lernziele“, wurden auf Struktur, Reliabilität und differenzielle Item-Funktionen hin überprüft, um erste Hinweise auf die Allgemeinheit bzw. die Spezifität des PPK zu erhalten.

Bezüglich der Struktur der ausgewählten Aufgaben zeigte sich für alle drei Stichproben – die Gesamt-, die EB- und die Schul-Stichprobe – gleichermaßen, dass ein zweidimensionales Modell, das die differenzierte Wissensstruktur widerspiegelt, einem eindimensionalen Modell überlegen ist. Es scheint sinnvoll(er) zu sein, konzeptionell und empirisch zwischen diesen beiden Wissensbereichen zu unterscheiden, auch wenn sich jeweils für alle drei Stichproben ähnlich hohe Zusammenhänge zwischen den latenten Faktoren zeigten; dass sich die hohen Zusammenhänge zwischen den latenten Faktoren über *alle* drei Stichproben hinweg zeigten, spricht zumindest nicht gegen die Allgemeinheit einer dieser Facetten. Weiterhin erscheinen diese Zusammenhänge plausibel, da Wissen über Lernprozess- und Individualdiagnostik dazu dient, zwischen Leistungs- bzw. Wissensständen zu differenzieren und Wissen über Lernziele die Anpassung verschiedener Lernziele an unterschiedliche Leistungs- bzw. Wissensstände ermöglicht. Eine offene Frage ist, welche Struktur des PPK sich unter Berücksichtigung der weiteren Testaufgaben der anderen Facetten zeigt; diese Frage wird anhand größerer Stichproben in einer derzeit durchgeführten Hauptstudie beantwortbar werden.

Für die Gesamtstichprobe der Pilotstudie liegen die Reliabilitäten für beide Facetten in einem akzeptablen bis befriedigenden Bereich. Für die beiden Teilgruppen unterscheiden sich diese teilweise deutlich: für die Schul-Lehrkräfte liegen die Reliabilitäten für beide Modellierungen in einem akzeptablen bis guten Bereich, für die EB-Lehrkräfte zeigten sich hier deutlich geringere Reliabilitäten, für die zweidimensionale Modellierung liegen diese in einem nicht mehr akzeptablen Bereich. Diese Befunde deuten auf einen erwartungsgemäß homogenen Wissensstand bei den Schul-Lehrkräften verglichen mit den EB-Lehrkräften hin, der durch die Differenz der Ausbildungen und Prüfungen (Staatsexamina) begünstigt sein könnte. Für die Weiterentwicklung des Tests und seinen Einsatz in der Hauptstudie ziehen wir daher den Schluss, die Anzahl der (für die Hauptstudie überarbeiteten bzw. selektierten) Aufgaben zu erhöhen, um die Reliabilität der Messungen zu stärken, sowie eine breite Streuung an Aufgabenschwierigkeiten anzustreben.

Für eine Aufgabe der Facette „Wissen über Individual- und Lernprozessdiagnostik“ fand sich ein schwacher DIF-Effekt zu Ungunsten der Schul-Lehrkräfte. Dies ist mit Blick auf die Ergebnisse der Expertenbefragung eher überraschend, da diese darauf hindeuteten, dass Aufgaben dieser Facette ggf. für EB-Lehrkräfte schwieriger zu beantworten sind, da dieses Wissen ggf. kein allgemeines, täglich geübtes ist; für das Ausbleiben von DIF-Effekten bei den restlichen Aufgaben dieser Facette wiederum spricht die Literaturstudie sowie die letztlich unumgängliche Notwendigkeit, im Rahmen von Lehr-Lernsituationen aller

Bildungsbereiche den aktuellen Lernstand zu diagnostizieren. Zur Bearbeitung der Aufgabe mit DIF-Effekt mussten Situationen auf Beurteilungsfehler der Lehrkraft bewertet werden. Bei der Analyse der einzelnen Antwortoptionen erscheint dabei einer der Distraktoren ausschlaggebend für die relative Benachteiligung der Schul-Lehrkräfte, der sich auf die Orientierung an den Leistungen der gesamten Lerngruppe bei der Bewertung der Lernenden bezieht. Ggf. wurde diese in der Schule bei der Notenvergabe gängige Praxis von den Schul-Lehrkräften fälschlicherweise als Beurteilungsfehler gewertet, da eine zu starke Orientierung an einer sozialen Bezugsnorm in verschiedenen pädagogischen Zusammenhängen mitunter kritisch gesehen wird. Mit Blick auf die Frage nach der Allgemeinheit bzw. der Spezifität des PPK lässt sich sagen, dass zumindest auf dieser Datengrundlage keine starken Argumente für die Ablehnung der Annahme der Allgemeinheit dieser Facette finden lassen; vielmehr ist es offenbar gelungen, Aufgaben zu entwickeln, die Phänomene erfassen, die sowohl in der EB als auch in Schulkontexten auftreten. Dieses Ergebnis ist jedoch unter dem starken Vorbehalt zu sehen, dass die Daten der Schul-Lehrkräfte auf einer sehr kleinen Stichprobe beruhen, was auch für alle weiteren Analysen einschränkend berücksichtigt werden muss.

Für die Aufgaben der Facette „Wissen über Lehr- und Lernziele“ zeigten sich für vier von fünf Aufgaben mittelstarke DIF-Effekte, jeweils für zwei Aufgaben pro Gruppe. Für die EB-Lehrkräfte zeigten sich DIF-Effekte für eine Aufgabe, bei der Lehr-Lernziele auf verschiedenen Niveaustufen frei formuliert werden mussten, und für eine Aufgabe, bei der eine Aussage getroffen werden musste, ob bzw. wann Lehrziele den Lernenden kommuniziert werden sollten. Für die Schul-Lehrkräfte zeigten sich DIF-Effekte bei einer Auswahlaufgabe, bei der korrekt formulierte Lehr- und Lernziele identifiziert werden mussten, sowie bei einer Aufgabe, bei der Lehr-Lernziel-Schwierigkeitsstufen in die richtige Reihenfolge gebracht werden mussten. Die Aufgaben mit DIF-Effekten für die Schul-Lehrkräfte sind insgesamt relativ leicht (Aufgabe 10 und 13 in der Wright-Map in Abb. 3), weswegen die Unterschiede zwischen den Personen(gruppen) bei diesen Aufgaben nicht so groß sind wie für die sehr schweren Aufgaben 11 und 12. Die Leichtigkeit der betreffenden Aufgaben generiert damit DIF-Effekte zu Ungunsten der Schul-Lehrkräfte, denn ein hoher Anteil der Testteilnehmenden – unabhängig von der Gruppenzugehörigkeit – konnte diese Aufgaben lösen, weswegen diese Aufgaben den „wahren“ Gruppenunterschied der gesamten Skala unterschätzen. Die DIF-Effekte in der Gruppe der EB-Lehrkräfte lassen sich zum einen durch die höhere Aufgabenschwierigkeit begründen, zum anderen könnte in Bezug auf den Aufgabeninhalt das freie Formulieren von Lernzielen für EB-Lehrkräfte ungewohnt sein.

Ob dieser Bereich nun auch genuines EB-Wissen ist (vom Anforderungsprofil an eine EB-Lehrkraft sein müsste bzw. normativ gedacht: sein sollte), wie die Expertenbefragung

zeigte, bleibt weiteren Untersuchungen vorbehalten bzw. kann in der derzeit durchgeführten Hauptstudie mit einer größeren Stichprobenzahl erneut überprüft werden. Jedoch ist aus (nicht nur schul-!) theoretischer Sicht sowie aus Sicht der Experten uneingeschränkt davon auszugehen, dass die Formulierung von Lernzielen auch für Lehr-Lernsituationen in der EB grundlegend ist.

Im Rahmen des vorgestellten Projekts sind weitere Studien geplant, die, dann unter Einbezug des überarbeiteten Instruments, die Erfassung des gesamten PPK vornehmen und der Frage nach seiner Struktur für Lehrkräfte aus EB und Schule an einer größeren (Gesamt-)Stichprobe nachgehen, wodurch weitere Hinweise auf die Allgemeinheit oder Spezifität des PPK gewonnen werden. Darüber hinaus soll das Testinstrument validiert werden, indem der Zusammenhang zwischen der durch Teilnehmerinnen und Teilnehmer bzw. Schülerinnen und Schüler wahrgenommenen Qualität von Lehr-Lernsituationen und dem PPK der Lehrkräfte untersucht wird. Des Weiteren wird erforscht, ob sich das allgemeine PPK von pädagogischen Überzeugungen diskriminieren lässt bzw. in welchem Zusammenhang es mit diesen steht. Starke Argumente für die Frage nach der Allgemeinheit oder Spezifität des PPK, also ob ggf. einzelne Facetten des PPK in einem der beiden Bildungsbereiche zusätzlich nötig, differenzierter erforderlich oder aber verzichtbar sind, liefern jedoch erst quasi-experimentelle Designs mit Extremgruppen (niedriges Wissen vs. hohes Wissen in den verschiedenen Facetten) von Lehrkräften verschiedener Bildungsbereiche bei Untersuchung des Effekts des PPK von Lehrkräften auf die wahrgenommene Unterrichtsqualität bzw. den Lernzuwachs von Lernenden in verschiedenen Bildungsbereichen. Hierfür ist jedoch ein reliables und valides Testinstrument nötig, das im vorgestellten Projekt entwickelt wird. Erst auf dieser Grundlage lassen sich die in der Literatur mit Comenius beginnenden, teils implizit, teils explizit formulierten Annahmen zur Allgemeinheit oder Spezifität des PPK empirisch prüfen und Beiträge zur von Terhart (2013) formulierten Aufgabe leisten, einen allgemeindidaktischen Ansatz zu entwickeln, der „[...] tatsächlich allgemein genug ist, um die unterschiedlichsten Felder des Lehrens und Lernens inklusive ihrer Rahmenstrukturen mit abzudecken und dabei sowohl Allgemeines als auch je Besonderes im Rahmen eines theoretischen Konzepts zu beschreiben [...]“ (Terhart 2013, S. 226). Ob sich daraus auch wissenschafts- und hochschulpolitische Debatten ergeben, bleibt abzuwarten.

Darüber hinaus bietet der Test in seiner finalen Form durch die Kombination video- und textbasierter Aufgaben die Möglichkeit, der Frage nach den häufig angenommenen Stärken videobasierter Aufgaben weiter systematisch nachzugehen. Videoaufgaben – jenseits von „klassisch deklarativen Wissensfragen“ und gerade, wenn man sie mit offenen Antwortformaten kombiniert – sollten besonders aussagekräftige Prädiktoren für den Unterrichtserfolg sein, da sie keine Relevanzstrukturen durch textbasierte

Situationsdarstellungen vorgeben. Im vorgestellten Projekt können die video- und textbasierten Aufgaben systematisch im Hinblick auf ihre psychometrischen Eigenschaften hin verglichen werden. Zudem adressieren einige der Aufgaben im Text- vs. Videoformat dieselben inhaltlichen Phänomene. Dies erlaubt zu prüfen, ob inhaltsunabhängig systematisch bestimmte Gruppen von Lehrkräften durch den Einsatz von videobasierten Aufgaben bevorzugt oder benachteiligt werden. Ebenso ist offen, welche Rolle es spielt, aus welchem Kontext (z. B. Bildungsbereich, Fach, Altersklasse) die Videos stammen. Videos aus Kontexten, die dem Kontext der jeweiligen Lehrkraft, die den Test bearbeitet, nahe sind, gehen vermutlich mit einem höheren Grad an subjektiver Identifikation mit dem Gesehenen einher. Welche Rolle dies für das Ausmaß an aufgabenbezogener Motivation, kognitiver Elaboration und das Ausmaß an korrekten Antworten spielt, ist bislang ungeklärt.

Das vorgestellte Projekt mit dem entwickelten Test, der verschiedene Aufgabenformate kombiniert, und dem Einsatz des Tests in verschiedenen Bildungsbereichen birgt somit das Potenzial, in weiterführenden Analysen hilfreiche Beiträge zur videobasierten Lehr-Lernforschung zu leisten.

8 Anhang

Tab. 9 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Lernprozesse von Lernenden“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens</i> <i>V</i>	<i>V</i> <i>critical</i>	<i>N</i>
a Wissen über kognitive Prozesse	EB	0	2	7	9 ^a	3,39	0,70	0,80	0,67	18
	SK	0	2	4	19 ^a	3,68	0,63	0,89	0,64	25
b Wissen über motivationale Prozesse	EB	0	1	4	13 ^a	3,67	0,59	0,89	0,67	18
	SK	0	1	3	21 ^a	3,80	0,50	0,93	0,64	25
c Wissen über Emotionen von Lernern	EB	0	2	3	12 ^a	3,59	0,71	0,86	0,67	17
	SK	0	2	8	15 ^a	3,52	0,65	0,84	0,64	25
d Wissen über Lerntheorien und deren Bedeutung für die Gestaltung von Lehr-Lern-Prozessen	EB	0	2	6	10 ^a	3,44	0,71	0,81	0,67	18
	SK	0	0	9	16 ^a	3,64	0,49	0,88	0,64	25

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext

^a Modus

Tab. 10 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Heterogenität der Lernenden“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens V</i>	<i>V critical</i>	<i>N</i>
a psychische Auffälligkeiten	EB	1	3	8	6 ^a	3,06	0,87	0,69	0,67	12
	SK	1	5	11 ^a	7	3,00	0,83	0,67	0,64	24
b Heterogenität hinsichtlich des Vorwissens und der Intelligenz	EB	0	2	5	11 ^a	3,50	0,71	0,83	0,67	18
	SK	0	1	7	17 ^a	3,64	0,57	0,88	0,64	25
c verschiedene Lernstile und Lerntypen	EB	0	1	8	9 ^a	3,44	0,62	0,81	0,67	18
	SK	0	5	13 ^a	7	3,08	0,70	0,69	0,64	25
d Unterschiede zwischen den Geschlechtern	EB	0	2	10 ^a	6	3,22	0,65	0,74	0,67	18
	SK	0	7	15 ^a	3	2,84	0,62	0,61	0,64	25
e kulturelle Besonderheiten, <i>religiöse Orientierungen^b, Einstellungen von Lernenden^b</i> sowie Migrationshintergrund und sozio- ökonomischen Status	EB	0	1	7	10 ^a	3,50	0,62	0,83	0,67	18
	SK	0	2	16 ^a	7	3,20	0,58	0,73	0,64	25
f Die Bedeutung unterschiedlicher Altersstufen der Lernenden und ihrer Lernbiographien (entwicklungspsycho- logisches Wissen)	EB	0	3	5	10 ^a	3,39	0,78	0,80	0,67	18
	SK	1	4	8	12 ^a	3,24	0,88	0,75	0,64	25
g besonderer Förderbedarf	EB	0	5	7 ^a	6	3,06	0,80	0,69	0,67	18
	SK	0	4	14 ^a	7	3,12	0,67	0,71	0,64	25

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext

^a Modus

^b Ergänzung aus der Expertenbefragung

Tab. 11 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Lehr-Lernmethoden und -Konzepte“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens</i> <i>V</i>	<i>V</i> <i>critical</i>	<i>N</i>
a Arbeits- und Sozialformen sowie deren Kombination und zieladäquaten und effektiven Einsatz in Lehr-Lernsituationen	EB	0	2	1	15 ^a	3,72	0,67	0,91	0,67	18
	SK	0	0	3	22 ^a	3,88	0,33	0,96	0,64	25
b Konzepte zu individualisierten, kooperativen oder offenen Formen des Lernangebots sowie deren Umsetzung	EB	0	2	1	14 ^a	3,71	0,69	0,90	0,67	17
	SK	0	0	4	21 ^a	3,84	0,37	0,95	0,64	25
c Wissen über Kriterien der Unterrichtsqualität ^b		-	-	-	-	-	-	-	-	-

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext

^a Modus

^b Ergänzung aus der Expertenbefragung

Tab. 12 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Führung von Lerngruppen“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens</i> <i>V</i>	<i>V</i> <i>critical</i>	<i>N</i>
a Monitoring/Allgegenwertigkeit in Lehr-Lernsituationen	EB	0	3	6	9 ^a	3,33	0,77	0,78	0,67	18
	SK	0	5	5	14 ^a	3,38	0,82	0,79	0,64	24
b Strategien zur Prävention von und Intervention bei Störungen	EB	0	2	5	11 ^a	3,50	0,71	0,83	0,67	18
	SK	0	1	2	22 ^a	3,84	0,47	0,95	0,64	25
c die Rolle und Etablierung von Regeln und Prozeduren	EB	0	2	7	9 ^a	3,39	0,70	0,80	0,67	18
	SK	0	0	6	19 ^a	3,76	0,44	0,92	0,64	25
d effektive Nutzung der Lernzeit/ Zeitmanagement	EB	0	2	5	11 ^a	3,50	0,71	0,83	0,67	18
	SK	0	2	7	15 ^a	3,54	0,66	0,85	0,64	24

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext

^a Modus

Tab. 13 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Interaktion und Kommunikation mit den Lernenden“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens V</i>	<i>V critical</i>	<i>N</i>
a Herstellung konstruktiver Dialoge und Diskussionen sowie Herstellung eines positiven Lernklimas	EB	0	0	3	15 ^a	3,83	0,38	0,94	0,67	18
	SK	0	0	3	22 ^a	3,88	0,33	0,96	0,64	25
b Gruppendynamiken	EB	0	2	1	15 ^a	3,72	0,67	0,91	0,67	18
	SK	0	0	7	18 ^a	3,72	0,46	0,91	0,64	25
c Gestaltung von Rückmeldungen	EB	0	0	1	17 ^a	3,94	0,24	0,98	0,67	18
	SK	0	1	4	20 ^a	3,76	0,52	0,92	0,64	25
d non-verbale Kommunikation	EB	0	1	6	11 ^a	3,56	0,62	0,85	0,67	18
	SK	0	2	12 ^a	11	3,36	0,64	0,79	0,64	25

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext
^a Modus

Tab. 14 Experteneinschätzungen zu den Subfacetten der Facette „Wissen über Gestaltung von Lernumgebungen“ auf die Frage, ob das in den Facetten und Subfacetten thematisierte PPK in ihrem jeweiligen Bildungsbereich relevant ist

		Trifft gar nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft völlig zu	<i>M</i>	<i>SD</i>	<i>Aikens</i> <i>V</i>	<i>V</i> <i>critical</i>	<i>N</i>
a die physische/räumliche Anordnung der Lernumgebung (indoor/outdoor, Tafeln, Flipchart, Tische, Stühle...)	EB	0	0	8	10 ^a	3,56	0,51	0,85	0,67	18
	SK	0	4	15 ^a	6	3,08	0,64	0,69	0,64	25
b die Gestaltung von Lernmaterialien	EB	0	0	3	15 ^a	3,83	0,38	0,94	0,67	18
	SK	0	1	8	16 ^a	3,60	0,58	0,87	0,64	25
c die Verwendung von Flipchart, Laptop, Tafel etc. (mediale Gestaltung)	EB	0	1	2	15 ^a	3,78	0,55	0,93	0,67	18
	SK	0	2	14	9 ^a	3,28	0,61	0,76	0,64	25

M Mittelwert, *SD* Standardabweichung, *N* Stichprobengröße, *EB* Experten für die Erwachsenenbildung; *SK* Experten für den Schulkontext
^a Modus

Literatur

- Aiken, L.R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142.
- Baumert, J. (2002). Deutschland im internationalen Bildungsvergleich. In N. Killius (Hrsg.), *Edition Suhrkamp: Die Zukunft der Bildung* (S. 100–150). Frankfurt am Main: Suhrkamp.
- Baumert, J., & Kunter, M. (2011a). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Baumert, J., & Kunter, M. (2011b). Das mathematikspezifische Wissen von Lehrkräften, kognitive Aktivierung im Unterricht und Lernfortschritte von Schülerinnen und Schülern. In J. Baumert, M. Kunter, W. Blum, U. Klusmann, S. Krauss, M. Neubrand, T. Kleickmann, D. Richter, K. Löwen, M. Brunner, T. Voss, A. Jordan, M. Besser, J. Elsner, Y. Anders, & A. Hachfeld (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 163–192). Münster: Waxmann.
- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.) (2000). *Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 1*. Opladen: Leske + Budrich.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P., Niedermann, R., Rogalla, M., & Vogt, F. (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010a). *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010b). *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich*. Münster: Waxmann.
- Brokmann-Nooren, C., Grieb, I., & Raapke, H.-D. (Hrsg.) (1994). *Handreichungen für die nebenberufliche Qualifizierung (NQ) in der Erwachsenenbildung*. Bonn: Bundeszentrale für Politische Bildung.

Bromme, R. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens.*

Bern: Hans Huber.

Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F.E. Weinert, N. Birbaumer, & C.F. Graumann (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 177–212). Göttingen: Hogrefe.

Cohen, J. (1968). Weighted kappa. Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34(1), 12–25.

Deutsches Institut für Erwachsenenbildung, Bundesinstitut für Berufsbildung, & Universität Duisburg-Essen (2016). *Personal in der Weiterbildung: Beschäftigungsverhältnisse und Tätigkeiten – wb-personalmonitor*. Bielefeld: Bertelsmann.

Fend, H. (2009). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen* (2. Aufl.). Wiesbaden: Springer VS.

Goeze, A., Hetfleisch, P., & Schrader, J. (2013). Wirkungen des Lernens mit Videofällen bei Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 16(1), 79–113.

Gold, B., & Holodynski, M. (2015). Development and construct validation of a situational judgment test of strategic knowledge of classroom management in elementary schools. *Educational Assessment*, 20(3), 226–248.

Grell, J., & Grell, M. (2010). *Unterrichtsrezepte*. Weinheim: Beltz.

Haladyna, T.M., & Downing, S.M. (1993). How many options for a multiple choice test item. *Educational and Psychological Measurement*, 53(3), 999–1010.

Hellermann, C., Gold, B., & Holodynski, M. (2015). Förderung von Klassenführungsfähigkeiten im Lehramtsstudium. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 97–109.

Hoffmann, N. (2011). Ratgeber-Didaktik in Didaktik-Ratgebern. Portrait eines Selbstlernmediums aus erwachsenenpädagogischer Perspektive. *Der pädagogische Blick*, 19(1), 4–12.

Jüttner, M., Boone, W., Park, S., & Neuhaus, B.J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, 25(1), 45–67.

- Kanning, U.P., Grewe, K., Hollenberg, S., & Hadouche, M. (2006). From the subjects' point of view: Reactions to different types of situational judgement items. *European Journal of Psychological Assessment, 22*(3), 168–176.
- Klafki, W. (1996). *Studien zur Bildungstheorie und Didaktik* (5. Aufl.). Weinheim: Beltz.
- König, J., & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 12*(3), 499–527.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education, 38*, 76–88.
- Körper, K., Kuhlenkamp, D., Peters, R., Schlutz, E., Schrader, J., & Wilckhaus, F. (1995). *Das Weiterbildungsangebot im Lande Bremen. Strukturen und Entwicklungen in einer städtischen Region*. Bremen: Universitätsbuchhandlung.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716–725.
- Kron, F.W. (2014). *Grundwissen Didaktik* (6. Aufl.). München: UTB.
- Kubinger, K.D., & Gottschall, C. (2007). Item difficulty of multiple choice tests dependent on different item response formats – An experiment in fundamental research on psychological assessment. *Psychology Science, 49*(4), 361–374.
- Kunina-Habenicht, O., Lohse-Bossenz, H., Kunter, M., Dicke, T., Förster, D., Gößling, J., Schulze-Stocker, F., Schmeck, A., Baumert, J., Leutner, D., & Terhart, E. (2012). Welche bildungswissenschaftlichen Inhalte sind wichtig in der Lehrerbildung? *Zeitschrift für Erziehungswissenschaft, 15*(4), 649–682.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kunter, M., Kunina-Habenicht, O., Baumert, J., Dicke, T., Holzberger, D., Lohse-Bossenz, H., Leutner, D., Schulze-Stocker, F., & Terhart, E. (2017). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung – Ergebnisse des Projekts BilWiss. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals*. Wiesbaden: Springer VS.
- Langemeyer, I., & Martin, A. (2014). Das Personal in der Erwachsenenbildung. Gemeinsamkeiten und Unterschiede zum Bildungs- und Dienstleistungssektor. *Hessische Blätter für Volksbildung, 63*(1), 17–28.

- Lauermann, F., & König, J. (2016). Teachers' professional competence and wellbeing. Understanding the links between general pedagogical knowledge, self-efficacy and burnout. *Journal of Research in Science Teaching*, 45, 9–19.
- Lenke, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., Liepertz, S., & Leutner, D. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht. *Zeitschrift für Erziehungswissenschaft*, 19(1), 211–233.
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personal Review*, 37(4), 426–441.
- Linninger, C., Kunina-Habenicht, O., Emmenlauer, S., Dicke, T., Schulze-Stocker, F., Leutner, D., Seidel, T., Terhart, E., & Kunter, M. (2015). Assessing teachers' educational knowledge. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 72–83.
- Lord, F.M. (1977). Optimal number of choices per item – A comparison of four approaches. *Journal of Educational Measurement*, 14(1), 33–38.
- Lüders, M. (2014). Erziehungswissenschaftliche Theorien. *Zeitschrift für Pädagogik*, 60(6), 832–849.
- Marx, C., Goeze, A., & Schrader, J. (2014). Pädagogisch-psychologisches Wissen zur Gestaltung von Lehr-Lernsituationen: (Wie) unterscheidet es sich in Erwachsenenbildung/Weiterbildung und Schule? *Hessische Blätter für Volksbildung*, 64(3), 238–251.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- Mayring, P. (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (10. Aufl.). Weinheim: Beltz.
- Meyer, J.W. (2005). *Weltkultur. Wie die westlichen Prinzipien die Welt durchdringen*. Frankfurt am Main: Suhrkamp.
- Riese, J., & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. Empirische Hinweise für eine Verbesserung des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft*, 15(1), 111–143.
- Robitzsch, A., & Rupp, A.A. (2008). Impact of missing data on the detection of differential item functioning. The case of Mantel-Haenszel and logistic regression analysis. *Educational and Psychological Measurement*, 69(1), 18–34.
- Rodriguez, M.C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issue and Practice*, 24(2), 3–13.

- Santelices, M.V., & Wilson, M. (2012). On the relationship between differential item functioning and item difficulty. An issue of methods? Item response theory approach to differential item functioning. *Educational and Psychological Measurement*, 72(1), 5–36.
- Schrader, J. (2003). Wissensformen in der Weiterbildung. In W. Gieseke (Hrsg.), *Institutionelle Innensichten der Weiterbildung* (S. 228–253). Bielefeld: Bertelsmann.
- Schrader, J. (2010). Mediengestützte Fallarbeit. Grundlagen und Zielsetzungen eines Forschungs- und Entwicklungsprojekts zur Kompetenzentwicklung von Lehrenden in der Erwachsenenbildung. In J. Schrader, R. Hohmann, & S. Hartz (Hrsg.), *Mediengestützte Fallarbeit. Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern* (S. 71–100). Bielefeld: Bertelsmann.
- Schrader, J. (2011). *Struktur und Wandel in der Weiterbildung*. Bielefeld: Bertelsmann.
- Schrader, J., Hohmann, R., & Hartz, S. (Hrsg.) (2010). *Mediengestützte Fallarbeit. Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern*. Bielefeld: Bertelsmann.
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik*, 60(6), 850–866.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). OBSERVE – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Leutner, & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*. 56. Beiheft der Zeitschrift für Pädagogik (S. 296–307). Weinheim: Beltz.
- Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4–14.
- Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review*, 57, 1–22.
- Siebert, H. (2006). *Didaktisches Handeln in der Erwachsenenbildung. Didaktik aus konstruktivistischer Sicht* (5. Aufl.). Augsburg: ZIEL.
- Statistisches Bundesamt (2015a). Bildung und Kultur: Allgemeinbildende Schulen. Schuljahr 2014/2015. https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100157004.pdf?__blob=publicationFile.
Zugegriffen: 10. Mai 2016.

- Statistisches Bundesamt (2015b). Bildung und Kultur: Berufliche Schulen. Schuljahr 2014/2015. https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/BeruflicheSchulen2110200127004.pdf?__blob=publicationFile. Zugegriffen: 10. Mai 2016.
- Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates. *Zeitschrift für Psychologie*, 223(1), 54–63.
- Syring, M., Bohl, T., Kleinknecht, M., Kuntze, S., Rehm, M., & Schneider, J. (2015). Videos oder Texte in der Lehrerbildung? Effekte unterschiedlicher Medien auf kognitive Belastung und die motivational-emotionalen Prozesse beim Lernen mit Fällen. *Zeitschrift für Erziehungswissenschaft*, 18(4), 667–685.
- Tenorth, H.-E. (1994). *Alle alles zu lehren. Möglichkeiten und Perspektiven allgemeiner Bildung*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Terhart, E. (2002). Fremde Schwestern. Zum Verhältnis von Allgemeiner Didaktik und empirischer Lehr-Lern-Forschung. *Zeitschrift für Pädagogische Psychologie*, 16(2), 77–86.
- Terhart, E. (2013). Neuansätze in der Allgemeinen Didaktik: Ein Kommentar. *Jahrbuch für Allgemeine Didaktik*, 3, 219–230.
- Terhart, E., Schulze-Stocker, F., Kunina-Habenicht, O., Dicke, T., Förster, D., Lohse-Bossenz, H., Gößling, J., Kunter, M., Baumert, J., & Leutner, D. (2012). Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung – Eine Kurzdarstellung des BiWiss-Projekts. *Lehrerbildung auf dem Prüfstand*, 5(2), 96–106.
- Tietgens, H. (1967). *Schule und Erwachsenenbildung*. Braunschweig: Westermann.
- Tietgens, H. (1981). *Angebotsplanung und -realisation. Kurseinheit 4: Möglichkeiten didaktischen Handelns II*. Hagen: Fernuniversität Hagen.
- Vinepac-Project (2008). *Handbook for the use of Validpack for the validation of psychopedagogical adult educators' competences*. Timisoara: Editura Mirton.
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969.
- 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.

- Voss, T., Kunter, M., Seiz, J., Hoehne, V., & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik*, 60(2), 184–201.
- Wu, M.L., Adam, R.J., Wilson, M.R., & Haldane, S.A. (2007). *ACER ConQuest version 2.0. Generalised item response modelling software*. Camberwell: ACER Press.
- Zeuner, C. (2003). Rückblicke. Erwachsenenbildung und Schule – (k)ein Verhältnis? *DIE Zeitschrift für Erwachsenenbildung*, 10(3), 43–44.
- Ziep, K.-D. (1990). *Der Dozent in der Weiterbildung. Professionalisierung und Handlungskompetenzen*. Weinheim: Deutscher Studien Verlag.
- Zierer, K. (2013). Eklektische Didaktik. *Jahrbuch für Allgemeine Didaktik*, 3, 203–218.

7 Studie 2 Das Wissen über Lehr-Lernmethoden und -konzepte von Lehrkräften der Erwachsenen- und Weiterbildung: Ist es ein- oder mehrdimensional konzeptualisierbar und erfassbar?

Das Manuskript wurde am 19.03.2018 erstmals eingereicht beim Journal for Educational Research Online / Journal für Bildungsforschung Online (JERO) und nach Begutachtung wiedereingereicht. Nachfolgend ist das wiedereingereichte Manuskript abgebildet. Das Manuskript wurde noch nicht angenommen und publiziert. Die in dieser Dissertationsschrift abgebildete Version wird voraussichtlich nicht mit der publizierten Version übereinstimmen. Das Manuskript wurde mit der folgenden Autorenanreihung eingereicht:

Marx, C., Goeze, A., Voss, T., Deutscher, V. K., Kelava, A. & Schrader, J. (eingereicht). Das Wissen über Lehr-Lernmethoden von Lehrkräften der Erwachsenen- und Weiterbildung: Ist es ein- oder mehrdimensional konzeptualisierbar und erfassbar? *Journal für Bildungsforschung Online*.

Zusammenfassung

Das Wissen über Lehr-Lernmethoden und -konzepte gilt als eine wichtige Facette des pädagogisch-psychologischen Wissens (PPK) von Lehrkräften der Erwachsenen- und Weiterbildung (EB/WB) für die Gestaltung von Lehr-Lernsituationen. Mit Blick auf die große Zahl an Lehr-Lernmethoden und -konzepten stellt der Beitrag die Frage, ob eine eindimensionale oder mehrdimensionale Konzeptualisierung dieses Wissens passend erscheint. Aus theoretisch-konzeptueller Perspektive ist die Annahme einer hierarchischen Struktur des Wissens über Lehr-Lernmethoden und -konzepte plausibel, die über verschiedene Lehr-Lernmethoden allgemeines Wissen und für einzelne Lehr-Lernmethoden spezifisches Wissen postuliert. Empirisch wurde überprüft, ob ein correlated-factors und ein eindimensionales Modell einerseits oder ein Bifaktor Modell, Testlet Modell oder Rasch Testlet Modell andererseits empirisch am besten auf die Daten passt. Annahmenkonform deuten die informationstheoretischen Maße auf die Überlegenheit einer hierarchischen Modellierung durch ein Rasch Testlet Modell hin. Abschließend werden die Befunde vor dem Hintergrund der Weiterentwicklung des Tests für die Forschung und Praxis der EB/WB diskutiert.

Schlüsselwörter

Pädagogisch-psychologisches Wissen; Lehr-Lernmethoden und -konzepte; Lehrkräfte; Trainer; Erwachsenenbildung/Weiterbildung

1 Gründe für die Entwicklung von Tests zur Erfassung von Aspekten der professionellen Kompetenz von Lehrkräften in der Erwachsenenbildung/Weiterbildung

Die allgemeine Erwachsenenbildung und die beruflich-betriebliche Weiterbildung (EB/WB) gelten in Deutschland nach dem Elementar- und Primarbereich, den weiterführenden Schulen, der Berufsausbildung sowie der tertiären Hochschulbildung als der quartäre Sektor des Bildungssystems, der den institutionalisierten Rahmen für das lebenslange Lernen abschließt. Veranstaltungen der EB/WB finden beispielsweise in Einrichtungen der Volkshochschulen, der Familienbildungsstätten, bei kommerziellen Sprachlernanbietern oder – vom Umfang her besonders bedeutsam – innerbetrieblich statt, können aber auch durch digitale Angebote verwirklicht werden. Mit dem Bedeutungsgewinn des lebenslangen Lernens (Bundesministerium für Arbeit und Soziales, 2014; Europäische Kommission, 2001, 2007) werden Fragen der Effektivität und Effizienz von Schule, Berufsbildung, Hochschule und zunehmend auch der EB/WB drängender gestellt (Kuper & Schrader, in Druck).

Wie für den Schulkontext vielfach empirisch gesichert (z.B. Hattie, 2009), wird auch für die EB/WB angenommen, dass Lehrkräfte durch ihre professionelle Kompetenz maßgeblich für den Lernerfolg von Teilnehmenden mitverantwortlich sind (Collins & Pratt, 2011; Kraft et al., 2009; Schrader et al., 2010; Siebert, 2012). Insgesamt arbeiten derzeit in Deutschland ca. 700.000 Personen in der EB/WB, davon ca. 530.000 Lehrkräfte, von denen ein großer Teil zumeist nicht in Vollzeit beschäftigt ist und hauptberuflich eine andere Tätigkeit ausübt oder aber noch nicht oder nicht (mehr) erwerbstätig ist (Autorengruppe wb-personalmonitor, 2016). Berücksichtigt man noch in der beruflich-betrieblichen Weiterbildung tätige Trainer, die in Erhebungen regelmäßig aufgrund ihrer häufig temporären Selbstständigkeit und des relativ geringen Organisationsgrads in Berufsverbänden schwer zu erfassen sind (Autorengruppe wb-personalmonitor, 2016), dann ist die Zahl der Lehrkräfte in der EB/WB (verglichen für die Jahre 2014/2015) vermutlich größer als die der Lehrkräfte im allgemeinen und beruflichen Schulsystem (Statistisches Bundesamt, 2015a, 2015b) und deutlich größer als im Hochschulsystem (Statistisches Bundesamt, 2016).

Der Berufszugang und die Berufsausübung sind in der EB/WB kaum staatlich, tarifvertraglich oder berufsständisch reglementiert und damit seit jeher „offen“ für Lehrkräfte mit unterschiedlichem Ausbildungshintergrund. Insofern existieren auch keine allgemeinverbindlichen Abschlüsse oder Zertifikate, die von der Praxis einheitlich als Indikatoren für vorhandene Kompetenzen interpretiert werden können (Goeze & Schneider, 2014). Nur knapp ein Drittel der in der EB/WB tätigen Lehrkräfte verfügt über einen erziehungs- und bildungswissenschaftlichen Haupt- oder Nebenfach-Studienabschluss (Autorengruppe wb-personalmonitor, 2016). Vor diesem doppelten Hintergrund – den Effektivitätserwartungen an die EB/WB einerseits und der großen Offenheit für Personen mit

vielfältigen beruflichen Hintergründen andererseits – ist es verständlich, dass in den vergangenen Jahren das Interesse an der Aus-, und Weiterbildung sowie der Erfassung der Kompetenzen des Lehrpersonals z.B. mit Blick auf die Lehrkräfterekrutierung in der Praxis (Goeze & Schneider, 2014) gestiegen ist. Nicht zufällig gehen zahlreiche Initiativen von der Praxis aus oder werden in enger Kooperation mit der Praxis umgesetzt: So werden standardisierte Befragungen, etwa zum Fortbildungsbedarf, oder Auswertungen der Ratgeberliteratur genutzt, um Einblicke in die vorhandenen und gewünschten Kompetenzen von Lehrkräften zu gewinnen (Hippel & Tippelt, 2009; Schöb et al., 2016; Schrader et al., 2010). Instrumente und Verfahren werden entwickelt, die non-formal oder informell erworbene Kompetenzen sichtbar machen und anerkennen sollen (z.B. Lencer & Strauch, 2016; Schläfli & Sgier, 2008; Steiner, 2010). Nicht zuletzt wurden (www.videofallarbeit.de; www.wb-web.de) und werden (Projekt EULE¹⁷) digitale Lernplattformen zur Aus- und Fortbildung für EB/WB-Lehrkräfte entwickelt, die u.a. das pädagogisch-psychologische Wissen (PPK) von Lehrkräften adressieren. Auffallend ist aber, dass erst langsam mit der Entwicklung empirisch fundierter Instrumente zur Erfassung der professionellen Kompetenz von EB/WB-Lehrkräften begonnen wurde. Bislang liegt der Fokus auf Facetten des PPK (Marx et al., 2017; Rohs et al., 2017). Instrumente z.B. zum Self-Assessment, zur Unterstützung bei der Trainerauswahl oder der informierten Wahl bzw. Zuweisung von Lehrkräften zu Online-Lernpfaden auf Lernplattformen zur Aus- und Fortbildung sollten das PPK in seiner konzeptuellen Breite erfassen können. Insbesondere für Lernplattformen, die zukünftig das PPK erschöpfend durch Lerninhalte abbilden möchten (Schöb et al., 2016), aber auch für die Forschung ist ein das PPK von EB/WB-Lehrkräften umfassend und differenziert erfassender Test nötig.

Erst ein solcher Test ermöglicht für die Weiterbildungsforschung in der Praxis implizit bleibende oder für die Theorie- und Modellbildung explizierbare Annahmen untersuchen zu können, die für Wissenschaft und Praxis gleichermaßen relevant sind; z.B. die in der Praxis verbreitete Annahme, dass mit zunehmender Erfahrung einer Lehrkraft deren PPK und darüber vermittelt ihr Erfolg in der Umsetzung von Weiterbildungsveranstaltungen steigt – wobei mit Erfolg in der EB/WB (anders als in der Schule) nicht nur der Lernerfolg der Teilnehmenden gemeint ist, sondern z.B. auch deren subjektive Erfüllung von Nutzenerwartungen, ihre Wiederteilnahmequote, die Lösung von Problemen aus Sicht der Auftraggeber von Inhouse-Trainings etc. Darüber hinaus ermöglicht solch ein Test, der sprachlich („Lehrkraft“, „Lernende“, „Lehr-Lernsituation“) wie inhaltlich fach- und

¹⁷ EULE (Entwicklung einer webbasierten Lernumgebung für Weiterbildung, Kompetenzerwerb und Professionalisierung von Lehrenden der Erwachsenenbildung) ist ein vom BMBF gefördertes Forschungs- und Entwicklungsprojekt, das vom Deutschen Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e. V. in Bonn und der Universität Tübingen bearbeitet wird.

bildungsbereichs-unspezifisch angelegt ist und auf das im Kern all(en)-gemeine PPK abhebt, bildungsbereichsvergleichende Untersuchungen zur Verallgemeinerbarkeit von Befunden; z.B. ob bzw. in wie weit die (prädiktive) Bedeutung von PPK, die die Forschung für Schul-Lehrkräfte bereits hat identifizieren können (z.B. Lenske et al., 2016) in der EB/WB – und damit ggf. auch in anderen Bildungsbereichen wie der Frühpädagogik oder der Hochschulbildung – (nicht) gelten.

Der vorliegende Beitrag setzt die im Projekt Think¹⁸ (Marx et al., 2017; Voss et al., 2017) begonnene Entwicklung und Erprobung eines Tests zur Erfassung von Facetten des PPK fort und fokussiert das Wissen über Lehr-Lernmethoden und -konzepte. Vom Wissen über Lehr-Lernmethoden und -konzepte wird angenommen, dass es für die Gestaltung von Lehr-Lernsituationen besonders relevant ist. Mit Blick auf die große Zahl an Lehr-Lernmethoden und -konzepten, die unterschiedlichen Zielen und Phasen innerhalb von Lehr-Lernprozessen dienlich sind (z.B. Peterßen, 2009), stellt sich die Frage, inwiefern eine eindimensionale Konzeptualisierung des Wissen über Lehr-Lernmethoden und -konzepte, wie sie z.B. im Projekt Think vorgenommen wurde, für die angesprochenen Verwendungszwecke adäquat erscheint oder ob theoretisch-konzeptuelle Überlegungen und empirische Befunde für eine mehrdimensionale Konzeptualisierung und Erfassung dieser Facette sprechen. Die Frage nach der ein- oder mehrdimensionalen Konzeptualisierbarkeit und Erfassbarkeit des Wissen über Lehr-Lernmethoden und -konzepte ist Gegenstand des vorliegenden Beitrags. Ziel dieses Beitrags ist es, durch die Bearbeitung dieser Frage Hinweise auf die Weiterentwicklung des Tests aus dem Projekt Think – hier mit dem Fokus auf der Facette Wissen über Lehr-Lernmethoden und -konzepte bei Lehrkräften der EB/WB – zu gewinnen.

2 Forschungsstand zum pädagogisch-psychologischen Wissen von Lehrkräften in der Erwachsenenbildung/Weiterbildung und in der Schule

PPK ist definiert als Wissen zur Gestaltung und Optimierung von Lehr-Lernsituationen in verschiedenen Fächern und Bildungsbereichen, das deklaratives und prozedurales Wissen beinhaltet, aber auch Erfahrungswissen, da es im Verlauf der beruflichen Karriere erworben wird (Voss et al., 2015; Voss, Kunter et al., 2011). Das PPK ist u.a. neben dem Fachwissen (content knowledge, CK) und dem fachdidaktischen Wissen (pedagogical content knowledge, PCK) ein Bereich des professionellen Wissens von Lehrkräften (Baumert & Kunter, 2011).

¹⁸ Think (Using Digital Media to Assess Generic Aspects of Teachers' Professional Knowledge in Different Educational Contexts) ist ein Projekt-Cluster des WissenschaftsCampus Tübingen, das vom Deutschen Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e. V. in Bonn, dem Hector-Institut für Empirische Bildungsforschung in Tübingen und der Universität Freiburg bearbeitet wird.

Das für den Schulkontext entwickelte und dort empirisch bewährte COACTIV-Kompetenzmodell (Kunter, Baumert et al., 2011) wird vermehrt auch für den Bereich der EB/WB adaptiert (z.B. Lencer & Strauch, 2016; Rohs et al., 2017), weil es generische und damit für die EB/WB relevant erscheinende Bereiche der professionellen Kompetenz von Lehrkräften thematisiert. Während das COACTIV-Kompetenzmodell im Schulbereich empirische Forschung angeregt hat, fehlen vergleichbare Studien in der EB/WB noch. Zwar wird das in konkreten Lehr-Lernsituationen anzuwendende PPK in Kompetenzmodellen in unterschiedlichen Begrifflichkeiten durchgehend berücksichtigt (z.B. Ziep, 1990), empirische Annäherungen beschränken sich bislang aber überwiegend auf Instrumente zur Selbst- und/oder Fremdeinschätzung (z.B. Vinepac-Project, 2008). Dafür lassen sich mindestens zwei Gründe benennen. Zum einen war die Forschung zur Professionalisierung der EB/WB lange auf das noch am ehesten im Sinne von Hauptberuflichkeit „professionalisierte“ planend-disponierend tätige Personal fokussiert (Schrader, 2010a). Zum anderen fand eine Auseinandersetzung mit schulpädagogischen sowie pädagogisch-psychologischen Modellen und Befunden zur professionellen Kompetenz von Lehrkräften bislang nur in geringem Ausmaß statt (Marx et al., 2017; Marx et al., 2014b).

In den vergangenen Jahren wuchs das Interesse an der Erforschung konkreter Lehr-Lernsituationen in der EB/WB (Kade et al., 2014), ihren Gelingensbedingungen und damit den professionellen Kompetenzen der Lehrkräfte (z.B. Lattke & Jütte, 2014). So widmeten sich Studien der Beschreibung individueller Professionalisierungsprozesse, dem didaktischen Handeln und den metakognitiven Kompetenzen des Lehrpersonals. Maier-Gutheil (2012) identifizierte anhand biographischer Fallstudien differente Prozesse der Herausbildung von Professionalität. Bastian (1997) erstellte qualitative Kompetenzprofile von KursleiterInnen, denen jeweils spezifische Wissensschwerpunkte und Themenprofile zugeordnet wurden. Mehrere Studien widmeten sich dem Planungshandeln von Lehrkräften: Hof (2001) fragte z.B. nach dem Zusammenhang des subjektiven Wissensverständnisses von Lehrkräften und deren Unterrichtskonzepten, Stanik (2016) nach den Entscheidungsfeldern für und den Einflussfaktoren auf mikrodidaktische Planungen, Haberzeth (2010) identifizierte die Planungs-Logiken des Handelns der Lehrkräfte, nach denen Inhalte für eigene Lehr-Lernveranstaltungen ausgewählt werden, Pachner (2013) untersucht die Kompetenz zur Selbstreflexion von Lehrkräften. Diese Studien kamen jedoch ohne kompetenztheoretische Modellierungen und Operationalisierungen des PPK aus.

Die bislang entwickelten Instrumente zur Erfassung (von Facetten) des PPK wurden für den Schulkontext konzeptualisiert (Beck et al., 2008; König & Blömeke, 2009; Lenske et al., 2015; Linninger et al., 2015; Seifert, Hilligus & Schaper, 2009; Voss, Kunter et al., 2011). Die Instrumente beziehen sich sprachlich bzw. bei videobasierten Instrumenten auch bildlich

nicht auf die EB/WB, sondern auf Schul-Lehrerinnen und Lehrer sowie auf Kinder und Heranwachsende als Lernende und auf Schulfächer wie z.B. Mathematik und Physik, welche in dieser Form wenig passend sind für Angebote der EB/WB, die sich eher an den Lebenswelten und Anwendungsinteressen der Adressaten als an einer Systematik von Fächern orientieren. Dies, und der Anspruch, die fach- und bildungsbereichsübergreifende Allgemeinheit des PPK untersuchbar zu machen, – was bisher implizit oder explizit zwar angenommen, aber insbesondere für den Bildungsbereichsvergleich noch nicht empirisch überprüft wurde – begründet die Entwicklung eines fach- und bildungsbereichsübergreifend einsetzbaren Instruments zur Erfassung des PPK im Projekt Think (siehe Marx et al., 2017).

Über verschiedene (schulbezogene) Forschungsprojekte hinweg stützen Befunde zur Dimensionalität des PPK die Annahme, dass mehrere Wissensfacetten einem gemeinsamen Konstrukt zugeordnet werden können. Die derzeit vorliegenden Instrumente zur Erfassung des PPK weisen eine große Schnittmenge hinsichtlich ihrer Facetten auf, bislang besteht allerdings keine Einigkeit darüber, aus wie vielen Wissensfacetten das Konstrukt besteht (König et al., 2017; Voss et al., 2015). Die empirischen Analysen zeigen, dass sich die Daten zum gesamten PPK zumeist mit Hilfe mehrdimensionaler Modelle am besten beschreiben lassen (Voss et al., 2015).

An das Kompetenzmodell von COACTIV und die Arbeiten von Voss und Kunter et al. (2011) zum PPK schließt das Projekt Think an mit dem Ziel, ein über verschiedene Fächer und Bildungsbereiche (z.B. EB/WB und Schule) einsetzbares Instrument zu entwickeln. Im Projekt Think wird PPK fach- und bildungsbereichsübergreifend konzeptualisiert und in acht Facetten untergliedert (Marx et al., 2017; Marx et al., 2014b; Voss et al., 2015):

1. Wissen über Lernprozesse von Lernenden
2. Wissen über Heterogenität der Lernenden
3. Wissen über Lehr-Lernmethoden und -konzepte
4. Wissen über Lehr- und Lernziele
5. Wissen über Führung von Lerngruppen
6. Wissen über Kommunikation und Interaktion mit Lernenden
7. Wissen über Gestaltung von Lernumgebungen
8. Wissen über Individual- und Lernprozessdiagnostik

Die Konzeptualisierung wurde anschließend an die Konzeptualisierung von Voss und Kunter et al. (2011) auf der Grundlage einer systematischen Literaturrecherche mit Inhaltsanalyse relevanter Quellen und einer Expertenbefragung entwickelt (Marx et al., 2017). Durch die acht Facetten geht die Konzeptualisierung von Marx et al. (2017) konform zur Annahme der Mehrdimensionalität des PPK (Voss et al., 2015). Erweitert wird diese Konzeptualisierung durch die Beschreibung der acht Facetten anhand von insgesamt 30 Subfacetten. Diese

stellten im Projekt ThinkK Hilfskonstrukte bei der Operationalisierung dar ohne den Anspruch, die Subfacetten empirisch zu differenzieren; gleichwohl deuten die Subfacetten aber eine etwaige Mehrdimensionalität Ebene der acht Facetten an.

Vor diesem Hintergrund wird im folgenden Kapitel auf der Grundlage theoretisch-konzeptueller Überlegungen die Frage bearbeitet, wie eine ein-, eine mehrdimensionale und eine hierarchische Konzeptualisierung für das Wissen über Lehr-Lernmethoden und -konzepte begründet werden kann.

3 Ein- und mehrdimensionale sowie hierarchische Konzeptualisierung des Wissens über Lehr-Lernmethoden und -konzepte

Lehr-Lernmethoden ergeben sich aus einer spezifischen Kombination von Arbeits- (z.B. Übung) und Sozialform (Einzel-, Partner-, Gruppen- oder Plenumsarbeit). Den Unterschied, der gern vermischt, vertauscht oder missverstanden wird, können zwei Fragen verdeutlichen (z.B. Jank & Meyer, 2008): Wer arbeitet mit wem zusammen? (=Sozialform). Welche Handlungsmuster (z.B. Vortrag halten, etwas wiedergeben) sollen ausgeführt werden? (=Arbeitsform). Blickt man in die Literatur, so findet sich eine (nahezu beliebig erweiterbar) große Zahl an Kombinationen aus Arbeits- und Sozialformen, die in verschiedenen Phasen für verschiedene Ziele in Lehr-Lernsituationen eingesetzt werden können (z.B. Peterßen, 2009; Weidenmann, 2008).

Bei einer eindimensionalen Konzeptualisierung des Wissen über Lehr-Lernmethoden und -konzepte stellt sich die Frage, was allgemeines Wissen über verschiedene Kombinationen von Arbeits- und Sozialformen hinweg sein könnte. Dies könnte das Wissen über die „Orchestrierung“ von Lehr-Lernsituationen, d.h. die Bedingtheit von Sicht- und Tiefenstrukturen (Oser & Baeriswyl, 2001) sein. Im Rahmen des Oser'schen Lehr-Lern-Konzepts werden durch den Begriff *Sichtstrukturen* sichtbare und niedriginferent zu beurteilende Merkmale von Lehr-Lernsituationen beschrieben. Dazu zählen Arbeits- und Sozialformen sowie das Material bzw. die Aufgaben und damit der Stoff, der Gegenstand einer Lehr-Lernsituation ist. Das Vorliegen bestimmter Sichtstrukturen erlaubt zunächst keine Rückschlüsse auf die Qualität eines Lehr-Lernprozesses, da innerhalb der gleichen Sichtstruktur die Interaktion zwischen Lehrenden und Lernenden völlig unterschiedlich gestaltet werden kann (Oser & Baeriswyl, 2001). *Tiefenstrukturen* beziehen sich auf die nicht direkt beobachtbaren Lernprozesse von Lernenden (Oser & Baeriswyl, 2001), die den Lernerfolg begründen. Die tiefenstrukturelle Qualität einer Lehr-Lernsituation ergibt sich aus dem Wissen der Lehrkraft um die Notwendigkeit einer pädagogisch-didaktisch urteilskräftigen Relationierung von Lehr-Lernziel einerseits und der Wahl der dafür ‚passenden‘ sichtstrukturellen methodischen Vorgehensweise andererseits. Das Wissen über diese Bedingtheit von Sicht- und Tiefenstrukturen stellt dann auf einer theoretisch-

konzeptuellen Ebene die Gemeinsamkeit über alle Aufgaben hinweg dar. Denn dieses Wissen könnte zur Beantwortung *aller* Aufgaben der Facette hilfreich sein: Weiß eine Lehrkraft, dass Sichtstrukturmerkmale (z.B. ein Lehrervortrag oder eine Gruppenarbeit) nicht pauschal wirksam oder unwirksam sind, sondern differenziert sie, *unter welchen Bedingungen* eine Sichtstruktur zur Ermöglichung welcher tiefenstruktureller und damit kognitiver Prozesse mehr oder weniger hilfreich ist, dann wird sie höhere Lösungswahrscheinlichkeiten bei allen Aufgaben dieser Facette aufweisen.

Das Wissen über Lehr-Lernmethoden und -konzepte wurde in Marx et al. (2017) in die zwei Subfacetten a) „Arbeits- und Sozialformen sowie deren Kombination und zieladäquater und effektiver Einsatz in Lehr-Lernsituationen“ sowie b) „Konzepte zu individualisierten, kooperativen oder offenen Formen des Lernangebots sowie deren Umsetzung“ untergliedert. Dies begründete sich dadurch, dass Arbeits- und Sozialformen zumeist kleinere Handlungs-/Interaktionseinheiten darstellen als die „größeren“ Lehr-Lernkonzepte. Eine Operationalisierung anhand dieser als Hilfskonstrukte gedachten Subfacetten in zwei Skalen erscheint nicht sinnvoll, da Subfacette a) eine Teilmenge von Subfacette b) darstellt. Eine bessere Möglichkeit, Wissen über Lehr-Lernmethoden und -konzepte (unter Berücksichtigung des Gedankens, dass dieses Wissen bei EB/WB-Lehrkräften „verinselt“ erworben und vorhanden sein könnte), mehrdimensional zu konzeptualisieren, ist es, zentrale und in der (Ratgeber-)Literatur häufig angesprochene Lehr-Lernmethoden als jeweils eine Subfacette dieses Wissens abzubilden. Dass verschiedene Lehr-Lernmethoden unterschiedliches Wissen zur Umsetzung dieser erforderlich machen, soll an drei zentralen (Kombinationen von) Arbeits- und Sozialformen exemplifiziert werden: dem Lehrervortrag, der Gruppenarbeit und der Feedback-Methode.

Wissen über die Gestaltung eines Lehrervortrag erfordert kognitionspsychologisches Wissen (z.B. Wissen über die Kapazität des Arbeitsspeichers des Gedächtnisses) in Kombination mit Wissen über Lehr-Lernmethoden und -konzepte (z.B. Weidenmann, 2008), das sich für die Beurteilung der Angemessenheit als praktisch relevantes Wissen an der didaktischen Zielsetzung relationiert: Durch den Lehrervortrag können Informationen zu einem Thema angeboten werden und dieser kann beim subjektiven Anlegen bzw. beim Erwerb einer kognitiven Grundstruktur hilfreich sein (Teilaufgaben a und b). Für die Ermöglichung komplexerer kognitiver Operationen (in der konkreten Situation) ist dieser zumeist aber weniger hilfreich (Teilaufgaben c, d, e). Das Wissen über den Lehrervortrag wurde im Projekt Think durch die Aufgabe in Abb.1 operationalisiert.

Bitte beurteilen Sie, welche der folgenden Aussagen stimmen oder nicht stimmen.		
<i>Kreuzen Sie bitte in jeder Zeile ein Kästchen an.</i>		
Der Lehrervortrag (auch als Vortragsmethode bezeichnet) ist dann geeignet, wenn...	Stimmt nicht	Stimmt
a) ... das Hauptziel in der Vermittlung von Informationen besteht.	<input type="checkbox"/>	<input type="checkbox"/>
b) ... eine Einführung in einen bestimmten Bereich gegeben werden soll.	<input type="checkbox"/>	<input type="checkbox"/>
c) ... das Behalten des Stoffs über einen längeren Zeitraum erwünscht ist.	<input type="checkbox"/>	<input type="checkbox"/>
d) ... die Beteiligung der Lernenden eine wesentliche Voraussetzung ist, um ein Lernziel zu erreichen.	<input type="checkbox"/>	<input type="checkbox"/>
e) ... der Stoff für die Lernenden komplex ist oder eine Menge Details enthält.	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 1 Aufgabe zur Erfassung des Wissens über den Lehrervortrag

Bei der Gestaltung von Gruppenarbeiten ist sozialpsychologisches Wissen in Kombination mit Wissen über Lehr-Lernmethoden und -konzepte Wissen nötig: hier ist z.B. Wissen über unterschiedliche Typen von Interdependenzen der beteiligten Lernenden wichtig und damit in Zusammenhang stehend die (Un)Eignung verschiedener Aufgabentypen (disjunktive, konjunktive oder additive Aufgaben) für Gruppenarbeiten, um für die Lernenden demotivierende Phänomene wie den Trittbrettfahreneffekt zu vermeiden (z.B. Wecker & Fischer, 2014). Dieses Wissen wurde durch die Aufgabe in Abb. 2 operationalisiert, die aus dem COACTIV-Test von Voss und Kunter et al. (2011) sprachlich adaptiert wurde.

Bei Gruppenarbeiten kommt es häufig vor, dass sich einzelne Gruppenmitglieder nicht anstrengen, während andere motiviert bei der Arbeit sind.

Welche der folgenden Aufgaben beugen diesem Phänomen vor und sind weniger gut, welche besonders gut für eine Gruppenarbeit geeignet?

Kreuzen Sie bitte in jeder Zeile ein Kästchen an.

	falsch	richtig
a) Die Lernenden sollen zum Thema Klimawandel ein Brainstorming in der Kleingruppe durchführen, wie man die Klima-Katastrophe verhindern kann. Die gefundenen Möglichkeiten sollen in einer Liste gesammelt werden.		
b) Die Lernenden sollen zum Thema Umweltschutz eine Collage in der Kleingruppe erarbeiten, für die jeder Lernende fünf Beispiele von Umweltsünden im Alltag, dokumentiert durch Bilder aus Zeitschriften,- finden muss.		
c) Die Lernenden sollen gemeinsam einen Text verfassen und jeder ist für jeweils einen Abschnitt verantwortlich.		
d) Die Lernenden sollen gemeinsam Erklärungen für den Ausgang eines durchgeführten physikalischen Experiments finden.		
e) Die Lernenden sollen gemeinsam eine mathematische Textaufgabe lösen.		

Abb. 2 Aufgabe zur Erfassung des Wissens über die Gestaltung von Gruppenarbeiten

Feedback geben ist eine der am häufigsten eingesetzten Lehr-Lernmethoden um Lernprozesse und Verhaltensänderungen zu unterstützen (Strijbos & Müller, 2014). Beim Feedback geben ist u.a. motivationspsychologisches Wissen in Kombination mit Wissen über Lehr-Lernmethoden und -konzepte nötig. Hier ist das Wissen über den Einfluss personaler Faktoren (z.B. Attributionen) auf die Verarbeitung von Feedback relevant. Die Attribution einer Rückmeldung ist bedeutsam, da mit den angenommenen Ursachen für ein Ereignis motivationale und emotionale Auswirkungen sowie Überzeugungen bezüglich möglicher individueller Handlungsspielräume verbunden sind. Zentral ist hier die Frage, ob ein bestimmtes Handlungsergebnis durch eine Person aktuell und zukünftig als beeinflussbar bewertet wird – dies entspricht einer internalen, variablen und kontrollierbaren Ursachenzuschreibung wie sie im Klassifikationsschema der Gründe für Handlungsergebnisse nach Weiner (1985) konzeptualisiert wird. Dieses Wissen wurde durch die Aufgabe in Abb. 3 operationalisiert, die aus dem COACTIV-Test von Voss und Kunter et al. (2011) sprachlich adaptiert wurde.

Ein Lerner zeigt im Vergleich zu seinen bisherigen Leistungen eine unterdurchschnittliche Leistung.		
Welche Rückmeldung nach Misserfolgen ist weniger gut, welche ist besonders gut geeignet, seine Motivation für die Beschäftigung mit den Kursinhalten zu erhöhen?		
<i>Kreuzen Sie bitte in jeder Zeile ein Kästchen an.</i>		
	weniger gut geeignet	besonders gut geeignet
a) „Sie haben diesmal zu flüchtig gearbeitet. Wenn Sie sich mehr anstrengen, können Sie das schaffen.“	<input type="checkbox"/>	<input type="checkbox"/>
b) „Die Aufgaben waren dieses Mal einfach zu schwer. Das nächste Mal klappt es bestimmt besser!“	<input type="checkbox"/>	<input type="checkbox"/>
c) „Das ist gar nicht schlimm, Sie hatten dieses Mal einfach Pech.“	<input type="checkbox"/>	<input type="checkbox"/>
d) „Machen Sie sich nichts daraus, Ihre Stärken liegen in anderen Bereichen.“	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 3 Aufgabe zur Erfassung des Wissens die Feedback-Methode

Die oben gegebenen Beispiele zeigen auf, dass für verschiedene Kombinationen von Arbeits- und Sozialformen unterschiedliches Wissen relevant ist. Eine dritte Möglichkeit, das Wissen über Lehr-Lernmethoden und -konzepte zu konzeptualisieren, ist eine hierarchische Konzeptualisierung. Eine hierarchische Konzeptualisierung beinhaltet für verschiedene Kombinationen aus Arbeits- und Sozialformen allgemeines *und* spezifisches Wissen. Allgemeines Wissen kann dabei das Wissen über die „Orchestrierung“ von Lehr-Lernsituationen, d.h. die Bedingtheit von Sicht- und Tiefenstrukturen (Oser & Baeriswyl, 2001) sein. Spezifisches Wissen kann z.B. das oben dargestellte Wissen zum Lehrervortrag, der Gruppenarbeit oder der Feedback-Methode sein. Unter Berücksichtigung des Gedankens, dass Wissen über Lehr-Lernmethoden und -konzepte bei EB/WB-Lehrkräften „verinselt“ erworben und vorhanden sein könnte, erscheint es plausibel, dass allgemeines Wissen und spezifische Subfacetten sowie jeweils die spezifischen Subfacetten unabhängig voneinander zu konzeptualisieren sind.

4 Fragestellung und Annahmen

Die Fragen des vorliegenden Beitrages sind, ob (1) eine mehrdimensionale, eine eindimensionale oder eine hierarchische Modellierung das Wissen über Lehr-Lernmethoden und -konzepte von Lehrkräften in der EB/WB besser abbildet und (2) wie reliabel dieses Wissen erfasst wird.

Wir nehmen an, dass eine hierarchische Konzeptualisierung die Daten am besten erklärt. Dies begründet sich dadurch, dass die Aufgaben sowohl auf der Basis der in Kapitel 3 ausgeführten theoretisch-konzeptuellen Überlegungen als auch, dass in den Aufgaben jeweils konkrete Lehr-Lernmethoden und damit spezifische Inhalte angesprochen werden, für die – zumindest bzw. im Besonderen für die EB/WB-Lehrkräfte – keine systematischen Zusammenhänge zu erwarten sind.

5 Methodisches Vorgehen

5.1 Stichprobe

An der Hauptstudie des Projekts Think nahmen insgesamt $N = 212$ EB-Lehrkräfte teil, dabei handelt es sich um eine Ad-hoc-Stichprobe. Die Lehrkräfte im Alter von 24 bis 77 Jahren ($M = 48.47$; $SD = 12.16$) verfügten zum Erhebungszeitpunkt über eine Lehrerfahrung zwischen 0.2 und 50 Jahren ($M=12.71$; $SD=10.38$) und lehrten pro Woche zwischen 1 und 50 Stunden ($M = 12.13$; $SD = 11.10$). Der Anteil an Akademikern (72.77%), die mehrheitlich nebenberufliche Beschäftigung bei mehreren Bildungseinrichtungen (64.34%) und der Anteil an weiblichen Lehrkräften (51.89%) weisen die vorliegende Stichprobe als typisch für die

EB/WB aus (siehe Autorengruppe wb-personalmonitor, 2016). Die Lehrkräfte dieser Stichprobe sind in verschiedenen Reproduktionskontexten (Schrader, 2010b) der EB/WB tätig („Unternehmen“ und freier „Markt“: 29%; werte- oder interessengebundene „Gemeinschaften“: 21%; öffentlich (mit)finanzierte Bildungseinrichtungen mit gesetzlichem Auftrag („Staat“): 50%) und bieten Veranstaltungen zu der gesamten Bandbreite an Themen an, von der (fachbezogenen) beruflichen Weiterbildung über Angebote der allgemeinen bis zur politischen Erwachsenenbildung. Nur etwa ein Drittel der in der EB/WB tätigen Lehrkräfte verfügt über einen formalen erziehungs- und bildungswissenschaftlichen Haupt- oder Nebenfachabschluss (Autorengruppe wb-personalmonitor, 2016), ein Anteil, der wiederholt für verschiedene Stichproben von EB/WB-Lehrkräften genannt wird (Bonnes & Hochholding, 2016; Schütz & Nittel, 2012). Zu dieser Gruppe mit erziehungs- und bildungswissenschaftlicher Vorbildung zählen Schullehrkräfte sowie durch erziehungs- und bildungswissenschaftliche Studiengänge ausgebildete Personen, deren Gemeinsamkeit ihre Ausbildung in bildungswissenschaftlichen Inhalten ist, die das PPK miteinschließen (Linninger et al., 2015). Eine zweite Gruppe bilden Lehrkräfte ohne erziehungs- und bildungswissenschaftlichen Haupt- oder Nebenfachabschluss, die in etwa zwei Drittel der Lehrkräfte in der EB/WB ausmachen und die sich für eine Lehrtätigkeit vor allem über ihre Fachexpertise qualifizieren (Autorengruppe wb-personalmonitor, 2016). Lehrkräfte mit bildungswissenschaftlichem Studium sind – diesen Anteilsverhältnissen fast typisch entsprechend – $N = 54$ (25.47%) in der Stichprobe enthalten, Lehrkräfte ohne bildungswissenschaftliches Studium $N = 158$ (74.53%).

5.2 Durchführung

Die Stichprobe wurde mithilfe von zwei großen Volkshochschulen in Bayern und Nordrhein-Westfalen sowie der IHK in Nordrhein-Westfalen und verschiedenen Trainernetzwerken rekrutiert. Die Teilnahme fand individuell an einem vorbereiteten Laptop in den Räumlichkeiten der Einrichtungen statt, war freiwillig und wurde durch Bezahlung honoriert. 42.45% der Stichprobe wurde in der beschriebenen Form rekrutiert, der übrige Teil der Stichprobe bearbeitete den Fragebogen online (www.videofallarbeit.de), ebenfalls freiwillig und auf Honorarbasis. Allen befragten Lehrkräften lagen die insgesamt 67 Testaufgaben des gesamten PPK-Test vor. In den Fällen, bei denen die Testzeit der EB-Lehrkräfte weit über den vereinbarten Rahmen von ca. 90 Minuten hinausging, wurde die Möglichkeit gegeben, restliche Testaufgaben zu überspringen und nur noch die soziodemographischen Angaben auszufüllen. Daher wurden die Testaufgaben in drei Versionen mit unterschiedlicher Reihenfolge präsentiert.

5.3 Instrument

Die in diesem Beitrag analysierten Aufgaben stellen nur eine Teilmenge des gesamten Tests dar. Gegenstand der Analysen sind Aufgaben aus dem Projekt ThinK, die zum Teil anschließend an die Aufgaben von Voss und Kunter et al. (2011) entwickelt wurden und somit Schnittmengen zu diesen aufweisen, aber nicht deckungsgleich sind. Alle fünf Aufgaben, die Gegenstand der Analysen dieses Beitrags sind, liegen im gleichen Aufgabenformat vor, wie die Aufgaben in Abb. 1-3, die vier oder fünf Teilaufgaben im Richtig-Falsch-Antwortformat umfassen. Die Richtig-Falsch-Antworten haben einen gemeinsamen Aufgabenstamm und ähneln damit Testlets (Wainer & Kiely, 1987). Jede Teilaufgabe wird mit null (nicht gelöst) oder eins (gelöst) bepunktet. Aus den nachfolgend berichteten Analysen wurden insgesamt acht Teilaufgaben ausgeschlossen. Der Ausschluss begründet sich empirisch (Teilaufgaben wurden von nahezu allen Personen gelöst oder wurden von einer großen Zahl an Personen nicht bearbeitet). Die fünf Aufgaben hatten ihren inhaltlichen Schwerpunkt auf verschiedenen lehr-lernmethodischen Vorgehensweisen.

5.4 Umgang mit fehlenden Werten

Vor allem in sogenannten low-stakes assessments überspringen Testteilnehmer Aufgaben (omitted items) oder brechen aus verschiedenen Gründen die Bearbeitung eines Tests bzw. Fragebogens ab und produzieren dabei not-reached items. Rubin (1976) schlug zur Klassifikation von fehlenden Werten die Unterscheidung in missing completely at random (MCAR), missing at random (MAR) und not missing at random (MNAR) vor, die sich in der empirischen Bildungsforschung weitestgehend durchgesetzt hat (vgl. Lüdtke, Robitzsch, Trautwein & Köller, 2007). In der Literatur zu Missing Data werden MAR und MCAR auch als ignorable nonresponse bezeichnet und full information maximum likelihood (FIML) oder Multiple Imputation (MI) werden derzeit als die state-of-the-art Methoden für den Umgang mit diesen beiden Fälle betrachtet (Lüdtke et al., 2007).

Ein Problem ist, dass – aufgrund des Fehlens der Werte – nicht mit Sicherheit identifiziert werden kann, welcher der drei Fälle nach Rubin (1976) vorliegt, lediglich Hinweise darauf können gefunden werden. In Analysen der PISA 2006 Daten zeigte sich eine negative Korrelation ($r = -.33$) zwischen dem Anteil an richtigen Antworten pro Item und dem Anteil an omitted values pro Item, was darauf hinweist, dass einfache Aufgaben häufiger beantwortet und schwere Aufgaben eher übersprungen wurden (Rose, Davier & Xu, 2010); für die in dieser Untersuchung betrachteten Aufgaben zeigte sich ebenfalls eine negative Korrelation, allerdings nicht signifikant ($r = -.39$). Für den Umgang mit MNAR gibt es unterschiedliche Vorschläge: Einerseits wird von der häufig praktizierten Null-Imputation

abgeraten, da von dieser empirisch fundiert angenommen wird, dass sie zu verzerrten Ergebnissen führt (Pohl, Gräfe & Rose, 2014; Rose et al., 2010); andererseits wird diese Annahme kritisch diskutiert und gezeigt, dass auch der Verzicht auf Null-Imputation wiederum zu verzerrten Ergebnissen führen kann (Robitzsch, 2016).

Da nicht sicher ist, welche Art von fehlenden Werten nach Rubin (1976) vorliegt, werden im Folgenden fehlende Werte als solche behandelt. Denn Simulationsstudien weisen darauf hin, dass IRT-Modelle relativ robust sind, wenn nur wenige Werte fehlen (Rose et al., 2010). Für die hier präsentierten fünf Aufgaben fehlen 114 von 3180 Werten und damit lediglich 3.6%, was als unproblematischer relativer Anteil an Missings betrachtet wird (Kline, 2016). Um Personen mit fehlenden Werten auf einzelnen Variablen nicht aus den Analysen ausschließen zu müssen, wurde in Mplus (Muthén & Muthén, 1998-2015) die Funktion full information maximum likelihood (FIML) verwendet, da ein listenweiser Fallausschluss neben einem Verlust an Effizienz bei der Parameterschätzung auch zu fehlerhaften Befunden führen kann (Lüdtke et al., 2007).

5.5 Modellspezifikation

Spezifikation des eindimensionalen Modells

Eine eindimensionale Konzeptualisierung wird durch ein eindimensionales Modell modelliert. Besteht die Annahme der Eindimensionalität, werden die Unterschiede der Personen in den Indikatoren durch eine gemeinsame latente Variable aufgeklärt und pro Indikator wird ein Messfehler modelliert. Von den Messfehlern nehmen wir an, dass diese unkorreliert sind (Abb. 4, Modell A). Von den Teilaufgaben der Aufgaben wird angenommen, dass diese gleichermaßen das jeweils thematisierte Wissen indizieren, was durch gleiche Faktorladungen spezifiziert wird. Berichtet man Personenscores, dann beziehen sich diese eindeutig auf die eine spezifizierte latente Variable.

Spezifikation des mehrdimensionalen Modells

Passend zur vorgeschlagenen mehrdimensionalen Konzeptualisierung in Kapitel 3 ist ein *correlated-factors Modell*, das voneinander trennbare, aber korrelierte Wissensdimensionen annimmt (Abb. 4, Modell B). In einem correlated-factors Modell wird häufig ein ursprünglich als eindimensional konzeptualisiertes Konstrukt in verschiedene latente Variablen aufgeteilt (z.B. Brown, 2006). Allerdings wird im correlated-factors Modell *keine* latente Variable in das Modell eingeführt, die das die Facetten *einende Wissen* über alle Items hinweg erfasst (Reise, Moore & Haviland, 2010).

Von den Teilaufgaben der Aufgaben nehmen wir an, dass diese gleichermaßen das jeweils thematisierte Wissen indizieren, was durch gleiche Faktorladungen spezifiziert wird. Von den Messfehlern der Indikatoren nehmen wir an, dass diese unkorreliert sind. Welche Zusammenhänge sich zwischen den Subfacetten bzw. den einzelnen Faktoren letztlich zeigen, ist offen. Aufgrund der unterschiedlichen Inhaltsbereiche, die in den Aufgaben thematisiert werden und der zumeist unsystematischen pädagogischen Ausbildung der EB/WB-Lehrkräfte, dürften diese aber gering ausfallen. Berichtet man Personenscores für einzelne latente Variablen eines correlated-factors Modells, dann handelt es sich bei diesen um ein Amalgam aus der nicht mitmodellierten, aber allen Indikatoren gemeinsamen latenten Variablen sowie den jeweils spezifischen, modellierten latenten Variablen (Reise, 2012; Reise et al., 2010).

Spezifikation der hierarchischen Modelle

Passend zur Annahme einer hierarchischen Konzeptualisierung mit allgemeinem und spezifischem Wissen sind Modelle der Bifaktor-Familie (Abb. 4, Modell C): "Bifactor models are potentially applicable when (a) there is a general factor that is hypothesized to account for the commonality of the items; (b) there are multiple domain specific factors, each of which is hypothesized to account for the unique influence of the specific domain over and above the general factor; and (c) researchers may be interested in the domain specific factors as well as the common factor that is of focal interest." (Chen, West & Sousa, 2006). Zusätzlich zum von Chen et al. (2006) skizzierten Verwendungszweck werden Bifaktor Modelle auch dazu verwendet, um für „nuisance factors“ zu kontrollieren, z.B. dann, wenn nur der Generalfaktor im Fokus steht, nicht aber die spezifischen Faktoren, die z.B. durch verschiedene Aufgabenstimuli begründet sind (DeMars, 2013). Pro Teilaufgabe werden bei Modellen der Bifaktor-Familie zwei Ladungsparameter spezifiziert, für die (mindestens) drei verschiedene Annahmen getroffen werden können:

- 1) alle Ladungsparameter werden frei geschätzt (z.B. Reise, 2012) – nachfolgend als Bifaktor Modell bezeichnet.
- 2) Die Ladungsparameter der Teilaufgaben sind für den Generalfaktor und den spezifischen Faktor proportional zueinander (z.B. Bradlow, Wainer & Wang, 1999) – nachfolgend als Testlet Modell bezeichnet.
- 3) Die Ladungsparameter sind für die Teilaufgaben eines Aufgabenstamms gleich, jeweils für den Generalfaktor und den spezifischen Faktor (Wang & Wilson, 2005b) – nachfolgend als Rasch Testlet Modell bezeichnet.

Von den Messfehlern der Indikatoren nehmen wir an, dass diese unkorreliert sind. Konform zu den Überlegungen in Kapitel 3 erscheint das Rasch-Testlet Modell am

passendsten zu sein, da von den Teilaufgaben der Aufgaben angenommen wird, dass diese gleichermaßen das jeweils thematisierte Wissen indizieren.

Bei der Interpretation der spezifischen Faktoren ist der oben von Chen et al. (2006) angesprochene Punkt b) zu beachten: die spezifischen Faktoren erklären zusätzliche Varianz zum Generalfaktor. Berichtet man Personenscores, dann ist zu beachten, dass die Scores des Generalfaktors sich auf das in Kapitel 3 skizzierte allgemeine Wissen beziehen und die Scores der spezifischen Faktoren sich auf das in Kapitel 3 skizzierte spezifische Wissen über Lehr-Lernmethoden beziehen würden – nicht aber auf allgemeine Wissen über Lehr-Lernmethoden.

5.6 Statistische Analysen

Alle Analysen wurden mit Mplus Version 7.4 durchgeführt (Muthén & Muthén, 1998-2015). Für kategoriale Daten, wie sie häufig bei Wissenstests vorliegen, bietet sich zur Analyse der Dimensionalität eines Konstrukts die Multidimensional Item Response Theory (MIRT) an (Reckase, 2009). In der vorliegenden Arbeit werden alle Analysen im MIRT-Framework durchgeführt und dementsprechend für den Modellvergleich die informationstheoretischen Maße AIC, AICc, CAIC, BIC, SABIC berichtet. Offen ist, wie die Straf-Funktion informationstheoretischer Maße idealerweise aussieht (Rost, 2004). Daher werden die beiden gängigsten Maße AIC und BIC sowie Varianten dieser berichtet (AICc, CAIC, SABIC), die u.a. die Stichprobengröße berücksichtigen.

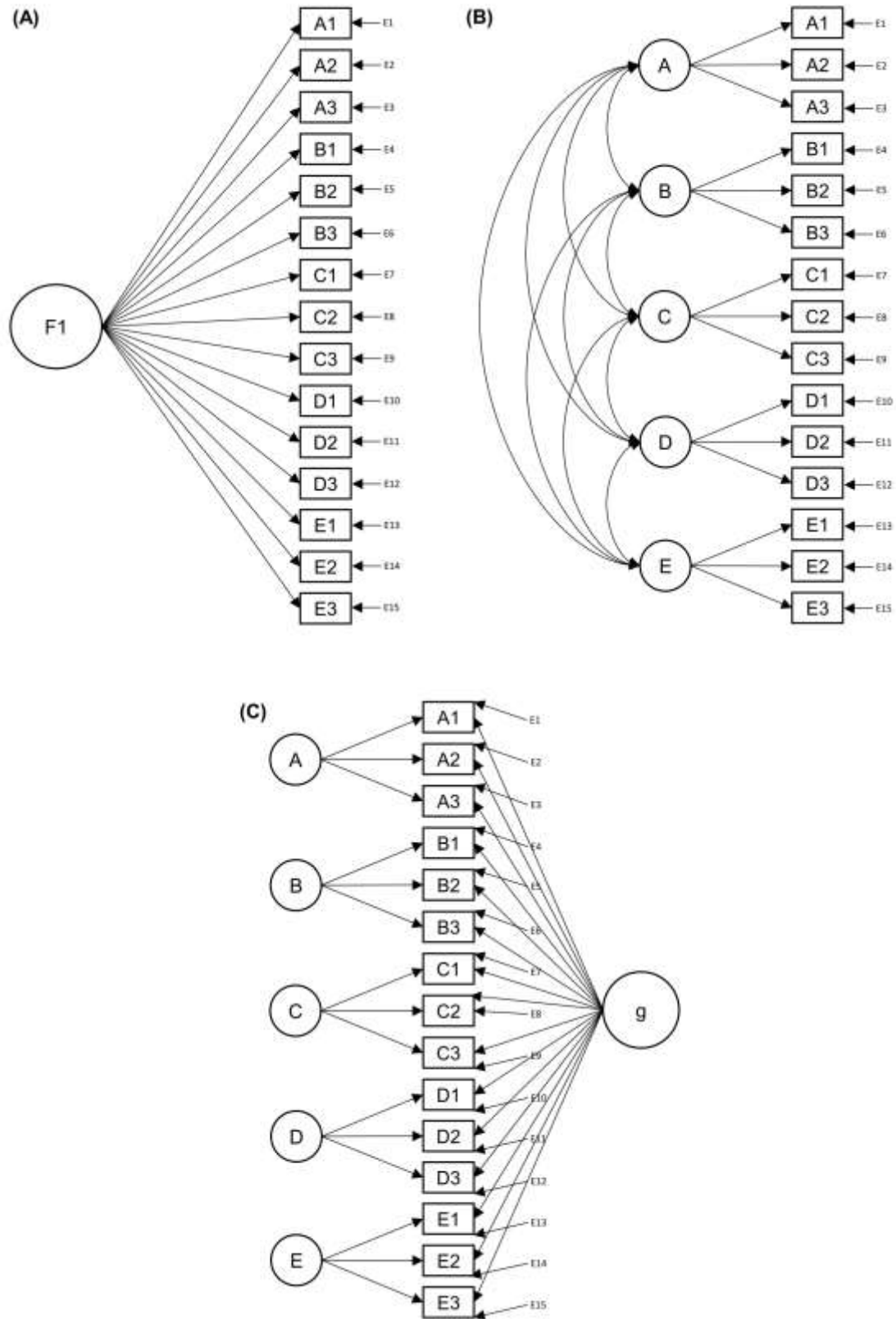


Abb. 4 Eindimensionales Modell (A), correlated-factors Modell (B) und hierarchisches Modell (C)

6 Ergebnisse

Zur Beantwortung der Fragestellung „Erklärt eine ein-, eine mehrdimensionale oder eine hierarchische Modellierung das Wissen über Lehr-Lernmethoden und -konzepte von Lehrkräften in der EB/WB besser?“ wurden die in Abschnitt 5.5 spezifizierten Modelle auf die Daten angepasst. Einerseits ein 1PL correlated-factors Modell und ein eindimensionales 1PL Modell, andererseits ein Bifaktor-Modell (Reise, 2012), ein Testlet-Modell (Bradlow et al., 1999) und ein Rasch-Testlet-Modell (Wang & Wilson, 2005b). Die Mehrzahl der informationstheoretischen Maße (AIC, AICc, CAIC und BIC) weisen auf die Überlegenheit des Rasch Testlet Modells gegenüber den anderen Modellen hin (siehe Tab. 1), der SABIC weist auf die Überlegenheit des 1PL correlated-factors Modell gegenüber den anderen Modellen hin.

Um ein möglichst vollständiges Bild der Ergebnisse darstellen zu können, werden nachfolgend für das 1PL correlated-factors Modell, das eindimensionale 1PL Modell und das Rasch Testlet Modell Itemschwierigkeiten und Standardfehler (der Teilaufgaben) sowie die Reliabilitäten abgebildet. Für alle mehrdimensionalen Modelle wird die Ladungsstruktur berichtet. Für das Bifaktor Modell und das Testlet Modell sind Itemschwierigkeiten, Standardfehler, die Ladungsstruktur sowie die Reliabilität im Anhang (Tab.7-9) zu finden.

Tab. 1 Modellvergleiche: Correlated-factors Modell, eindimensionales Modell und hierarchische Modelle (Bifaktor Modell, Testlet Modell und Rasch Testlet Modell)

<i>Correlated-factors Modell und eindimensionales Modell</i>							
Modell	Loglikeli- hood	Freie Parameter	AIC	AICc	BIC	SABIC	CAIC
1PL correlated- factors Modell	-1662	30	3384	3394	3484	3389	3514
Eindimen- sionales 1PL Modell	-1707	16	3447	3448	3501	3450	3515
<i>Hierarchische Modelle (Bifaktor Modell, Testlet Modell und Rasch Testlet Modell)</i>							
Modell	Loglikeli- hood	Freie Parameter	AIC	AICc	BIC	SABIC	CAIC
Bifaktor Modell	-1644	45	3379	3402	3602	3459	3574
Testlet Modell	-1658	30	3376	3386	3525	3429	3506
Rasch Testlet Modell	-1664	21	3371	3374	3476	3409	3461

Anmerkungen: Fette Markierung weist auf den niedrigsten Wert eines informationstheoretischen Maßes hin.

AIC = Akaike information criterion; AICc = bias corrected AIC; CAIC = consistent AIC; BIC = Bayesian information criterion; SABIC = sample-size adjusted BIC

Correlated-factors Modell und eindimensionales Modell

Tab. 2 standardisierte Ladungsstruktur und Itemschwierigkeiten des correlated-factors Modells und Itemschwierigkeiten des eindimensionalen 1PLModells

Teilaufgaben	1PL correlated-factors Modell					Eindimensionales 1PL Modell	
	A	B	C	D	E	Itemschwierigkeit in logits	Itemschwierigkeit in logits
A1	0.475** (0.070)					-1.105 (0.194)	-1.565 (0.292)
A2	0.475** (0.070)					-0.525 (0.174)	-0.737 (0.253)
A3	0.475** (0.070)					-1.865 (0.233)	-2.661 (0.382)
B1		0.553** (0.081)				-2.924 (0.310)	-3.981 (0.576)
B2		0.553** (0.081)				-1.904 (0.261)	-2.547 (0.405)
B3		0.553** (0.081)				-0.993 (0.205)	-1.320 (0.272)
C1			0.715** (0.046)			0.112 (0.220)	0.131 (0.235)
C2			0.715** (0.046)			-1.563 (0.259)	-1.693 (0.302)
C3			0.715** (0.046)			-1.222 (0.250)	-1.319 (0.285)
D1				0.507** (0.069)		-0.124 (0.176)	-0.165 (0.244)
D2				0.507** (0.069)		-0.562 (0.179)	-0.768 (0.257)
D3				0.507** (0.069)		-0.856 (0.187)	-1.168 (0.283)
E1					0.720** (0.062)	-4.115 (0.637)	-4.708 (0.641)
E2					0.720** (0.062)	0.997 (0.250)	1.091 (0.265)
E3					0.720** (0.062)	0.835 (0.243)	0.922 (0.265)

Anmerkungen: ** $p \leq .01$; Werte in Klammern sind Standardfehler

Die Ladungsstruktur des correlated-factors Modell sowie die Itemschwierigkeiten und Standardfehler auch des eindimensionalen Modells sind in Tab. 2 abgebildet. Für die fünf Faktoren zeigen sich Ladungsparameter zwischen .475 und .720, die alle statistisch signifikant sind. In Verbindung mit den (in Abschnitt 5.5 hypothetisierten) überwiegend geringen latenten Zusammenhängen zwischen den Faktoren (zwischen $r = .011$ und $r = .391$) – drei von zehn erwiesen sich als statistisch signifikant (siehe Tab. 3) – ist dies ein Hinweis darauf, dass die Annahme der Eindimensionalität des Wissens fraglich und Mehrdimensionalität wahrscheinlich erscheint. Um Eindimensionalität anzunehmen bzw. die Aufgaben zu einem Gesamtwert zusammenzufassen, müssten sich deutlich höhere (latente) Zusammenhänge zwischen den Faktoren zeigen, die sich nicht substantiell von 1 unterscheiden.

Die Itemschwierigkeit liegt für das correlated-factors Modell zwischen $-4.115 \text{ logits} \leq \text{Itemschwierigkeit} \leq 0.997 \text{ logits}$. Für das eindimensionale 1PL Modell liegt die Itemschwierigkeit zwischen $-4.708 \text{ logits} \leq \text{Itemschwierigkeit} \leq 1.091 \text{ logits}$.

Tab. 3 Korrelationsmatrix (standardisiert) des correlated-factors Modell

	Faktor A	Faktor B	Faktor C	Faktor D
Faktor B	0.391* (0.199)			
Faktor C	0.267 (0.155)	0.267 (0.137)		
Faktor D	0.011 (0.213)	0.062 (0.191)	0.317* (0.149)	
Faktor E	0.195 (0.192)	0.346* (0.161)	0.167 (0.142)	0.169 (0.147)

Anmerkungen: * $p \leq .05$. Werte in Klammern sind Standardfehler

Für das eindimensionale 1PL und das 1PL correlated-factors Modell wurde zur Ermittlung der internen Konsistenz der Teilaufgaben die Kuder-Richardson 20 Formula (KR-20) verwendet. KR-20 entspricht im Wesentlichen Cronbachs Alpha für dichotome Daten, wie sie hier gegeben sind (z.B. Lienert & Raatz, 1998). Für das eindimensionale Modell zeigte sich mit .56 eine niedrige interne Konsistenz (vgl. Tab. 4). Dies ist ein weiterer Hinweis darauf, dass die Eigenschaft der Eindimensionalität für den Test nicht vorliegt. Auch die interne Konsistenz der Faktoren A-E ist niedrig (zwischen .29 und .60), allerdings gilt es hier bei der Interpretation zu bedenken, dass lediglich drei Teilaufgaben pro Faktor in die Analysen gingen, was zu einer niedrigen internen Konsistenz beitragen kann.

Tab. 4 Reliabilitäten für das eindimensionale Modell und das correlated-factors Modell

Eindimensionales Modell	Correlated-factors Modell				
	Faktor A	Faktor B	Faktor C	Faktor D	Faktor E
KR-20 = .56	KR-20 = .29	KR-20 = .52	KR-20 = .60	KR-20 = .49	KR-20= 0.54

Rasch Testlet Modell

Die Ladungsstruktur und die Itemschwierigkeiten des Rasch Testlet Modells sind in Tab. 5 abgebildet. Die Itemschwierigkeit liegt zwischen $-4.066 \text{ logits} \leq \text{Itemschwierigkeit} \leq 0.987 \text{ logits}$. Es zeigen sich für den Generalfaktor (g) für alle Teilaufgaben statistisch signifikante Faktorladungen, ebenso für die fünf spezifischen Faktoren (A-E). Die Faktorladungen der spezifischen Faktoren sind durchweg höher als die des Generalfaktors. Diese Befunde weisen auch hier darauf hin, dass nicht von Eindimensionalität sondern vielmehr von einem Generalfaktor auszugehen ist.

Tab. 5 Ladungsstruktur (standardisiert) des Rasch Testlet Modells und Itemschwierigkeiten mit Standardfehler

	g	A	B	C	D	E	Itemschwierigkeit in logits
A1	0.289** (0.048)	0.374** (0.099)					-1.100 (0.192)
A2	0.289** (0.048)	0.374** (0.099)					-0.521 (0.172)
A3	0.289** (0.048)	0.374** (0.099)					-1.859 (0.233)
B1	0.277** (0.049)		0.459** (0.107)				-2.885 (0.310)
B2	0.277** (0.049)		0.459** (0.107)				-1.876 (0.256)
B3	0.277** (0.049)		0.459** (0.107)				-0.981 (0.202)
C1	0.234** (0.043)			0.661** (0.056)			0.114 (0.214)
C2	0.234** (0.043)			0.661** (0.056)			-1.538 (0.251)
C3	0.234** (0.043)			0.661** (0.056)			-1.203 (0.243)
D1	0.278** (0.045)				0.455** (0.080)		-0.119 (0.179)
D2	0.278** (0.045)				0.455** (0.080)		-0.566 (0.182)
D3	0.278** (0.045)				0.455** (0.080)		-0.867 (0.190)
E1	0.231** (0.045)					0.671** (0.075)	-4.066 (0.621)
E2	0.231** (0.045)					0.671** (0.075)	0.987 (0.247)
E3	0.231** (0.045)					0.671** (0.075)	0.830 (0.239)

Anmerkungen: ** $p \leq .01$; Werte in Klammern sind Standardfehler

Für den Generalfaktor und die spezifischen Faktoren stellt sich die Frage, zu welchen Anteilen diese die Varianz im Antwortverhalten aufklären. Um diese Frage zu beantworten, werden die Koeffizienten omega hierarchical (ω_h) und omega subscale (ω_s) herangezogen (siehe z.B. Reise, 2012; Revelle & Zinbarg, 2009). Bei ω_h werden die Effekte der spezifischen Faktoren (und des Messfehlers) kontrolliert, bei ω_s werden der Effekt des Generalfaktors sowie die Effekte der weiteren spezifischen Faktoren (und des Messfehlers) kontrolliert. ω_h ist ein wichtiger Indikator dafür, wie reliabel ein Test ein Konstrukt misst (Revelle & Zinbarg, 2009).

Für das Rasch Testlet Modell zeigte sich mit $\omega_h = .41$ für den Generalfaktor – das allgemeine Wissen über verschiedene Lehr-Lernmethoden hinweg – ein vergleichsweise geringer Wert (vgl. Tab. 6). Dies ist ein weiterer Hinweis darauf, dass die Eindimensionalität des Tests fraglich ist. ω_s indiziert, wie reliabel die spezifischen Faktoren – das spezifische Wissen über eine Lehr-Lernmethode – gemessen werden. Es zeigten sich mit ω_s zwischen .29 bis .67 niedrige bis befriedigende Werte, wobei bei der Interpretation zu berücksichtigen ist, dass lediglich drei Teilaufgaben pro Faktor in die Analysen eingingen, was zu einer niedrigen internen Konsistenz beitragen kann.

Omega total (ω_t) repräsentiert den Anteil der Varianz, der auf alle Faktoren zurückgeht. $\omega_t = .75$ zeigt an, dass 75% der Unterschiede im Antwortverhalten durch die insgesamt sechs Faktoren aufgeklärt werden und 25% Messfehler sind. Omega (ω) ist die Reliabilität der spezifischen Faktoren, allerdings *nicht* wie ω_s bereinigt um den Effekt des Generalfaktors (Reise, 2012). ω ist mit Werten zwischen .46 und .75 niedrig bis gut.

Tab. 6 Reliabilitäten Rasch Testlet Modell

	General-faktor	Gesamt-modell	Faktor A	Faktor B	Faktor C	Faktor D	Faktor E
Rasch Testlet Modell	$\omega_h = .41$	$\omega_t = .75$	$\omega = .46$	$\omega = .55$	$\omega = .74$	$\omega = .54$	$\omega = .75$
			$\omega_s = .29$	$\omega_s = .40$	$\omega_s = .66$	$\omega_s = .39$	$\omega_s = .67$

Anmerkungen. ω_h = Omega hierachical; ω_t = Omega total; ω = Omega; ω_s = Omega subscale

7 Zusammenfassung und Diskussion der Ergebnisse

Der vorliegende Beitrag widmete sich der Forschungsfrage, ob eine ein-, eine mehrdimensionale oder eine hierarchische Modellierung des Wissens über Lehr-Lernmethoden und -konzepte von EB/WB-Lehrkräften aus theoretisch-konzeptueller Sicht plausibler ist und empirisch besser auf die Daten passt. Diese Frage stellt sich vor dem Hintergrund einer großen Zahl an Lehr-Lernmethoden und -konzepten, die unterschiedlichen Zielen und Phasen innerhalb von Lehr-Lernprozessen dienlich sind, und einer vermutlich eher erfahrungsbasierten und unsystematischen Wissensaneignung durch mehrheitlich nicht

formal pädagogisch qualifizierte Lehrkräfte in der EB/WB. Ziel dieses Beitrags ist es, Hinweise auf die Weiterentwicklung des vorliegenden Tests – hier mit dem Fokus auf der Facette Wissen über Lehr-Lernmethoden und -konzepte – zu gewinnen.

In Kapitel 3 wurde eine ein-, eine mehrdimensionale und eine hierarchische Konzeptualisierung des Wissens über Lehr-Lernmethoden und -konzepte vorgestellt. Aus theoretisch-konzeptueller Perspektive erscheint die Annahme einer hierarchischen Struktur des Wissens über Lehr-Lernmethoden und -konzepte plausibel, die über verschiedene Lehr-Lernmethoden allgemeines Wissen und für einzelne Lehr-Lernmethoden spezifisches Wissen postuliert. Empirisch wurde überprüft, ob ein correlated-factors Modell oder ein eindimensionales Modell einerseits oder ein Bifaktor Modell, Testlet Modell oder Rasch Testlet Modell andererseits empirisch am besten auf die Daten passt. Annahmenkonform deuten vier von fünf informationstheoretischen Maßen auf die Überlegenheit einer hierarchischen Modellierung durch ein Rasch Testlet Modell hin. Eines der informationstheoretischen Maßen deutet auf die Überlegenheit des correlated-factors Modell hin. Mit Blick auf die Frage nach der Dimensionalität dieses Wissens geben beide Modelle Hinweise auf die Mehrdimensionalität des Wissens.

Für das Rasch Testlet Modell zeigte sich für den Generalfaktor – das allgemeine Wissen über verschiedene Lehr-Lernmethoden hinweg – ein geringes ω_h . Eine Erklärung hierfür kann sein, dass nur wenige Aufgaben zur Erfassung des Wissens verwendet wurden und dabei heterogene Themen angesprochen wurden. Dies ist ein Hinweis darauf, dass die zukünftige Berücksichtigung sowie Weiterentwicklung der Subskalen sinnvoll ist. Diese Befunde sind plausibel, wenn man in Rechnung stellt, dass Lehrkräfte der EB/WB über „verinselt“ Wissen verfügen vor dem Hintergrund der zumeist unsystematischen pädagogischen Ausbildung der EB/WB-Lehrkräfte. Angesichts dieses „verinselten“ Wissens, i.S. eines unkorrelierten Wissens aus Einzelaspekten über spezifische Lehr-Lernmethoden, erscheint die Berücksichtigung der Subskalen und die Weiterentwicklung dieser zu mehreren eindimensionalen Tests sinnvoll zu sein. Dies trifft umso mehr zu, wenn der Test im Rahmen einer Lernplattform für EB/WB-Lehrkräfte eingesetzt werden soll.

Für hierarchische Modelle stellt sich allerdings immer die Frage, ob bzw. welche Konstrukte im Generalfaktor und den spezifischen Faktoren modelliert werden. Ob der Generalfaktor Wissen über die Relationierungsnotwendigkeit von Sicht- und Tiefenstrukturen tatsächlich diese oder aber domänenunabhängige kognitive Fähigkeiten abbildet, kann in diesem Beitrag nicht abschließend beantwortet werden. Gleiches gilt für die Gruppenfaktoren: ob diese tatsächlich der Varianz des wahren Wertes zuzuordnen sind oder konstruktirrelevant sind, muss durch Validierungsstudien weiter plausibilisiert werden; erste

Hinweise, dass im Generalfaktor und den Gruppenfaktoren konstruktrelevante Varianz modelliert wird, gibt die Studie von Marx, Goeze, Kelava und Schrader (2018).

Bei Modellierung der Daten durch ein correlated-factors Modell zeigen sich bei zehn möglichen Zusammenhängen zwischen den Faktoren lediglich für drei statistisch signifikante Zusammenhänge. Dies kann als ein weiterer Hinweis dafür interpretiert werden, dass das in den Aufgaben thematisierte Wissen nicht eindimensional konzeptualisiert und erfasst werden sollte. Zudem zeigte sich bei einer eindimensionalen Modellierung eine geringe interne Konsistenz, was auf eine geringe Homogenität der Aufgaben hindeutet und ein Hinweis auf die Nützlichkeit/Notwendigkeit einer mehrdimensionalen Konzeptualisierung und Erfassung ist.

Alle Befunde einschränkend ist die insgesamt kleine Stichprobe zu nennen, die für die Studie zur Verfügung stand. Eine Konsequenz der kleinen Stichprobe ist, dass nur eine Auswahl an Aufgaben in die Analysen eingegangen ist, um die Anzahl an zu schätzenden Parametern in einem angemessenen Verhältnis zur Stichprobengröße zu halten (Bentler & Chou, 1987). Welche Struktur sich für eine größere Stichprobe und unter Einbezug weiterer Aufgaben und Facetten zeigt, ist offen.

Für die Testweiterentwicklung für den zukünftigen Einsatz in bildungsbereichsvergleichenden Studien oder in einer Lernplattform für EB/WB-Lehrkräfte, mit Hilfe dessen Lehrkräften schnell und präzise angemessene Lerninhalte entsprechend ihrer Fähigkeiten und Bedarfe zugewiesen werden können, erscheint die Entwicklung mehrerer eindimensionaler Tests zur Erfassung des Wissens über Lehr-Lernmethoden und -konzepte sinnvoll. Eine Herausforderung dabei wird die Vielzahl verschiedener Lehr-Lernmethoden- und konzepte sein. Eine Möglichkeit das Wissen über Lehr-Lernmethoden und -konzepte für Lehrkräfte in der EB/WB mehrdimensional zu konzeptualisieren ist, zentrale Lehr-Lernmethoden als jeweils eine Dimension dieses Wissens abzubilden. Welches die zentralen und am häufigsten verwendeten Lehr-Lernmethoden und -konzepte sind, ist Gegenstand einer Befragung von EB/WB-Lehrkräften, die derzeit vom Deutschen Institut für Erwachsenenbildung, Leibniz Zentrum für Lebenslanges Lernen e.V. (DIE) durchgeführt wird. Eine weitere Herausforderung wird sein, schwierigere Aufgaben zu generieren, um zwischen Personen mit viel Wissen über Lehr-Lernmethoden und -konzepte differenzieren zu können. Die vergleichsweise geringe Schwierigkeit des Tests kann einerseits durch das Antwortformat und andererseits durch den Verzicht auf Fachbegriffe im Test erklärt werden.

Der vorgestellte bzw. weiterentwickelte Test muss dann weiteren Validierungsschritten unterzogen werden, z.B. indem der Zusammenhang zwischen dem Wissen der Lehrkräfte vermittelt über die Qualität von Lehr-Lernsituationen auf den (Lern-) Erfolg von TeilnehmerInnen der Erwachsenen- und Weiterbildung untersucht wird. Zudem

bietet es sich für die Erwachsenen- und Weiterbildungsforschung an zu untersuchen, ob sich (Facetten des) PPK von pädagogischen Überzeugungen diskriminieren lassen bzw. in welchem Zusammenhang es mit diesen Überzeugungen, anderen Aspekten des professionellen Wissens oder mit der Berufserfahrung steht.

Insgesamt kann die vorliegende Arbeit einen Beitrag zur Entwicklung von Instrumenten zur Erfassung von Aspekten der professionellen Kompetenz von Lehrkräften in der EB/WB leisten, die im Rahmen bildungspolitischer und bildungspraktischer Initiativen benötigt werden. Bei konstruktinhärenter Interpretation der Gruppenfaktoren sowie des Generalfaktors kann eine Kurzskala zur Erfassung des Wissens über Lehr-Lernmethoden von EB/WB-Lehrkräften angeboten werden. Diese empfehlen wir allerdings – bevor weitere im Beitrag aufgezeigte Validierungsschritte durchgeführt wurden – lediglich für formatives Assessment und nicht für summative „high-stakes“ Assessments.

8 Literaturverzeichnis

- Autorengruppe wb-personalmonitor (Hrsg.). (2016). *Personal in der Weiterbildung: Beschäftigungsverhältnisse und Tätigkeiten – wb-personalmonitor*. Bielefeld: Bertelsmann.
- Bastian, H. (1997). *Kursleiterprofile und Angebotsqualität* (Theorie und Praxis der Erwachsenenbildung). Bad Heilbrunn: Klinkhardt.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29-54). Münster: Waxmann.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P. et al. (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.
- Bentler, P. M. & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16 (1), 78-117.
- Bonnes, C. & Hochholdinger, S. (2016). Die berufliche Entwicklung und die Wege des qualifikations- und Kompetenzerwerbs von Lehrenden in der berufsbezogenen Weiterbildung. *bwp@ Berufs- und Wirtschaftspädagogik - online* (29), 1-18. Verfügbar unter www.bwpat.de/ausgabe29/bonnes_hochholdinger_bwpat29.pdf.
- Bradlow, E., Wainer, H. & Wang, X. (1999). A bayesian random effects model for testlets. *Psychometrika*, 64 (2), 153-168.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research* (Methodology in the social sciences). New York: Guilford Press.
- Bundesministerium für Arbeit und Soziales. (2014). *Fortschrittsreport „Altersgerechte Arbeitswelt“. Ausgabe 4: Lebenslanges Lernen und betriebliche Weiterbildung*. Bonn: BMAS.
- Chen, F. F., West, S. G. & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41 (2), 189-225.
- Collins, J. B. & Pratt, D. D. (2011). The teaching perspectives inventory at 10 years and 100,000 respondents. *Adult Education Quarterly*, 61 (4), 358-375.
- DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, 13 (4), 354-378.

- Europäische Kommission. (2001). *Mitteilung der Kommission: Einen europäischen Raum des lebenslangen Lernens schaffen*. Verfügbar unter https://www.bibb.de/dokumente/pdf/foko6_neues-aus-euopa_04_raum-III.pdf
- Europäische Kommission. (2007). *Mitteilung der Kommission an den Rat, das europäische Parlament, den europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen*. Verfügbar unter <http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=LEGISSUM:c11102>
- Goeze, A. & Schneider, D. (2014). What creates and regulates access to the adult education profession? Research on recruiting practices. In S. Lattke & W. Jütte (Hrsg.), *Professionalisation of adult educators: International and comparative perspectives* (S. 201-213). Frankfurt am Main: Peter Lang.
- Haberzeth, E. (2010). *Thematisierungsstrategien im Vermittlungsprozess. Empirische Analysen zum Umgang mit Wissen im Planungsprozess von Weiterbildungsangeboten*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hippel, A. v. & Tippelt, R. (Hrsg.). (2009). *Fortbildung der Weiterbildungner/innen. Eine Analyse der Interessen und Bedarfe aus verschiedenen Perspektiven*. Weinheim: Beltz.
- Hof, C. (2001). *Konzepte des Wissens. Eine empirische Studie zu den wissenstheoretischen Grundlagen des Unterrichtens*. Bielefeld: Bertelsmann.
- Jank, W. & Meyer, H. (2008). *Didaktische Modelle* (10. Aufl). Berlin: Cornelsen Scriptor.
- Kade, J., Nolda, S., Dinkelaker, J. & Herrle, M. (Hrsg.). (2014). *Videographische Kursforschung. Empirie des Lehrens und Lernens Erwachsener*. Stuttgart: Kohlhammer.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (Fourth Edition). New York: The Guilford Press.
- König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 12 (3), 499-527.
- König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M. et al. (2017). Pädagogisches Wissen versus fachdidaktisches Wissen? Struktur des professionellen Wissens bei angehenden Deutsch-, Englisch- und Mathematiklehrkräften im Studium. *Zeitschrift für Erziehungswissenschaft*, 21 (3), 1-28.

- Kraft, S., Seitter, W. & Kollwe, L. M. (2009). *Professionalitätsentwicklung des Weiterbildungspersonals*. Bielefeld: Bertelsmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kuper, H. & Schrader, J. (in Druck). Wirkungen der Weiterbildung und der Steuerung von Weiterbildung. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, H. Solga et al. (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale*. Bad Heilbrunn: UTB.
- Lattke, S. & Jütte, W. (Hrsg.). (2014). *Professionalisation of adult educators: International and comparative perspectives*. Frankfurt am Main: Peter Lang.
- Lencer, S. & Strauch, A. (2016). *Das GRETA-Kompetenzmodell für Lehrende in der Erwachsenen- und Weiterbildung*, Deutsches Institut für Erwachsenenbildung. Verfügbar unter <http://www.die-bonn.de/doks/2016-erwachsenenbildung-02.pdf>
- Lenske, G., Thillmann, H., Wirth, J., Dicke, T. & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften. Evaluation des ProwiN-Tests. *Zeitschrift für Erziehungswissenschaft*, 18 (2), 225-245.
- Lenske, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., Liepertz, S. et al. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht. *Zeitschrift für Erziehungswissenschaft*, 19 (1), 211-233.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Linninger, C., Kunina-Habenicht, O., Emmenlauer, S., Dicke, T., Schulze-Stocker, F., Leutner, D. et al. (2015). Assessing teachers' educational knowledge. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47 (2), 72-83.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (Zeitschrift für Pädagogik, Beiheft. 51, S. 47-70). Weinheim u.a: Beltz.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58 (2), 103-117.

- Maier-Gutheil, C. (2012). Das Lernen Professioneller. Selbstbeobachtung als konstitutives Merkmal von Professionalitätsentwicklung. In H. v. Felden (Hrsg.), *Erwachsenenbildung und Lernen. Dokumentation der Jahrestagung der Sektion Erwachsenenbildung der Deutschen Gesellschaft für Erziehungswissenschaft vom 22.-24. September 2011 an der Universität Hamburg* (S. 131-142). Baltmannsweiler: Schneider Verlag Hohengehren.
- Marx, C., Goeze, A., Kelava, A. & Schrader, J. (2018). Lehrkräfte in der Erwachsenen- und Weiterbildung – Zusammenhänge zwischen Vorbildung und Erfahrung mit dem Wissen über Lehr-Lernmethoden und -konzepte. *Zeitschrift für Weiterbildungsforschung*, 41 (1), 57-77.
- Marx, C., Goeze, A. & Schrader, J. (2014). Pädagogisch-psychologisches Wissen zur Gestaltung von Lehr-Lernsituationen: (Wie) unterscheidet es sich in Erwachsenenbildung/Weiterbildung und Schule? *Hessische Blätter für Volksbildung*, 64 (3), 238-251.
- Marx, C., Goeze, A., Voss, T., Hoehne, V., Klotz, V. K. & Schrader, J. (2017). Pädagogisch-psychologisches Wissen von Lehrkräften aus Schule und Erwachsenenbildung. Entwicklung und Erprobung eines Testinstruments. *Zeitschrift für Erziehungswissenschaft*, 20 (S1), 165-200.
- Muthén, L. K. & Muthén, B. O. (1998-2015). *Mplus user's guide*. (Seventh Edition). Los Angeles, CA: Muthén & Muthén.
- Oser, F. & Baeriswyl, F. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Hrsg.), *Handbook of research on teaching* (Fourth Edition, S. 1031-1065). Washington, DC: American Educational Research Association.
- Peterßen, W. H. (2009). *Kleines Methoden-Lexikon* (2., aktualisierte Aufl., 1. Dr.). München: Oldenbourg.
- Pohl, S., Gräfe, L. & Rose, N. (2014). Dealing with omitted and not-reached items in competence tests. *Educational and Psychological Measurement*, 74 (3), 423-452.
- Reckase, M. (2009). *Multidimensional item response theory* (Statistics for social and behavioral sciences). New York: Springer.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47 (5), 667-696.
- Reise, S. P., Moore, T. M. & Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *Journal of personality assessment*, 92 (6), 544-559.

- Research voor Beleid. (2008). *ALPINE - Adult learning professions in Europe. A study of the current situation, trends and issues. Final report.* Zoetermeer.
- Revelle, W. & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74 (1), 145-154.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73 (2), 417-458.
- Robitzsch, A. (2016). Zu nichtignorierbaren Konsequenzen des (partiellen) Ignorierens fehlender Item Responses im Large-Scale Assessment. In B. Suchan, C. Wallner-Paschon & C. Schreiner (Hrsg.), *Die Kompetenzen in Lesen, Mathematik und Naturwissenschaften am Ende der Volksschule. Österreichischer Expertenbericht* (S. 55-64). Graz: Leykam.
- Rohs, M., Rott, K. J., Schmidt-Hertha, B. & Bolten, R. (2017). Medienpädagogische Kompetenzen von ErwachsenenbildnerInnen. *Magazin erwachsenenbildung.at*, 30, 1-14.
- Rose, N., Davier, M. von & Xu, X. (2010). *Modeling nonignorable missing data with item response theory (IRT). Research Report No. RR-10-11.* Princeton: Educational Testing Service.
- Rost, J. (2004). *Testtheorie – Testkonstruktion* (2., vollst. überarb. und erw. Aufl.). Bern: Huber.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Schläfli, A. & Sgier, I. (2008). *Porträt Weiterbildung Schweiz* (Länderporträts Weiterbildung). Bielefeld: Bertelsmann.
- Schöb, S., Rohs, M., Biel, C. & Scholze, T. (2016). Professionalisierung von Lehrhandeln in einer digital geprägten Lernkultur – Bedarf und Ansatzpunkte der Entwicklung einer Online-Lernumgebung. In O. Dörner, C. Iller, H. Pätzold & S. Robak (Hrsg.), *Differentielle Lernkulturen – regional, national, transnational* (S. 171-183). Opladen: Barbara Budrich.
- Schrader, J. (2010a). Fortbildung von Lehrenden in der Erwachsenenbildung: Notwendig? Sinnvoll? Möglich? Bedarf und Angebote im Überblick. In J. Schrader, R. Hohmann & S. Hartz (Hrsg.), *Mediengestützte Fallarbeit. Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern* (S. 25-70). Bielefeld: Bertelsmann.
- Schrader, J. (2010b). Reproduktionskontexte der Weiterbildung. *Zeitschrift für Pädagogik*, 56 (2), 267-284.

- Schrader, J., Hohmann, R. & Hartz, S. (Hrsg.). (2010). *Mediengestützte Fallarbeit. Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern*. Bielefeld: Bertelsmann.
- Schütz, J. & Nittel, D. (2012). Von der Heterogenität zur Vielfalt. Akademische Professionalisierung im Blick einer komparativen pädagogischen Berufsgruppenforschung. In R. Egetenmeyer & I. Schüßler (Hrsg.), *Akademische Professionalisierung in der Erwachsenenbildung / Weiterbildung* (Grundlagen der Berufs- und Erwachsenenbildung, Bd. 70, S. 229–244). Baltmannsweiler: Schneider Hohengehren.
- Seifert, A., Hilligus, A. H. & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstrumentes zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2 (1), 82-103.
- Siebert, H. (2012). *Didaktisches Handeln in der Erwachsenenbildung. Didaktik aus konstruktivistischer Sicht* (Grundlagen der Weiterbildung, 7., überarbeitete Aufl.). Augsburg: ZIEL.
- Stanik, T. (2016). Mikrodidaktische Planungen von Lehrenden in der Erwachsenenbildung. *Zeitschrift für Weiterbildungsforschung – Report*, 39 (3), 317-330.
- Statistisches Bundesamt. (2015a). *Bildung und Kultur: Allgemeinbildende Schulen. Schuljahr 2014/2015*, Statistisches Bundesamt. Fachserie 11: Reihe 1. Zugriff am 10.05.2016. Verfügbar unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100157004.pdf?__blob=publicationFile
- Statistisches Bundesamt. (2015b). *Bildung und Kultur: Berufliche Schulen. Schuljahr 2014/2015*, Statistisches Bundesamt. Fachserie 11: Reihe 2. Zugriff am 10.05.2016. Verfügbar unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/BeruflicheSchulen2110200127004.pdf?__blob=publicationFile
- Statistisches Bundesamt. (2016). *Bildung und Kultur: Personal an Hochschulen 2015.*, Statistisches Bundesamt. Verfügbar unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PersonalHochschulen2110440157004.pdf?__blob=publicationFile

- Steiner, P. (2010). Die Weiterbildungsakademie Österreich (wba). Erfahrungen und Zugänge zu Beratung und Kompetenzanerkennung für ErwachsenenbildnerInnen. *Magazin erwachsenenbildung.at* (14), 1-6.
- Strijbos, J.-W. & Müller, A. (2014). Personale Faktoren im Feedbackprozess. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 83-134). Münster: Waxmann.
- Vinepac-Project. (2008). *Handbook for the use of Validpack for the validation of psychopedagogical adult educators' competences*. Timisoara: Editura Mirton.
- Voss, T., Goeze, A., Marx, C., Hoehne, V., Klotz, V. K. & Schrader, J. (2017). Using digital media to assess and promote school and adult education teacher competence. In J. Buder & F. W. Hesse (Hrsg.), *Informational Environments* (S. 125-148). Cham: Springer International Publishing AG.
- 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.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103 (4), 952-969.
- Wainer, H. & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24 (3), 185-201.
- Wang, W.-C. & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement*, 29 (2), 126-149.
- Wecker, C. & Fischer, F. (2014). Lernen in Gruppen. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (S. 277-296). Weinheim: Beltz.
- Weidenmann, B. (2008). *Handbuch active Training. Die besten Methoden für lebendige Seminare* (2. Aufl.). Weinheim: Beltz.
- Ziep, K.-D. (1990). *Der Dozent in der Weiterbildung. Professionalisierung und Handlungskompetenzen*. Weinheim: Deutscher Studien Verlag.

Tab. 7 Ladungsstruktur (standardisiert) des Bifaktor Modells und Itemschwierigkeiten mit Standardfehler

	g	A	B	C	D	E	Itemschwierigkeit in logits
A1	0.115 (0.164)	0.581 (0.372)					-0.546 (0.112)
A2	0.213 (0.142)	0.241 (0.163)					-0.246 (0.091)
A3	0.245 (0.180)	0.510 (0.335)					-0.916 (0.119)
B1	0.768** (0.088)		0.638** (0.106)				-1.377 (0.066)
B2	0.430* (0.188)		0.273 (0.239)				-0.869 (0.114)
B3	0.031 (0.167)		0.553 (0.325)				-0.460 (0.097)
C1	0.488* (0.173)			0.536* (0.203)			0.043 (0.091)
C2	0.323 (0.211)			0.797** (0.246)			-0.615 (0.104)
C3	0.513* (0.191)			0.399* (0.175)			-0.467 (0.098)
D1	0.082 (0.167)				0.655 (0.341)		-0.058 (0.100)
D2	0.294 (0.164)				0.464* (0.227)		-0.268 (0.096)
D3	0.171 (0.147)				0.342* (0.165)		-0.396 (0.094)
E1	0.010 (0.298)					0.149 (0.302)	-1.568 (0.188)
E2	0.303 (0.177)					0.755 (0.682)	0.390 (0.103)
E3	0.253 (0.160)					0.808 (0.718)	0.329 (0.101)

Anmerkungen: * $p \leq .05$; ** $p \leq .01$; Werte in Klammern sind Standardfehler

Tab. 8 Ladungsstruktur (standardisiert) des Testlet Modells und Itemschwierigkeiten mit Standardfehler

	g	A	B	C	D	E	Itemschwierigkeit in logits
A1	0.268* (0.113)	0.268* (0.113)					-0.523 (0.086)
A2	0.238* (0.105)	0.238* (0.105)					-0.245 (0.080)
A3	0.357* (0.126)	0.357* (0.126)					-0.899 (0.100)
B1	0.670** (0.054)		0.670** (0.054)				-1.343 (0.123)
B2	0.397** (0.097)		0.397** (0.097)				-0.868 (0.100)
B3	0.183 (0.096)		0.183 (0.096)				-0.437 (0.084)
C1	0.547** (0.072)			0.547** (0.072)			0.042 (0.083)
C2	0.493** (0.064)			0.493** (0.064)			-0.589 (0.088)
C3	0.480 (0.073)			0.480 (0.073)			-0.459 (0.086)
D1	0.301** (0.103)				0.301** (0.103)		-0.058 (0.082)
D2	0.384** (0.108)				0.384** (0.108)		-0.264 (0.085)
D3	0.270** (0.102)				0.270** (0.102)		-0.394 (0.086)
E1	0.094 (0.141)					0.094 (0.141)	-1.568 (0.170)
E2	0.604** (0.071)					0.604** (0.071)	0.376 (0.089)
E3	0.547** (0.075)					0.547** (0.075)	0.312 (0.087)

Anmerkungen: * $p \leq .05$; ** $p \leq .01$; Werte in Klammern sind Standardfehler

Tab. 9 Reliabilitäten Bifaktor Modell und Testlet Modell

	General-faktor	Gesamtmodell	Faktor A	Faktor B	Faktor C	Faktor D	Faktor E
<i>Bifaktor Modell</i>	$\omega_h = .47$	$\omega_t = .78$	$\omega = .49$	$\omega = .71$	$\omega = .78$	$\omega = .53$	$\omega = .67$
			$\omega_s = .41$	$\omega_s = .42$	$\omega_s = .49$	$\omega_s = .47$	$\omega_s = .60$
<i>Testlet Modell</i>	$\omega_h = .67$	$\omega_t = .81$	$\omega = .37$	$\omega = .64$	$\omega = .76$	$\omega = .43$	$\omega = .65$
			$\omega_s = .19$	$\omega_s = .32$	$\omega_s = .38$	$\omega_s = .22$	$\omega_s = .33$

Anmerkungen. ω_h = Omega hierachical; ω_t = Omega total; ω = Omega; ω_s = Omega subscale

8 Studie 3 Lehrkräfte in der Erwachsenen- und Weiterbildung: Zusammenhänge zwischen Vorbildung und Erfahrung mit dem Wissen über Lehr-Lernmethoden und -konzepte

Die folgende Version des Manuskripts wurde am 02.03.2018 eingereicht bei der Zeitschrift für Weiterbildungsforschung. Das Manuskript wurde noch nicht angenommen und publiziert. Die in dieser Dissertationsschrift abgebildete Version wird voraussichtlich nicht mit der publizierten Version übereinstimmen. Das Manuskript wurde mit der folgenden Autorenanreihung eingereicht:

Marx, C., Goeze, A., Kelava, A. & Schrader, J. (eingereicht). Lehrkräfte in der Erwachsenen- und Weiterbildung: Zusammenhänge zwischen Vorbildung und Erfahrung mit dem Wissen über Lehr-Lernmethoden und -konzepte. *Zeitschrift für Weiterbildungsforschung*.

Zusammenfassung

Das Wissen über Lehr-Lernmethoden und -konzepte gilt als ein wichtiger Aspekt des pädagogisch-psychologischen Wissens von Lehrkräften in der Erwachsenen- und Weiterbildung (EB/WB) für die Gestaltung von Lehr-Lernsituationen. Der Beitrag untersucht, ob und inwiefern dieses Wissen mit der Vorbildung und der Erfahrung von $N = 212$ Lehrkräften aus der EB/WB zusammenhängt. Die Ergebnisse weisen darauf hin, dass der Umfang an Fort- und Weiterbildungen zum Thema Lehren und Lernen / Methodik / Didaktik und die Tätigkeit in der beruflich-betrieblichen Weiterbildung in einem positiven Zusammenhang mit dem erfassten Wissen stehen. Keine (direkten) Zusammenhänge zu diesem Wissen zeigten sich entgegen der Annahmen zur Lehrerfahrung sowie einem absolvierten erziehungs- und bildungswissenschaftlichen Studium. Die Befunde werden an den aktuellen Diskurs der EB/WB angeschlossen.

Schlüsselwörter

Pädagogisch-psychologisches Wissen; Lehr-Lernmethoden und -konzepte; Lehrkräfte; Trainer; Erwachsenenbildung/Weiterbildung

1 Einleitung: Zur Entwicklung professioneller Kompetenz von Lehrkräften in der Erwachsenen- und Weiterbildung als Forschungsdesiderat

Befragungsdaten aus der Erwachsenen- und Weiterbildung (EB/WB) (z.B. Hartz, 2011), Expertisen für die Politik (Guerriero, 2017; Kraft et al., 2009) und auch empirische Studien aus dem Schulbereich (Rivkin, Hanushek & Kain, 2005) verweisen auf den großen Einfluss des pädagogischen Handelns von Lehrkräften auf den Lernerfolg der Lernenden. Demgegenüber wird den institutionellen Rahmenbedingungen bildungsbereichsunabhängig zumeist deutlich weniger Einfluss zugeschrieben. Folglich wuchs in der EB/WB in den vergangenen Jahren das Interesse an der Erforschung einzelner Bestandteile der wie folgt angenommenen Wirkungskette (Terhart, 2012): (1) Aus- und Weiterbildung von Lehrkräften führen zu (2) professioneller Kompetenz und darüber vermittelt zu (3) hoher Qualität bei Lehr-Lernprozessen, die (4) den Lernerfolg bei Teilnehmenden befördert.

An der konkreten Aus- bzw. Weiterbildung von Lehrkräften in der EB/WB setzen zahlreiche Initiativen für die Praxis an (Avramovska, Czerwinski & Lattke, 2015; Lencer & Strauch, 2016; Schläfli & Sgier, 2008; Schöb et al., 2016; Steiner, 2010), verschiedene Bereiche der professionellen Kompetenz von Lehrkräften sind in den Fokus empirischer Forschung gerückt (Avramovska et al., 2015; Hof, 2010; Lattke & Jütte, 2014; Pachner, 2013; Rohs et al., 2017; Stanik, 2016; Vinepac-Project, 2008) ebenso konkrete Lehr-Lernsituationen (Herrle, Dinkelaker, Nolda & Kade, 2014) sowie die Kompetenzen der Teilnehmenden (Rosenblatt & Lehmann, 2013). Bislang gibt es allerdings nur wenig Forschung in der EB/WB, die einzelne Glieder in ihrer Interdependenz untersucht (Goeze, Hetfleisch und Schrader, 2013).

Betrachtet man die ersten beiden Glieder der Kette, so kann man – anders als im Schulkontext, aber ähnlich dem Hochschulkontext – relativ wenig systematische und institutionalisierte Einflussnahme auf die Entwicklung der professionellen Kompetenz von Lehrkräften in der EB/WB konstatieren. Weil der Ausbildungs- und Berufszugang sowie die Berufsausübung hier nur in geringem Maße staatlich, tarifvertraglich oder berufsständisch reglementiert sind und damit seit jeher „offen“ für Lehrkräfte mit unterschiedlichem Ausbildungshintergrund, bleibt die Entwicklung von professioneller Kompetenz eine (größtenteils) individuelle „berufs-“biographische Entscheidung jeder einzelnen Lehrkraft und ist daher vergleichsweise schwer zu untersuchen. Umso mehr ist es von Interesse für Forschung und Praxis, welche Vorbildungen und Erfahrungen mit der professionellen Kompetenz(entwicklung) der Lehrkräfte („on the job“) in Zusammenhang stehen könnten.

Das Wissen um die berufsbiographische, berufliche und soziale Situation von EB/WB-Lehrkräften konnte in den vergangenen Jahren durch mehrere Studien zum Personal in der Weiterbildung deutlich verbessert werden (Autorengruppe Bildungsberichterstattung, 2016;

Autorengruppe wb-personalmonitor, 2016; Dobischat et al., 2018). Dabei wurden grundlegende Merkmale sowie deren typische Ausprägungen identifiziert, von denen angenommen werden darf, dass sie die Entwicklung der professionellen Kompetenz von Lehrkräften und in der Folge die Gestaltung von Lehr-Lernsituationen begünstigen könnten.

Neben den Überzeugungen von Lehrkräften (z.B. Collins & Pratt, 2011), die relativ zeitstabil sind (z.B. zusammenfassend Skott, 2015), wird ein Bereich der professionellen Kompetenz von Lehrkräften derzeit innerhalb der EB/WB verstärkt thematisiert: Das vergleichsweise gut entwickelbare professionelle Wissen von Lehrkräften (Marx et al., 2017; Rohs et al., 2017). Welche ausbildungs- und tätigkeitsbezogenen Merkmale für die Lehrtätigkeit einschlägige Lern- und damit Kompetenzentwicklungsgelegenheiten darstellen können (z.B. ein erziehungs- und bildungswissenschaftliches Studium, Fort- und Weiterbildungen, Tätigkeitsumfang und Lehrerfahrung sowie die Tätigkeitsschwerpunkte) und wie diese mit dem professionellen Wissen von EB/WB-Lehrkräften zusammenhängen, wurde bislang noch nicht untersucht. Ein Grund hierfür kann sein, dass es an standardisierten Instrumenten insgesamt bislang mangel(t)e, (siehe aber das Teaching Perspectives Inventory (Collins & Pratt, 2011; Vinepac-Project, 2008). Insbesondere Instrumente zur Erfassung (einzelner Facetten) des professionellen Wissens sind erst seit jüngster Zeit in der Entwicklung und Validierung (Marx et al., 2017; Rohs et al., 2017).

Ein im Rahmen der genannten Lern- und Kompetenzentwicklungsgelegenheiten vielthematisierter Bereich des professionellen Wissens von Lehrenden ist das *Wissen über Lehr-Lernmethoden und -konzepte* (Schrader et al., in Druck), das für die Gestaltung von Lehr-Lernprozessen in der EB/WB allgemein als zentral betrachtet wird und seit jeher zum Kernbestand der didaktischen Diskussion in der EB/WB gehört (Arnold et al., 2005; Kraft et al., 2009; Tietgens, 1967). Die Bedeutsamkeit dieser Wissensfacette findet Ausdruck in der Häufigkeit, mit der sie in print (siehe Hoffmann, 2011) und digitaler Form (z.B. wb-web.de) sowie in einschlägigen Fort- und Weiterbildungen betont wird. Daher bietet sie sich für die Untersuchung der Frage an, ob und inwiefern dieses Wissen mit den verschiedenen genannten ausbildungs- und tätigkeitsbezogenen Merkmalen der Lehrkräfte zusammenhängt.

Diese Frage greift der vorliegende Beitrag auf und bearbeitet damit ein Desiderat in der Erforschung der eingangs vorgestellten Wirkungskette (Terhart, 2012); die hier dargestellte Untersuchung stellt zudem einen weiteren Validierungsschritt eines im Projekt ThinkK entwickelten Tests zur Erfassung des Wissens über Lehr-Lernmethoden und -konzepte dar (vgl. Kapitel 4.3).

2 Das Wissen über Lehr-Lernmethoden und -konzepte als Facette des pädagogisch-psychologischen Wissens von Lehrkräften

Im Rahmen der Forschungen zur professionellen Kompetenz von Lehrkräften in der EB/WB wird das ursprünglich für den Schulkontext entwickelte und dort empirisch bewährte Rahmenmodell der professionellen Kompetenz von Lehrkräften von Baumert und Kunter (2011) vermehrt auch für den Bereich der EB/WB adaptiert (Lencer & Strauch, 2016; Marx et al., 2017; Rohs et al., 2017), weil es generische und damit auch für die EB/WB relevant erscheinende Bereiche der professionellen Kompetenz thematisiert. Grund für die Verwendung dieses Modells ist, dass für den Bereich der EB/WB zwar zahlreiche Konzeptualisierungen von Teilbereichen der professionellen Kompetenz von Lehrkräften entwickelt wurden (z.B. Vinepac-Projekt, 2008; Ziep, 1990), diesen allerdings kein (gemeinsames) theoretisches Rahmenmodell zugrunde liegt, welches empirisch überprüft wäre (Marx et al., 2014a).

Das Modell der professionellen Kompetenz von Lehrkräften nach Baumert und Kunter (2011) unterscheidet basierend auf der Kompetenzdefinition von Weinert (2001) in Überzeugungen und Werthaltungen, Selbstregulation und motivationale Orientierungen und integriert die Unterscheidung des professionellen Wissens in Fachwissen, fachdidaktisches Wissen und pädagogisch-psychologisches Wissen (Shulman, 1986). Dieses Rahmenmodell liegt auch dem Projekt ThinK zugrunde, in dem das Wissen über Lehr-Lernmethoden und -konzepte als Teil des pädagogisch-psychologischen Wissens (PPK) konzeptualisiert wird und ein Instrument zur fach- und bildungsbereichsübergreifenden Erfassung des PPK von Lehrkräften verschiedener Bildungsbereiche entwickelt wird (Marx et al., 2017; Marx et al., 2014b; Voss et al., 2017; Voss et al., 2015).

3 Fragestellung und Hypothesen zum Zusammenhang von ausbildungs- und tätigkeitsbezogenen Merkmalen und dem professionellen Wissen von EB/WB-Lehrkräften

Im Folgenden sollen ausbildungs- und tätigkeitsbezogenen Merkmale vorgestellt werden, die daraufhin beleuchtet werden, inwiefern sie im Zusammenhang zum professionellem Wissen, genauer zum Wissen über Lehr-Lernmethoden und -konzepte als eine Facette des PPK, stehen. Diese Ausführungen begründen die Hypothesenbildung am Ende jedes Abschnitts und die Fragestellung der Untersuchung: In welchem Zusammenhang stehen die im Folgenden erläuterten ausbildungs- und tätigkeitsbezogenen Merkmale mit dem Wissen über Lehr-Lernmethoden und -konzepte?

3.1 Merkmal: Erziehungs- und bildungswissenschaftliches Studium

Die berufsbiographisch betrachtet zumeist erste Möglichkeit zur Entwicklung professioneller Lehrkompetenz sind einschlägige *erziehungs- und bildungswissenschaftliche Studiengänge*, z.B. mit dem Schwerpunkt EB/WB, Wirtschaftspädagogik oder ein Lehramtsstudium mit anschließendem Referendariat. Auf den ersten Blick erscheint es plausibel, dass AbsolventInnen solcher Studiengänge, die ca. ein Drittel der Lehrkräfte in der EB/WB ausmachen (Autorengruppe wb-personalmonitor, 2016; Schütz & Nittel, 2012), auf eine Lehrtätigkeit durch entsprechende Studieninhalte besser vorbereitet sind als AbsolventInnen anderer Studiengänge. Allerdings werden Studierende dieser Studiengänge – abgesehen von den Schul-Lehrkräften – nicht ausschließlich und vornehmlich für eine Lehrtätigkeit ausgebildet, sondern im Falle der EB/WB-Studierenden für (die bei hauptberuflich in der EB/WB Beschäftigten zentralen) Tätigkeiten des Weiterbildungsmanagements oder der Programmplanung (Wißhak & Hochholding, 2016). Dennoch sollte die Auseinandersetzung mit Inhalten der (Allgemeinen) Didaktik, selbst wenn dies nur in grundlegenden Veranstaltungen der Fall ist, aus einer theoretisch-konzeptuellen Perspektive für die Entwicklung des PPK dienlich sein.

Hinweise darauf, dass berufstätige und damit durch erstes und zweites Staatsexamen ausgebildete Schul-Lehrkräfte verglichen mit in der Regel nicht systematisch für eine Lehrtätigkeit ausgebildeten EB/WB-Lehrkräfte ein größeres Wissen in Facetten des PPK aufweisen, finden sich bei Marx et al. (2017). Zudem konnte für Schul-Lehrkräfte gezeigt werden, dass das professionelle Wissen am Ende des Studiums höher war als zu Beginn (Sorge et al., 2017).

Hypothese 1: Vor diesem Hintergrund wird auch für die EB/WB-Lehrkräfte angenommen, dass sich Unterschiede zwischen EB/WB-Lehrkräften mit und ohne erziehungs- und bildungswissenschaftlichem Studium hinsichtlich ihres PPK zeigen, hier operationalisiert am Wissen über Lehr-Lernmethoden und -konzepte: Lehrkräfte mit einem erziehungs- und bildungswissenschaftlichen Studium sollten über mehr Wissen verfügen als Lehrkräfte ohne ein solches Studium.

3.2 Merkmal: Fort- und Weiterbildungen für Lehrkräfte

Die Mehrzahl der Lehrkräfte beginnt ohne ein einschlägiges Studium ihre Lehrtätigkeit in der EB/WB (Autorengruppe wb-personalmonitor, 2016). Für diese Gruppe sind dann – über das Selbststudium und das Beobachtungslernen als Seminarteilnehmende/r hinaus – Fort- und Weiterbildungsmöglichkeiten sowie die konkrete Lehrtätigkeit die ersten Lern- und Kompetenzentwicklungsgelegenheiten hinsichtlich ihrer Lehrtätigkeit.

Die Vielfalt an *Fort- und Weiterbildungen* ist groß und reicht vom kollegialen Gespräch bis hin zu zertifizierten Trainerausbildungen (Hohmann, 2010). Blickt man auf die Nutzung

der non-formalen Angebote für EB/WB-Lehrkräfte, so zeigt sich für das Weiterbildungspersonal, welches im Bereich der Lehre bzw. Training oder Coaching tätig ist, eine Beteiligungsquote an Weiterbildung von 81.8 Prozent (Autorengruppe wb-personalmonitor, 2016). Im Durchschnitt nahmen die EB/WB-Lehrkräfte im Zeitraum eines Jahres knapp 55 Stunden an Weiterbildungsveranstaltungen teil (Autorengruppe wb-personalmonitor, 2016), wobei allerdings unklar ist, welche Inhalte in diesen thematisiert wurden und ob diese überhaupt auf die Lehrtätigkeit bezogen sind.

Fort- und Weiterbildungsangebote für EB/WB-Lehrkräfte zum Thema Lehren und Lernen / Methodik / Didaktik sind in der Regel spezifischer auf die Lehrtätigkeit und die Bedürfnisse der Lehrkräfte in der EB/WB ausgerichtet, als dies die Mehrzahl der erziehungs- und bildungswissenschaftlichen Studiengänge sein können (Wißhak, 2017). Zudem liegen empirische Befunde vor, die zeigen, dass einschlägige Fortbildungen das Wissen und die Kognitionen von Lehrkräften positiv verändern können (Digel, 2013; Goeze, 2016).

Hypothese 2: Für das Wissen über Lehr-Lernmethoden und -konzepte wird ein positiver Zusammenhang zu Fort- und Weiterbildungen angenommen: Je mehr Stunden an spezifischer Fort- und Weiterbildung zum Thema Lehren und Lernen / Methodik / Didaktik absolviert wurden, über desto mehr Wissen zu Lehr-Lernmethoden und -konzepten verfügen EB/WB-Lehrkräfte.

3.3 Merkmale: Lehrerfahrung und Lehrumfang

Von praktischen Lehrerfahrungen wird angenommen, dass sie beim Erwerb professioneller Kompetenz, die sich auch aus Erfahrungswissen speist, eine wichtige Rolle spielen (Voss et al., 2011). Eigene Lehrerfahrungen sind die Voraussetzung dafür, dass das z.B. in Fort- und Weiterbildungen oder im Selbststudium erworbene theoretisch-konzeptuelle Wissen angewendet werden kann und nicht träge bleibt (grundlegend z.B. Gruber, Mandl & Renkl, 2000). Insgesamt liegen bislang nur wenige Studien vor, die den Einfluss der Lehrerfahrung auf die professionelle Kompetenz untersucht haben (Voss et al., 2015) oder gar der Bedeutung der Lehrerfahrung für den Lernfortschritt der Lernenden (Rivkin, Hanushek & Kain, 2005). Dieses Erfahrungswissen dürfte bei längerer Lehrerfahrung bzw. bei höherem aktuellem Lehrumfang zu mehr Auseinandersetzungen mit Lehr-Lernmethoden und -konzepten und damit zu mehr informell erworbenem Wissen in diesem Bereich geführt haben.

Hypothesen 3a und 3b: Für den aktuellen Lehrumfang in Wochenstunden (H3a) sowie die Lehrerfahrung in Jahren (H3b) wird daher ein positiver Zusammenhang zum Wissen über Lehr-Lernmethoden und -konzepte angenommen: Je höher der aktuelle Lehrumfang in Wochenstunden bzw. je mehr Lehrerfahrung in Jahren, desto mehr Wissen über Lehr-Lernmethoden und -konzepte zeigt sich bei den EB/WB-Lehrkräften.

Während die Ausbildung von Schul-Lehrkräften mit Praktikum und Referendariat (mindestens zwei) Praxisphasen vorsieht, deren Verzahnung mit „Theoriephasen“ als bewährt gilt (z.B. European Commission, 2015), wird die Verzahnung von Theorie und Praxis bei Lehrkräften in der EB/WB häufig berufsbegleitend realisiert: einerseits durch die praktische Lehrtätigkeit und andererseits durch den Besuch einschlägiger Fort- und Weiterbildungen. Im Vergleich zur Ausbildung von Schul-Lehrkräften fehlt Lehrkräften in der EB/WB häufig weniger die praktische Lehrerfahrung, sondern eher die theoretisch-konzeptuelle Wissensgrundlage zu didaktischen Konzepten und Prinzipien des Lehrens und Lernens.

Hypothese 4: Vor diesem Hintergrund wird ein Interaktionseffekt zwischen Lehrerfahrung und dem Besuch einschlägiger Fort- und Weiterbildungen erwartet: Lehrkräfte mit wenig Lehrerfahrung dürften stärker von zusätzlichen Stunden absolvierter Fort- und Weiterbildung hinsichtlich des Wissens profitieren als Lehrkräfte mit viel Lehrerfahrung. Für Lehrkräfte mit weniger Lehrerfahrung in Jahren wird also ein stärkerer Zusammenhang zwischen den absolvierten Stunden an Fort- und Weiterbildung und dem Wissen über Lehr-Lernmethoden und -konzepten erwartet als für Lehrkräfte mit mehr Lehrerfahrung.

Der angenommene Interaktionseffekt impliziert auch, dass Lehrkräfte, die weniger Stunden in Fort- und Weiterbildung absolviert haben, hinsichtlich ihres Wissens stärker von zusätzlichen Jahren an Lehrerfahrung profitieren als Lehrkräfte, die viele Stunden in Fort- und Weiterbildung absolviert haben: Für Lehrkräfte mit weniger absolvierten Fort- und Weiterbildungsstunden wird ein stärkerer Zusammenhang zwischen den Jahren an Lehrerfahrung und dem Wissen über Lehr-Lernmethoden und -konzepten erwartet als für Lehrkräfte mit mehr absolvierten Stunden in Fort- und Weiterbildung.

3.4 Merkmal: Schwerpunkt der Lehrtätigkeit in der allgemeinen Erwachsenenbildung oder in der beruflich-betrieblichen Weiterbildung

Ein weiterer Einfluss auf die Entwicklung der professionellen Kompetenz der Lehrkräfte kann deren *Tätigkeitsschwerpunkt* sein. Eine gängige Typologie zur Unterscheidung von Bildungsangeboten differenziert nach der inhaltlichen Ausrichtung der Bildungsangebote in allgemeine Erwachsenenbildung und beruflich-betriebliche Weiterbildung wie sie z.B. im wb-personalmonitor vorgenommen wurde (Autorengruppe wb-personalmonitor, 2016).

Die Erwartungshaltung der Teilnehmenden in der beruflich-betrieblichen WB an das lehr-lernmethodische Wissen der Lehrkräfte angesichts des eigenen potentiellen „Konsequenzenreichtums“ bei möglichem ausbleibenden Lernerfolg im beruflichen Kontext sowie die höhere finanzielle Vergütung der TrainerInnen in der beruflich-betrieblichen WB (im Vergleich zur allgemeinen EB) könnten eine höhere „Anspruchshaltung“ der Teilnehmenden in diesem Weiterbildungsbereich begründen. Die Folge könnte sein, dass die

TrainerInnen sich stärker selbständig mit methodischem und didaktischen Wissen auseinandersetzen, um die Lehr-Lernveranstaltungen zur Zufriedenheit der Teilnehmenden gestalten zu können. Dies scheint insbesondere dann plausibel, wenn die Teilnehmerzufriedenheit Geschäfts- und Existenzgrundlage ist für die TrainerInnen, wie es in der beruflich-betrieblichen WB häufig der Fall ist im Vergleich zur den meist nebenberuflich ausgeübten Lehrtätigkeiten in der allgemeinen EB. Insofern könnten die allgemeine EB und die beruflich-betriebliche WB „differenzielle Entwicklungsmilieus“ (Neumann et al., 2007) für das Lehrkräfte-Wissen darstellen.

Hypothese 5: Vor diesem Hintergrund wird folgender Unterschied angenommen: Im Schwerpunkt in der beruflich-betrieblichen WB tätige Lehrkräfte verfügen über mehr Wissen über Lehr-Lernmethoden und -konzepte als schwerpunktmäßig in der allgemeinen EB tätige Lehrkräfte.

4 Methodisches Vorgehen

4.1 Stichprobe

An der Hauptstudie des Projekts Think nahmen insgesamt $N = 212$ EB/WB-Lehrkräfte teil, dabei handelte es sich um eine ad-hoc-Stichprobe. Für den vorliegenden Beitrag wurden drei Fälle aufgrund von extremen Ausreißern in den Prädiktorvariablen ausgeschlossen, so dass für die nachfolgenden Analysen $N = 209$ EB/WB-Lehrkräfte zur Verfügung standen. Um Personen mit fehlenden Werten auf einzelnen Variablen nicht aus den Analysen ausschließen zu müssen, wurde in Mplus (Muthén & Muthén, 1998-2015) die Funktion full information maximum likelihood (FIML) verwendet, da ein listenweiser Fallausschluss neben einem Verlust an Effizienz bei der Parameterschätzung auch zu fehlerhaften Befunden führen kann (Lüdtke, Robitzsch, Trautwein & Köller, 2007). Die Lehrkräfte im Alter von 24 bis 77 Jahren ($M = 48.39$; $SD = 12.19$) verfügten zum Erhebungszeitpunkt über eine Lehrerfahrung zwischen 1 und 50 Jahren ($M = 12.70$; $SD = 10.44$) und lehrten pro Woche zwischen 1 und 50 Stunden ($M = 11.97$; $SD = 11.06$). Der Anteil an Akademikern (73.54%), die mehrheitlich nebenberufliche Beschäftigung (63.02%) bei mehreren Bildungseinrichtungen, der Anteil an weiblichen Lehrkräften (48.44%) sowie der Anteil an Personen mit einem erziehungs- und bildungswissenschaftlichen Studium (25.84%) weisen die vorliegende Stichprobe als typisch für die EB/WB aus (Autorengruppe wb-personalmonitor, 2016). Die Lehrkräfte nahmen insgesamt zwischen 0 und 520 Stunden an Weiterbildungen zum Thema Lehren und Lernen / Methodik / Didaktik teil ($M = 95.85$; $SD = 134.18$). Die Anteil der Lehrkräfte in der allgemeinen EB und der beruflich-betrieblichen WB war nahezu ausgewogen (EB: 46.32%). Nach den ausdifferenziert abgefragten Themenschwerpunkten (z.B. IT mit berufs- oder nicht berufsbezogenem Schwerpunkt) konnten die Lehrkräfte den beiden Bereichen zugeordnet werden.

4.2 Erfassung des Wissens über Lehr-Lernmethoden und -konzepte: Aufgaben

Das Wissen über Lehr-Lernmethoden und -konzepte wurde im Projekt Think durch (vorläufig) fünf Aufgaben erfasst (Marx et al., eingereicht). Diese Aufgaben lagen im gleichen Aufgabenformat vor wie die Beispielaufgabe in Abb. 1 und umfassen bis zu fünf Antwortoptionen im Richtig-Falsch-Antwortformat. Die Richtig-Falsch-Antworten hatten einen gemeinsamen Aufgabenstamm und ähnelten damit Testlets (Wainer & Kiely, 1987). Die fünf Aufgaben hatten ihren inhaltlichen Schwerpunkt auf fünf verschiedenen lehr-lernmethodischen Vorgehensweisen. Eine Aufgabe (A) thematisierte die Anstrengungsbereitschaft bei Gruppenarbeiten, (B) die Vortragsmethode (vgl. Abb. 1), (C) die Vermeidung aufgabenbedingter Trittbrettfahreneffekte bei Gruppenarbeiten, (D) die Feedback-Methode und (E) die Aktivierung von Vorwissen in Lehr-Lernsituationen.

Bitte beurteilen Sie, welche der folgenden Aussagen stimmen oder nicht stimmen.		
<i>Kreuzen Sie bitte in jeder Zeile ein Kästchen an.</i>		
Der Lehrervortrag (auch als Vortragsmethode bezeichnet) ist dann geeignet, wenn...	Stimmt nicht	Stimmt
... das Hauptziel in der Vermittlung von Informationen besteht.		
... eine Einführung in einen bestimmten Bereich gegeben werden soll.		
... das Behalten des Stoffs über einen längeren Zeitraum erwünscht ist.		
... die Beteiligung der Lernenden eine wesentliche Voraussetzung ist, um ein Lernziel zu erreichen.		
... der Stoff für die Lernenden komplex ist oder eine Menge Details enthält.		

Abb. 1 Beispiel-Aufgabe zur Erfassung des Wissens über Lehr-Lernmethoden und -konzepte

4.3 Erfassung des Wissens über Lehr-Lernmethoden und -konzepte: Vorbereitende Validierungsschritte

Da die fünf Aufgaben aus theoretisch-konzeptueller Perspektive zwar einerseits alle der Wissensfacette „Wissen über Lehr-Lernmethoden und -konzepte“ zuzuordnen sind und damit allen Aufgaben gemeinsames Wissen zu ihrer Beantwortung erfordern, andererseits aber jeweils einen spezifischen Themenschwerpunkt setzen, stellt sich die Frage nach der

Anzahl der Wissensdimensionen, die für die Beantwortung der Aufgaben erforderlich sind. Diese Frage wurde an anderer Stelle gesondert untersucht (Marx et al., eingereicht), so dass hier von einer erprobten Struktur ausgegangen werden kann. Ein Rasch Testlet Modell (Wang & Wilson, 2005) erklärte das Antwortverhalten der EB/WB-Lehrkräfte am besten, das die beschriebene Dimensionalitätsstruktur mit gemeinsamem Wissen über verschiedene Aufgaben hinweg sowie spezifischem Wissen für die jeweilig einzelne Aufgabe berücksichtigt (Abb. 2, Modell (a)). Allerdings stellt sich die Frage, ob v.a. die modellierten aufgabenspezifischen Wissensdimensionen nicht letztlich konstruktirrelevante Varianz abbilden, also beispielsweise innerhalb von Personen über die Zeit hinweg inkonsistente Merkmale, die z.B. auf die konkrete Befragungssituation zurückzuführen sind oder auf den Test als solchen und damit letztlich Effekte der Methode der Erfassung darstellen (z.B. Kelava & Schermelleh-Engel, 2012). Daher wurde für den vorliegenden Beitrag die in Marx et al. (eingereicht) vorgenommene Modellierung einem weiteren Validierungsschritt unterzogen, der hier nur zusammenfassend skizziert werden kann. Für diesen Validierungsschritt stellt sich die Frage, ob sich für die insgesamt sechs (eine gemeinsame (g) und fünf spezifischen (A-E)) Wissensdimensionen jeweils ein aus theoretisch-konzeptueller Sicht plausibles Zusammenhangsmuster mit den oben genannten ausbildungs- und tätigkeitsbezogenen Merkmalen zeigt.

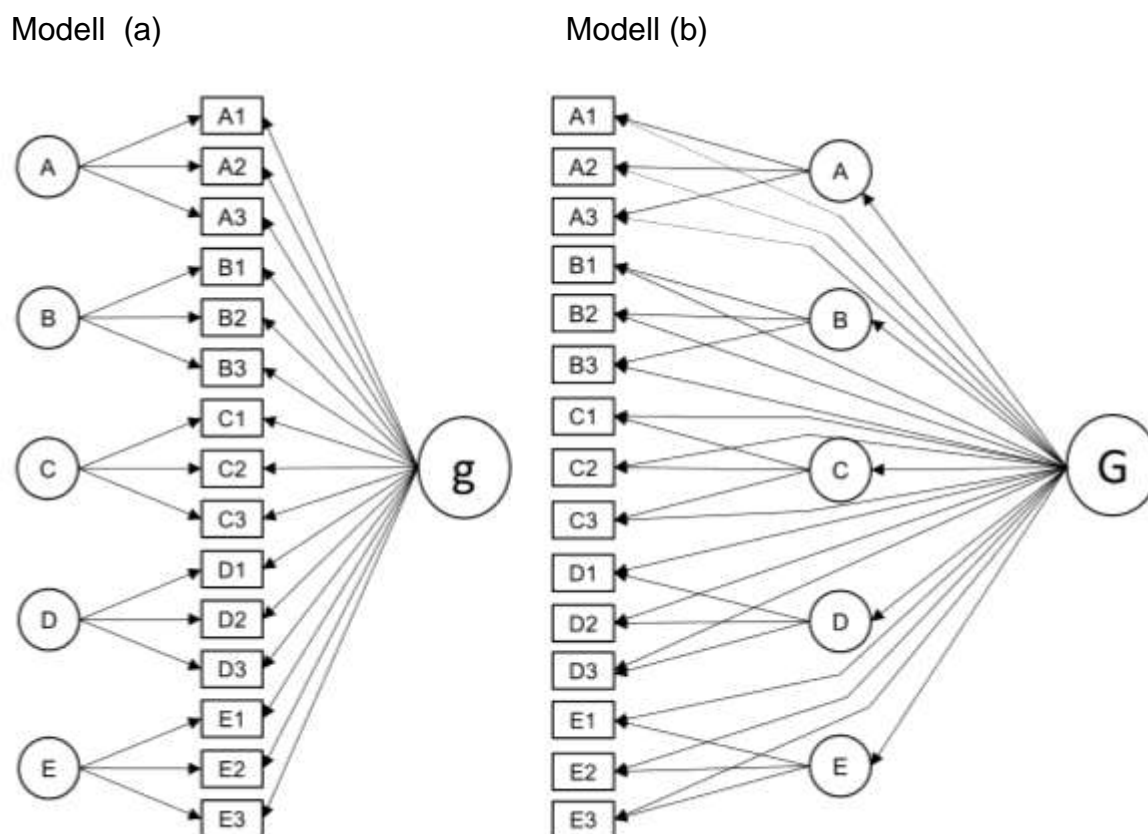


Abb. 2 Rasch Testlet Modell nach Wang und Wilson (2005) und full second-order Modell nach Chen, West and Sousa (2006)

Das in Marx et al. (eingereicht) vorgestellte Modell wurde in einem ersten Schritt in Mplus Version 7.4 spezifiziert.¹⁹ In einem zweiten Schritt wurde pro Wissensdimension durch eine latente multiple Regressionsanalyse auf die ausbildungs- und tätigkeitsbezogenen Merkmale regressiert (die vorher auf Konfundierungen überprüft wurden, siehe Kapitel 5). Es zeigte sich *cum grano salis* über alle sechs Wissensdimensionen hinweg ein weitgehend plausibles Zusammenhangsmuster (siehe Tab. 3 im Anhang). Hierdurch eröffnen sich zwei Möglichkeiten – *vorläufig bzw. solange keine weiteren Validierungsstudien dagegen sprechen*: Erstens können die Zusammenhänge zwischen ausbildungs- und tätigkeitsbezogenen Merkmalen einerseits und dem Wissen über Lehr-Lernmethoden und -konzepte andererseits als ebensolche Zusammenhänge aus einer theoretisch-konzeptuellen Perspektive interpretiert werden; zweitens können die insgesamt sechs Wissensdimensionen des spezifizierten Rasch Testlet Modells in einem gemeinsamen Faktor (G) in einem full second-order Modell zusammengefasst werden, der zwar hinsichtlich (etwaiger)

¹⁹ Gleiche Faktorladungen für alle Indikatoren auf die allen Aufgaben gemeinsame Wissensdimension und gleiche Faktorladungen der jeweils spezifischen Indikatoren auf die jeweils spezifischen Wissensdimensionen; die Varianz aller Wissensdimensionen wurde auf eins fixiert, vgl. zur Spezifikation von Rasch Modellen in Mplus Muthén und Muthén (1998-2015).

verschiedener Wissensbereiche ein weniger differenzierbares, aber überschaubareres Bild der Zusammenhänge anbieten kann (Abb. 2, Modell (b)). Dieser gemeinsame Faktor G bildet im Ergebnisbericht das Wissen über Lehr-Lernmethoden und -konzepte ab. Das full second-order Modell wurde spezifiziert wie von Chen, West und Sousa (2006) vorgeschlagen. Bei diesem Vorgehen zeigte sich für die Erfassung des Wissens über Lehr-Lernmethoden und -konzepte eine Reliabilität von $\omega^2 = .76$ (Marx et al., eingereicht).

Die Erfassung der ausbildungs- und tätigkeitsbezogenen Merkmale wurde mit Instrumenten aus den Erhebungen des wb-personalmonitors durchgeführt (Autorengruppe wb-personalmonitor, 2016). Die formulierten Hypothesen werden unter Berücksichtigung der Kontrollvariablen Alter und Geschlecht überprüft. Der Diskussion des Beitrags vorgehend soll an dieser Stelle bereits betont werden, dass aufgrund der für die zugrundeliegenden Analyseverfahren kleinen und ausgewählten Stichprobengröße der vorliegende Beitrag trotz der Überprüfung von Hypothesen einen inhaltlich explorierenden Ansatz verfolgt und nicht den Anspruch erhebt, repräsentativ-generalisierbare Aussagen zu treffen.

5 Ergebnisse

In einem ersten Schritt, und damit dem oben berichteten Validierungsschritt vorangehend, wurden die Prädiktorvariablen auf Konfundierungen überprüft. Hierbei zeigte sich ein erwartbarer signifikanter Zusammenhang zwischen den Variablen Alter und Lehrerfahrung in Jahren (siehe Tab. 1).

Tab. 1 Korrelationsmatrix der Variablen Alter, Lehrerfahrung in Jahren, Stunden Fort- und Weiterbildung in Lehren und Lernen / Methodik / Didaktik und aktueller Lehrumfang in Wochenstunden

	Alter	Lehrerfahrung in Jahren	Stunden Fort- und Weiterbildung
Lehrerfahrung in Jahren	$r = 0.560^*$ (0.119)		
Stunden Fort- und Weiterbildung	$r = 0.170$ n.s. (0.270)	$r = 0.365$ n.s. (0.353)	
Aktueller Lehrumfang in Wochenstunden	$r = 0.001$ n.s. (0.132)	$r = 0.197$ n.s. (0.114)	$r = 0.225$ n.s. (0.117)

Anmerkungen: n.s. = $p > 0.05$; * = $p \leq 0.05$;
 r = Pearson Korrelation
 Werte in Klammern: Standardfehler

Auch zeigte sich ein signifikanter Zusammenhang zwischen den Variablen erziehungs- und bildungswissenschaftliches Studium und Geschlecht, wobei mehr Frauen ($n = 35$) als Männer ($n = 19$) ein erziehungs- und bildungswissenschaftliches Studium absolviert haben bzw. weniger Frauen ($n = 64$) als Männer ($n = 74$) kein erziehungs- und bildungswissenschaftliches Studium absolviert haben; $\chi^2(1) = 4.571$, $p = 0.033$. Zudem zeigte sich ein signifikanter Zusammenhang zwischen den Variablen Geschlecht und Tätigkeitsschwerpunkt in der EB/WB, wobei mehr Frauen ($n = 65$) als Männer ($n = 37$) in der EB und mehr Männer ($n = 56$) als Frauen ($n = 29$) in der WB arbeiteten; $\chi^2(1) = 15.095$, $p = 0.000$. Für Lehrkräfte in der allgemeinen EB ($M = 51.56$; $SD = 11.46$) zeigte sich ein signifikant höheres Durchschnittsalter als für Lehrkräfte in der beruflich-betrieblichen WB ($M = 44.42$; $SD = 11.84$); ($t(188) = 4.216$, $p = 0.000$; $g_{Hedges} = -0.614$). Lehrkräfte ohne erziehungs- und bildungswissenschaftliches Studium ($M = 11.67$, $SD = 9.75$) verfügten über eine signifikant geringere Lehrerfahrung als Lehrkräfte mit erziehungs- und bildungswissenschaftlichem Studium ($M = 15.38$; $SD = 11.74$); $t(190) = -2.219$, $p = 0.028$; $g_{Hedges} = -0.360$). Für übrigen Prädiktorvariablen zeigten sich keine bedeutsamen Zusammenhänge.

Um die Frage zu beantworten, in welchem Zusammenhang die ausbildungs- und tätigkeitsbezogenen Merkmale von Lehrkräften aus der EB/WB mit ihrem Wissen über Lehr-Lernmethoden und -konzepte stehen, wurde zur Modellierung dieses Wissens ein full second-order Modell spezifiziert (Chen, West & Sousa, 2006). Insgesamt wurden 90 Modellparameter frei geschätzt (ESTIMATOR = MLF, vgl. Muthén & Muthén, 1998-2015). Alle dichotomen Variablen (Geschlecht, Schwerpunkt EB/WB, erziehungs- und bildungswissenschaftliches Studium) wurden als Dummy-Variablen in die Analysen aufgenommen. In Tab. 2 werden die unstandardisierten²⁰ und die standardisierten Regressionsgewichte berichtet, Standardfehler, t -Wert und p -Wert beziehen sich auf die standardisierten Regressionsgewichte ((STANDARDIZED (STDY), vgl. Muthén & Muthén, 1998-2015).

Alter stellte eine Kontrollvariable dar, für die sich ein signifikanter negativer Zusammenhang mit dem Wissen über Lehr-Lernmethoden und -konzepte zeigte. Es zeigten sich – entgegen der formulierten Hypothese 3a – keine signifikanten Zusammenhänge zwischen dem Wissen über Lehr-Lernmethoden und -konzepte und dem aktuellen Lehrumfang in Wochenstunden. Das Ausmaß an diesem Wissen scheint zudem wider Erwarten in keinem statistisch bedeutsamen Zusammenhang damit zu stehen, ob die bisherige Lehrerfahrung kurz oder lang ist (H3b) und ob ein erziehungs- und bildungswissenschaftliches Studium absolviert wurde oder nicht (H1). Konform zu den

²⁰ Die Variablen absolvierte Stunden an Fort- und Weiterbildung und Lehrerfahrung in Jahren wurden aufgrund der Schiefe der Verteilungen z-standardisiert.

Hypothesen 2 und 4 zeigten sich zum Wissen über Lehr-Lernmethoden und -konzepte ein signifikanter positiver Zusammenhang zwischen der Anzahl der Stunden in Fort- und Weiterbildung (H2) sowie ein signifikanter negativer Interaktionseffekt zwischen Lehrerfahrung in Jahren und der absolvierten Stunden an Fort- und Weiterbildung (H4). Um den Interaktionseffekt besser interpretieren zu können, sind in Abb. 3 für die Gruppe der Lehrkräfte mit einer Standardabweichung unter dem Durchschnitt ($B = 0.402$), Durchschnitt ($B = 0.214$) und einer Standardabweichung über dem Durchschnitt ($B = 0.026$) an Lehrerfahrung in Jahren die Regressionsgeraden des Wissens über Lehr-Lernmethoden und -konzepte für den absolvierten Umfang an Fort- und Weiterbildung abgebildet. Im vorliegenden Beitrag wurde der Interaktionseffekt mit der Variablen Lehrerfahrung in Jahren als Moderatorvariable abgebildet. Für Lehrkräfte mit einer Standardabweichung über dem Durchschnitt an Lehrerfahrung zeigte sich ein weniger enger Zusammenhang zwischen dem absolvierten Umfang an Fort- und Weiterbildung zum Wissen über Lehr-Lernmethoden und -konzepte als für die Gruppe die durchschnittlich bzw. wenig Lehrerfahrung hat. Dies bedeutet, dass für Lehrkräfte mit wenig Lehrerfahrung der Umfang an absolvierter Fort- und Weiterbildung einen größeren Effekt zu haben scheint als für Lehrkräfte mit durchschnittlicher Lehrerfahrung oder einer Standardabweichung unter dem Durchschnitt – Lehrkräfte mit wenig Lehrerfahrung scheinen also stärker von Fort- und Weiterbildung zu profitieren als Lehrkräfte mit durchschnittlicher oder viel Lehrerfahrung. Bei Abbildung der Variablen Umfang an Fort- und Weiterbildung als Moderatorvariable zeigte sich u.a. für die Lehrkräfte mit weniger Fort- und Weiterbildung ein stärkerer positiver Zusammenhang zwischen Lehrerfahrung und dem Wissen über Lehr-Lernmethoden und -konzepte als für Lehrkräfte mit durchschnittlichem Fort- und Weiterbildung (siehe Abb. 4 im Anhang).

Konform zur Hypothese 5 zeigt sich ein signifikanter Unterschied im Wissen über Lehr-Lernmethoden und -konzepte zugunsten der Lehrkräfte mit Schwerpunkt in der beruflich-betrieblichen WB gegenüber Lehrkräften in der allgemeinen EB. Für das gesamte Modell zeigte sich eine Varianzaufklärung im Wissen über Lehr-Lernmethoden und -konzepte von 62.6%.

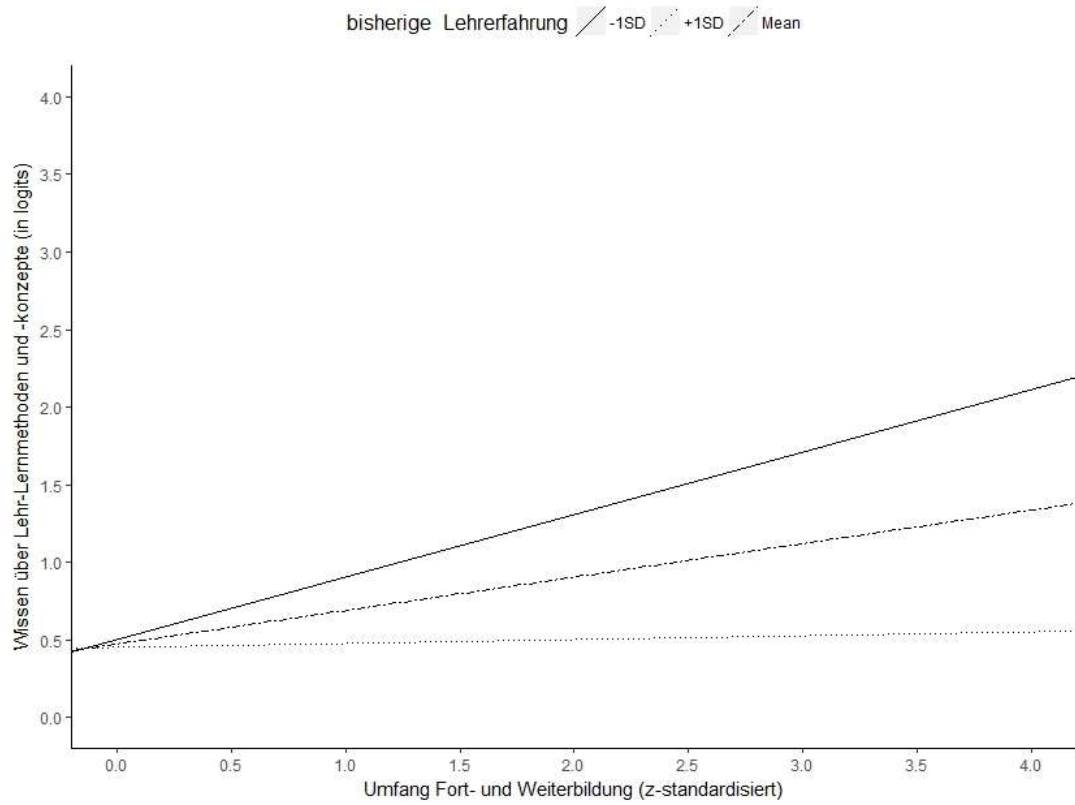


Abb. 3 Interaktionseffekt zwischen Lehrerfahrung in Jahren und den absolvierten Stunden in Fort- und Weiterbildung.

Tab. 2 Ergebnisse der multiplen linearen Regressionsanalyse zur Vorhersage des „Wissens über Lehr-Lernmethoden und -konzepte“ von Lehrkräften aus der Erwachsenen- und Weiterbildung aus ausbildungs- und tätigkeitsbezogenen Merkmalen ($N = 209$)

	<i>B</i>	<i>SE</i>	β	<i>SE</i>	<i>t</i>	<i>p</i>
Alter	-0.017	0.011	-0.413	0.161	-2.568	0.010
Geschlecht	-0.130	0.170	-0.132	0.154	-0.859	0.391
Tätigkeitsschwerpunkt EB oder WB	0.414	0.227	0.420	0.147	2.851	0.004
Aktueller Lehrumfang in Wochenstunden	-0.004	0.007	-0.087	0.153	-0.567	0.571
Erziehungs- und bildungswissenschaftliches Studium	-0.051	0.157	-0.045	0.138	-0.328	0.743
Lehrerfahrung in Jahren	-0.027	0.082	-0.055	0.170	-0.321	0.748
Stunden Fort- und Weiterbildung	0.214	0.144	0.438	0.208	2.110	0.035
Interaktion Lehrerfahrung und Fort- und Weiterbildung	-0.188	0.123	-0.452	0.221	-2.040	0.041

Gesamtmodell: $R^2 = 62.6\%$

Anmerkungen:

Geschlecht: Frauen = 0, Männer = 1;

EB/WB: Tätigkeitsschwerpunkt in der allgemeinen Erwachsenenbildung = 0, Tätigkeitsschwerpunkt in der beruflich-betrieblichen Weiterbildung = 1;

Erziehungs- und bildungswissenschaftliches Studium: nein = 0, ja = 1

6 Zusammenfassung und Diskussion der Ergebnisse

Dieser Beitrag widmete sich der für den Bereich der EB/WB bislang unbeantworteten Frage, welche Zusammenhänge sich zwischen ausbildungs- und tätigkeitsbezogenen Merkmalen und dem Wissen über Lehr-Lernmethoden und -konzepte von EB/WB-Lehrkräften zeigen. Die beobachteten Zusammenhänge werden nachfolgend vor dem Hintergrund der für diesen Beitrag zur Verfügung stehenden Stichprobe in ihrer Bedeutsamkeit eingeordnet.

Es zeigten sich keine Unterschiede zwischen EB/WB-Lehrkräften mit und ohne erziehungs- und bildungswissenschaftlichem Studium im Wissen über Lehr-Lernmethoden und -konzepte, aber Unterschiede in Abhängigkeit vom Umfang an absolvierten Stunden an Fort- und Weiterbildungen zum Thema Lehren und Lernen / Methodik / Didaktik. Diese Resultate können als Konsequenz der empirischen Befundlage gedeutet werden, dass Trainerausbildungen und erziehungswissenschaftliche Studiengänge recht unterschiedliche Inhalte vorsehen (vgl. Wißhak, 2017): Während einschlägige Fortbildungen einen direkten Mehrwert für das Wissen über Lehr-Lernmethoden und -konzepte intendieren und auch zu haben scheinen (wenn man diese kausale Wirkrichtung unterstellen mag), erbringen erziehungs- und bildungswissenschaftliche Studiengänge keinen hier belegbaren Mehrwert für dieses Wissen. Letzteres Ergebnis sollte allerdings vorsichtig interpretiert werden. Denn die vergleichsweise undifferenzierte Abbildung der Lehrkräfte mit unterschiedlichen erziehungs- und bildungswissenschaftlichen Studienabschlüssen in einer gemeinsamen Gruppe kann dazu führen, dass sich etwaige doch vorhandene Effekte ggf. verlieren. Die aus theoretisch-konzeptueller Perspektive plausibel erscheinende Annahme, dass sich (wenigstens bestimmte) erziehungs- und bildungswissenschaftliche Studiengänge positiv – ggf. aber weniger stark als einschlägige Fort- und Weiterbildung – auf das Wissen über Lehr-Lernmethoden und -konzepte auswirken könnten, wird daher trotz der Hinweise, die der vorliegende Beitrag gibt, nicht vollständig aufgegeben, sondern sie bedarf weiterer empirischer Klärung.

Dass mehr Stunden an spezifischer Fort- und Weiterbildung zum Thema Lehren und Lernen / Methodik / Didaktik mit mehr Wissen zu Lehr-Lernmethoden und -konzepten einhergehen und dass insbesondere weniger Lehrerfahrene, wenn sie umfangreichere Fort- und Weiterbildungen besucht haben, über mehr methodisches Wissen verfügen, darf als ermutigender Befund gedeutet werden: Die eingangs angesprochenen vielfältigen Initiativen zur Fort- und Weiterbildung von Lehrkräften setzen an der richtigen Stelle an.

Für die Lehrkräfte mit vergleichsweise wenig Fort- und Weiterbildung zeigte sich ein stärkerer positiver Zusammenhang zwischen Lehrerfahrung und dem Wissen über Lehr-Lernmethoden und -konzepte als für Lehrkräfte mit durchschnittlichem Fort- und Weiterbildung (siehe Abb. 4 im Anhang). Vor dem Hintergrund der Annahme, dass mit einem höheren Wissen über Lehr-Lernmethoden und -konzepte eine größere Chance auf eine

höhere Unterrichtsqualität einhergeht, kann eine praktische Implikation dieses Befunds sein, dass es sich trotz der „kompensierenden“ Wirkung von Lehrerfahrung lohnt, zu einem frühen Zeitpunkt, ggf. noch vor Beginn der Lehrtätigkeit, in die Ausbildung der EB/WB-Lehrkräfte zu investieren. Denn interessant ist, dass sich kein direkter Zusammenhang zwischen der Lehrerfahrung, erfasst durch den aktuellen Lehrumfang pro Woche in den vergangenen 12 Monaten, bzw. der bisherigen Lehrerfahrung insgesamt, erfasst in Jahren, mit dem Wissen über Lehr-Lernmethoden und -konzepte zeigt. Lehrkräfte werden also weder durch das kurz- noch durch das längerfristige Anhäufen von Lehrerfahrung (wobei offenbleiben muss, in welchem Gesamtvolumen) kenntnisreicher bezogen auf Lehr-Lernmethoden und -konzepte (ähnlich Rivkin, Hanushek & Kain, 2005). Dieser Befund relativiert die praktische Bedeutung von „viel Erfahrung“ einer Lehrkraft z.B. für die Trainerauswahl – immer vorausgesetzt, dass methodische Kenntnis und praktisches Lehrhandeln nicht völlig unabhängig voneinander sind.

Ein erwartungskonform signifikanter Unterschied im Wissen über Lehr-Lernmethoden und -konzepte zeigte sich zu Gunsten der Lehrkräfte mit Arbeitsschwerpunkt in der beruflich-betrieblichen Weiterbildung gegenüber Lehrkräften in der allgemeinen Erwachsenenbildung. Ob die hypothesenleitende Annahme zutrifft, dass eine ausgeprägtere „Anspruchshaltung“ der Teilnehmenden an Methodenvielfalt und didaktischer Professionalität in ersterem Weiterbildungsbereich größer ist (sowie üblicherweise eine höhere existentielle Abhängigkeit der TrainerInnen von der Zufriedenheit der „Kunden“) als im letzteren und dies „differenzielle Entwicklungsmilieus“ für das Lehrkräfte-Wissen begründet, kann hier in einer korrelativen Studie nicht als belegt gewertet werden, es liegt jedoch nahe. Der berichtete Befund kann als Anregung dazu dienen, im Rahmen zukünftiger Untersuchungen das Merkmal des Tätigkeitsschwerpunkts – und wenn möglich auch des Themenschwerpunkts – differenzierter zu betrachten.

Überraschend zeigte sich für die Kontrollvariable Alter ein negativer Zusammenhang zum Wissen über Lehr-Lernmethoden und -konzepte. Hier könnte es jedoch aufgrund der Stichprobengröße zu einer Konfundierung von Alter mit Lehrerfahrung gekommen sein (vgl. Tab. 1).

Die ausbildungs- und tätigkeitsbezogenen Merkmale im aufgestellten Regressionsmodell erklären insgesamt ungefähr zwei Drittel der Unterschiede im Wissen über Lehr-Lernmethoden und -konzepte der EB/WB-Lehrkräfte, was prinzipiell ein bedeutsamer Hinweis darauf ist, dass diese Merkmale der Lehrkräfte zur Entwicklung des Wissens über Lehr-Lernmethoden und -konzepte beitragen *könnten* – inwiefern es sich hierbei um kausale Effekt handelt, kann mit den vorliegenden Querschnittsdaten nicht beantwortet werden. Denn für den Bereich der EB/WB gibt es derzeit keine Studien, die im Rahmen der derzeit eben nur angenommenen, aber nicht belegten Wirkungskette zwischen dem Wissen der

Lehrkräfte und z.B. dem Lernerfolg der Teilnehmenden eine Verbindung herstellen können. Neben der begrenzten Stichprobengröße, die sich auch auf die Teststärke auswirkt, sind die Ergebnisse zudem unter dem Vorbehalt zu sehen, dass die Erfassung des Wissens über Lehr-Lernmethoden und -konzepte über die bereits vorgenommenen Validierungsschritte hinaus selbst noch weiter validiert werden muss.

Dennoch können die berichteten Befunde und methodischen Herausforderungen wichtige Anregungen geben für zukünftige und bestenfalls längsschnittliche Untersuchungen zur Frage nach der Entwicklung (von Aspekten) der professionellen Kompetenz von Lehrkräften.

7 Literaturverzeichnis

- Arnold, R., Krämer-Stürzl, A. & Siebert, H. (2005). *Dozentenleitfaden. Planung und Unterrichtsvorbereitung in Fortbildung und Erwachsenenbildung*. Berlin: Cornelsen.
- Autorengruppe Bildungsberichterstattung. (2016). *Bildung in Deutschland 2016 Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration* (1. Aufl.). Bielefeld: Bertelsmann.
- Autorengruppe wb-personalmonitor (Hrsg.). (2016). *Personal in der Weiterbildung: Beschäftigungsverhältnisse und Tätigkeiten – wb-personalmonitor*. Bielefeld: Bertelsmann.
- Avramovska, M., Czerwinski, T. & Lattke, S. (2015). *Curriculum globALE. Global curriculum for Adult Learning and Education*. (Second Edition). Bonn: DVV International, DIE.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29-54). Münster: Waxmann.
- Chen, F. F., West, S. G. & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41 (2), 189-225.
- Collins, J. B. & Pratt, D. D. (2011). The teaching perspectives inventory at 10 years and 100,000 respondents. *Adult Education Quarterly*, 61 (4), 358-375.
- Digel, S. (2013). Von der Diagnose zum Handeln: Befunde aus einer Interventionsstudie zum Lernen mit Videofällen. In S. Digel & J. Schrader (Hrsg.), *Diagnostizieren und Handeln von Lehrkräften. Lernen aus Videofällen in Hochschule und Erwachsenenbildung* (S. 25-61). Bielefeld: Bertelsmann.
- Dobischat, R., Elias, A. & Rosendahl, A. (Hrsg.). (2018). *Das Personal in der Weiterbildung. Im Spannungsfeld von Professionsanspruch und Beschäftigungsrealität*. Wiesbaden: Springer.
- European Commission. (2015). *Strengthening teaching in Europe. New evidence from teachers compiled by Eurydice and CRELL*. Verfügbar unter http://ec.europa.eu/education/library/policy/teaching-profession-practices_en.pdf
- Goeze, A. (2016). *Professionalitätsentwicklung von Lehrkräften durch videofallbasiertes Lernen – Voraussetzungen, Prozesse, Wirkungen* (Theorie und Praxis der Erwachsenenbildung). Bielefeld: Bertelsmann.

- Goeze, A., Hetfleisch, P. & Schrader, J. (2013). Wirkungen des Lernens mit Videofällen bei Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 16 (1), 79-113.
- Gruber, H., Mandl, H. & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln: Empirische und theoretische Lösungsansätze* (S. 139-156). Göttingen: Hogrefe.
- Guerriero, S. (Hrsg.). (2017). *Pedagogical knowledge and the changing nature of the teaching profession*. Paris: OECD Publishing.
- Hartz, S. (2011). *Qualität in Organisationen der Weiterbildung. Eine Studie zur Akzeptanz und Wirkung von LQW* (1. Aufl). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Herrle, M., Dinkelaker, J., Nolda, S. & Kade, J. (2014). Kursforschung und Videographie. In J. Kade, S. Nolda, J. Dinkelaker & M. Herrle (Hrsg.), *Videographische Kursforschung. Empirie des Lehrens und Lernens Erwachsener* (S. 13-40). Stuttgart: Kohlhammer.
- Hof, C. (2010). Wissen als Thema der Erwachsenenbildung. In B. Dewe (Hrsg.), *Professionswissen und erwachsenenpädagogisches Handeln. Dokumentation der Jahrestagung 2001 der Sektion Erwachsenenbildung der Deutschen Gesellschaft für Erziehungswissenschaft* (S. 9-17). Bielefeld: Bertelsmann.
- Hoffmann, N. (2011). Ratgeber-Didaktik in Didaktik-Ratgebern. Portrait eines Selbstlernmediums aus erwachsenenpädagogischer Perspektive. *Der pädagogische Blick*, 19 (1), 4-12.
- Hohmann, R. (2010). Fortbildungsbedarf von Lehrkräften aus Sicht der KBE. In J. Schrader, R. Hohmann & S. Hartz (Hrsg.), *Mediengestützte Fallarbeit. Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern* (S. 9-24). Bielefeld: Bertelsmann.
- Kelava, A. & Schermelleh-Engel, K. (2012). Latent-State-Trait-Theorie (LST). 364-381. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl.). Heidelberg: Springer.
- Kraft, S., Seitter, W. & Kollwe, L. M. (2009). *Professionalitätsentwicklung des Weiterbildungspersonals*. Bielefeld: Bertelsmann.
- Lattke, S. & Jütte, W. (Hrsg.). (2014). *Professionalisation of adult educators: International and comparative perspectives*. Frankfurt am Main: Peter Lang.
- Lencer, S. & Strauch, A. (2016). *Das GRETA-Kompetenzmodell für Lehrende in der Erwachsenen- und Weiterbildung*, Deutsches Institut für Erwachsenenbildung. Verfügbar unter <http://www.die-bonn.de/doks/2016-erwachsenenbildung-02.pdf>

- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58 (2), 103-117.
- Marx, C., Goeze, A. & Schrader, J. (2014a). Adult education teachers' pedagogical psychological knowledge: Elements and test development. In S. Lattke & W. Jütte (Hrsg.), *Professionalisation of adult educators: International and comparative perspectives* (S. 162-182). Frankfurt am Main: Peter Lang.
- Marx, C., Goeze, A. & Schrader, J. (2014b). Pädagogisch-psychologisches Wissen zur Gestaltung von Lehr-Lernsituationen: (Wie) unterscheidet es sich in Erwachsenenbildung/Weiterbildung und Schule? *Hessische Blätter für Volksbildung*, 64 (3), 238-251.
- Marx, C., Goeze, A., Voss, T., Deutscher, V. K., Kelava, A. & Schrader, J. (eingereicht). Das Wissen über Lehr-Lernmethoden von Lehrkräften der Erwachsenen- und Weiterbildung – Zur validen Messung eines mehrdimensionalen Konstrukts. *Journal for Educational Research Online*.
- Marx, C., Goeze, A., Voss, T., Hoehne, V., Klotz, V. K. & Schrader, J. (2017). Pädagogisch-psychologisches Wissen von Lehrkräften aus Schule und Erwachsenenbildung. Entwicklung und Erprobung eines Testinstruments. *Zeitschrift für Erziehungswissenschaft*, 20 (S1), 165-200.
- Muthén, L. K. & Muthén, B. O. (1998-2015). *Mplus user's guide*. (Seventh Edition). Los Angeles, CA: Muthén & Muthén.
- Neumann, M., Schnyder, I., Trautwein, U., Niggli, A., Lüdtke, O. & Cathomas, R. (2007). Schulformen als differenzielle Lernmilieus. *Zeitschrift für Erziehungswissenschaft*, 10 (3), 399-420.
- Pachner, A. (2013). Selbstreflexionskompetenz. Voraussetzung für Lernen und Veränderung in der Erwachsenenbildung? *Magazin erwachsenenbildung.at* (20), 1-7.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73 (2), 417-458.
- Rohs, M., Rott, K. J., Schmidt-Hertha, B. & Bolten, R. (2017). Medienpädagogische Kompetenzen von ErwachsenenbildnerInnen. *Magazin erwachsenenbildung.at*, 30, 1-14.
- Rosenblatt, B. von & Lehmann, R. (2013). Grade der Schriftbeherrschung und subjektiver Lernerfolg bei Teilnehmenden an Alphabetisierungskursen. *Zeitschrift für Erziehungswissenschaft*, 16 (1), 55-77.
- Schläfli, A. & Sgier, I. (2008). *Porträt Weiterbildung Schweiz* (Länderporträts Weiterbildung). Bielefeld: Bertelsmann.

- Schöb, S., Rohs, M., Biel, C. & Scholze, T. (2016). Professionalisierung von Lehrhandeln in einer digital geprägten Lernkultur – Bedarf und Ansatzpunkte der Entwicklung einer Online-Lernumgebung. In O. Dörner, C. Iller, H. Pätzold & S. Robak (Hrsg.), *Differente Lernkulturen – regional, national, transnational* (S. 171-183). Opladen: Barbara Budrich.
- Schrader, J., Anders, Y. & Richter, D. (in Druck). Fortbildung des pädagogischen Personals in der frühen Bildung, der Schule und der Erwachsenen- und Weiterbildung. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, H. Solga et al. (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale*. Bad Heilbrunn: UTB.
- Schütz, J. & Nittel, D. (2012). Von der Heterogenität zur Vielfalt. Akademische Professionalisierung im Blick einer komparativen pädagogischen Berufsgruppenforschung. In R. Egetenmeyer & I. Schüßler (Hrsg.), *Akademische Professionalisierung in der Erwachsenenbildung / Weiterbildung* (Grundlagen der Berufs- und Erwachsenenbildung, Bd. 70, S. 229–244). Baltmannsweiler: Schneider Hohengehren.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4-14.
- Skott, J. (2015). The promises, problems, and prospects of research on teachers' beliefs. In H. Fives & M. G. Gill (Hrsg.), *International handbook of research on teachers' beliefs* (Educational psychology handbook series, S. 13-30). New York [u.a.]: Routledge.
- Sorge, S., Kröger, J., Petersen, S. & Neumann, K. (2017). Structure and development of pre-service physics teachers' professional knowledge. *International Journal of Science Education*, Advanced online publication.
- Stanik, T. (2016). Mikrodidaktische Planungen von Lehrenden in der Erwachsenenbildung. *Zeitschrift für Weiterbildungsforschung – Report*, 39 (3), 317-330.
- Steiner, P. (2010). Die Weiterbildungsakademie Österreich (wba). Erfahrungen und Zugänge zu Beratung und Kompetenzanerkennung für ErwachsenenbildnerInnen. *Magazin erwachsenenbildung.at* (14), 1-6.
- Terhart, E. (2012). Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung*, 2 (1), 3-21.
- Tietgens, H. (1967). *Schule und Erwachsenenbildung*. Braunschweig: Georg Westermann Verlag.
- Vinepac-Project. (2008). *Handbook for the use of Validpack for the validation of psychopedagogical adult educators' competences*. Timisoara: Editura Mirton.

- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103 (4), 952-969.
- 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.
- Voss, T., Goeze, A., Marx, C., Hoehne, V., Klotz, V. K. & Schrader, J. (2017). Using digital media to assess and promote school and adult education teacher competence. In J. Buder & F. W. Hesse (Hrsg.), *Informational Environments* (S. 125-148). Cham: Springer International Publishing AG.
- Wainer, H. & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24 (3), 185-201.
- Wang, W.-C. & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement*, 29 (2), 126-149.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45-65). Seattle: Hogrefe & Huber.
- Wißhak, S. (2017). *Qualifizierung und professionelles Wissen von Lehrpersonen in der berufsbezogenen Weiterbildung*. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:bsz:352-2--qjhzazjd0xb83>
- Wißhak, S. & Hochholding, S. (2016). Analyse der Inhalte erziehungswissenschaftlicher Studiengänge im Hinblick auf eine spätere Tätigkeit in der berufsbezogenen Weiterbildung. *Zeitschrift für Weiterbildungsforschung – Report*, 39 (1), 97-115.
- Ziep, K.-D. (1990). *Der Dozent in der Weiterbildung. Professionalisierung und Handlungskompetenzen*. Weinheim: Deutscher Studien Verlag.

8 Anhang

Tab. 3 Ergebnisse multipler latenter Regressionsanalysen (standardisierte Regressionsgewichte (STDY))

		g	A	B	C	D	E
Alter	β	-0.012	-0.024	-0.024	-0.001	-0.043	0.004
	SE	(0.053)	(0.163)	(0.173)	(0.275)	(0.219)	(0.294)
Geschlecht	β	-0.135	0.131	-0.163	0.064	0.068	-0.098
	SE	(0.130)	(0.221)	(0.298)	(0.149)	(0.290)	(0.311)
Tätigkeitsschwerpunkt EB oder WB	β	0.239	0.117	0.134	0.121	0.034	0.002
	SE	(0.133)	(0.212)	(0.318)	(0.317)	(0.224)	(0.170)
Aktueller Lehrumfang in Wochenstunden	β	0.015	0.017	0.163	-0.234	0.033	0.026
	SE	(0.155)	(0.397)	(0.354)	(0.150)	(0.302)	(0.251)
Erziehungs- und bildungswiss. Studium	β	-0.086	-0.114	0.080	0.128	0.092	-0.016
	SE	(0.153)	(0.232)	(0.406)	(0.153)	(0.409)	(0.172)
Lehrerfahrung in Jahren	β	-0.075	0.223	-0.182	0.087	-0.226	0.022
	SE	(0.204)	(0.269)	(0.400)	(0.184)	(0.531)	(0.502)
Stunden Fort- und Weiterbildung	β	0.232	-0.279	0.181	0.091	0.186	0.042
	SE	(0.329)	(0.203)	(0.513)	(0.547)	(0.428)	(0.906)
Interaktion Lehrerfahrung und Fort- Weiterbildung	β	-0.265	0.048	-0.036	-0.088	-0.102	-0.029
	SE	(0.318)	(0.260)	(0.581)	(0.532)	(0.657)	(0.557)

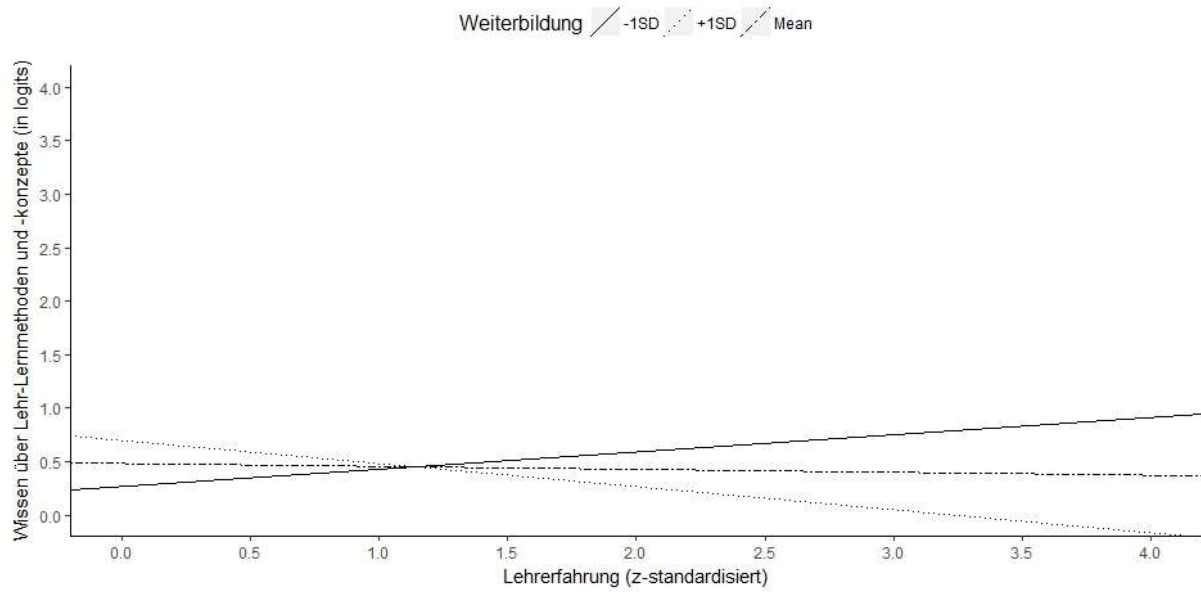


Abb. 4 Interaktionseffekt zwischen Lehrerfahrung in Jahren und den absolvierten Stunden in Fort- und Weiterbildung.

9 Gesamtdiskussion – Fragestellungen, Befunde und Perspektiven

Die vorliegende Dissertationsschrift berichtete über die Konzeptualisierung, Entwicklung und Validierung eines fach- und bildungsbereichsübergreifend einsetzbaren Tests zur Erfassung des PPK. Ein solcher Test kann in der pädagogischen Praxis diverser Bildungsbereiche, u.a. dem der EB/WB und in der auf diesen Bereich bezogenen empirischen Lehr-Lernforschung Anwendung finden. Hilfreich kann der für diese fach- und bildungsbereichsübergreifenden Zwecke neu entwickelte Test insbesondere für die Beantwortung der Frage nach der Allgemeinheit des PPK sein. Diese theoretisch ebenso wie praktisch relevante Frage der vollständigen bzw. teilweisen Allgemeinheit (einzelner Facetten) des PPK ist aus unterschiedlichen Perspektiven bisher entweder implizit unterstellt oder explizit angenommen worden (vgl. Studie 1). Bisher lag jedoch noch kein psychometrisch geprüft Instrument vor, das in der Lage gewesen wäre, diese Frage empirisch zu untersuchen.

Die berichteten Schritte und Ergebnisse sind Teil eines längerfristig angelegten Vorhabens zur Erforschung der Allgemeinheit des PPK und wurden im Projekt Think umgesetzt. Die Frage nach der fach- und bildungsbereichsübergreifenden Allgemeinheit des PPK wurde immer wieder (implizit) thematisiert (z.B. Aebli, 2011) und gestellt (z.B. Terhart, 2013) und ist für die Allgemeine Didaktik und die empirische Lehr-Lernforschung eine aus theoretisch-konzeptueller sowie für die Ausbildung von Lehrkräften auch aus praktisch-curriculer Sicht eine relevante Frage. Bislang wurde aber nicht begonnen, diese bildungsbereichsübergreifend empirisch zu überprüfen (siehe für eine fachbereichsübergreifende Bearbeitung der Frage jedoch jüngst Lohse-Bossenz, Holzberger, Kunina-Habenicht, Seidel & Kunter, 2018). Kurz gefasst meint die Frage nach der Allgemeinheit des PPK, ob bzw. inwiefern Prinzipien des Lehrens und Lernens über verschiedene Fach- und Bildungsbereiche hinweg „gültig“ im Sinne von „wirksam“ sind. Das Wissen über diese Prinzipien, das in verschiedene Facetten unterteilt werden kann, konstituiert schließlich das PPK. Um die Frage nach der Allgemeinheit des PPK zu beantworten, gibt es prinzipiell verschiedene Möglichkeiten, von denen nachfolgend zwei skizziert werden.

Angenommen, es liegt für alle Bildungsbereiche ein weitestgehend vergleichbarer, ausreichend hoher Stand an empirischer-Lehr-Lernforschung vor, dann könnten Befunde, z.B. zur Wirksamkeit eines lehr-lernmethodischen Vorgehens über verschiedene Bildungsbereiche verglichen werden und durch diesen Vergleich eine Aussage über die Spezifik (Wirksamkeit zeigt sich nur in Bildungsbereich x und oder y) oder die Allgemeinheit („wirkt“ in allen Bildungsbereichen) dieser Wirkungsbeziehung gemacht werden. Ist man an einem Vergleich zwischen Schule und EB/WB interessiert, so ist diese Voraussetzung jedoch nicht erfüllt, da sich für die EB/WB ein deutlich eingeschränkterer entsprechender Forschungsstand vorfindet (z.B. Kuper & Schrader, in Druck).

Eine zweite Möglichkeit ist es, Hinweise auf die Frage nach der fach- und bildungsbereichsübergreifenden Allgemeinheit „indirekt“ über den Zusammenhang des PPK von Lehrkräften und dem Lernerfolg der Lernenden zu gewinnen. Aufgaben zur Erfassung des PPK werden auf der Grundlage empirischer Befunde aus dem Schulkontext entwickelt und der Zusammenhang zwischen dem PPK der Lehrkräfte aus der EB/WB und dem Lernerfolg der Teilnehmer wird untersucht. Zeigt sich hierbei unter Kontrolle relevanter Drittvariablen ein positiver Zusammenhang, dann wäre dies ein Hinweis auf die Allgemeinheit des PPK, da dieses Wissen sowohl im Schulkontext, aus dem und für den die Befunde gewonnen wurden, als auch in der EB/WB „Gültigkeit“ zu haben scheinen, da das Wissen auch in diesem anderen Bildungsbereich offenbar genauso „wirksam“ werden kann. Um den beschriebenen Wirkungszusammenhang zwischen PPK der Lehrkraft und dem Lernerfolg der Teilnehmenden überprüfen zu können, ist zunächst ein Test zur Erfassung (von Facetten) des PPK nötig, dessen Konzeptualisierung, Entwicklung und Validierung im Fokus der vorliegenden Arbeit stand.²¹

9.1 Zusammenfassung zentraler Ergebnisse

In Studie 1 der vorliegenden Dissertationsschrift lag der Fokus auf der Frage, ob und inwiefern das PPK fach- und bildungsbereichsübergreifend konzeptualisiert werden kann und ob und inwiefern zwei Facetten des PPK bei Lehrkräften aus der EB/WB und dem Schulkontext reliabel und valide erfasst werden können. Studie 2 legte ihren Schwerpunkt auf die Konzeptualisierung des Wissens über Lehr-Lernmethoden und -konzepte sowie deren empirische Überprüfung. Studie 3 berichtete im Schwerpunkt über die Validierung eines Tests zur Erfassung des Wissens über Lehr-Lernmethoden und -konzepte bei Lehrkräften in der EB/WB.

Da, wie in Kapitel 1.2 angesprochen, die Allgemeinheit des PPK eine Grundvoraussetzung ist, um einen „gültigen“ Test für die Erfassung des PPK bei Lehrkräften in der EB/WB anbieten zu können, werden zunächst die theoretisch-konzeptuellen und empirischen Befunde zusammengefasst, die Hinweise auf die Allgemeinheit des PPK geben. In einem zweiten Schritt werden die Ergebnisse der Erfassung bzw. Validierung des Tests zur Erfassung dieses Wissens vorgestellt. Vor dem Hintergrund der theoretisch-konzeptuell und empirisch begründeten „gültigen“ Erfassung von Facetten des PPK werden schließlich in einem dritten Schritt für die Erwachsenenbildungsforschung relevante Befunde vorgestellt.

²¹ Mit diesem Vorgehen ist allerdings ein grundlegendes Problem verbunden: Zeigt sich kein Zusammenhang (vermittelt z.B. über die Unterrichtsqualität) zwischen Facetten des PPK und dem Lernerfolg der TN, dann könnte nicht eindeutig gesagt werden, ob dieser Nulleffekt dadurch begründet ist, dass es in der Realität einfach keinen Effekt gibt, z.B. zwischen dem Wissen über Lehr-Lernmethoden und -konzepten und dem Lernerfolg der Teilnehmenden, oder ob das als allgemein angenommene Wissen bzw. vielmehr die Befunde, die aus der Schule auf die EB übertragen wurden, einfach nicht übertragbar sind, d.h. die Wirkungsbeziehungen in Schule eben nicht diesselben wie in der EB/WB sind – kurz: es ist nicht klar, ob die „falschen“ Themen (Facetten) identifiziert wurden oder diese lediglich durch „falsche“ Wirkungsbeziehungen (empirische Befunde) operationalisiert wurden.

Zur Beantwortung der Frage nach der Allgemeinheit des PPK konnte die vorliegende Arbeit durch die fach- und bildungsbereichsübergreifende Konzeptualisierung des PPK einen Beitrag leisten. Die Konzeptualisierung kann eine qualitativ-empirisch fundierte Legitimationsgrundlage darstellen für das, was Lehrkräfte alles wissen und können (sollten) im Rahmen einer tatsächlich allgemeinen (im Sinne von bildungsbereichsübergreifenden) Didaktik, die bislang aus einer theoretisch-konzeptuellen (und empirischen) Perspektive zumeist aus schulbezogener Sicht gedacht wurde (siehe aber Aebli, 2011; Terhart, 2013), weshalb „allgemein“ häufig lediglich als fachübergreifend verstanden wurde bzw. wird. Ein weiterer Beitrag kann die Überprüfung der Plausibilität der Befunde aus der schulbezogenen empirischen Lehr-Lernforschung auf ihre Übertragbarkeit auf die EB/WB darstellen. Diese Überprüfung wurde im Zuge der in Studie 1 und Studie 2 entwickelten Aufgaben vorgenommen. Neben des als plausibel erscheinenden „Übertrags“ von Befunden aus der schulbezogenen empirischen Lehr-Lernforschung in die EB/WB kann die empirische Überprüfung der Struktur von zwei Facetten des PPK, die für Lehrkräfte in Schule und EB/WB übereinstimmt, als ein Beitrag zur Beantwortung der Frage nach der Allgemeinheit des PPK betrachtet werden. Diese Befunde begründen und rechtfertigen (zunächst) das Vorgehen, einen fach- und bildungsbereichsübergreifend einsetzbaren Test auf der Grundlage von empirischen Befunden weitestgehend aus dem Schulkontext zu entwickeln.

Eine weitere Voraussetzung eines solchen Tests, ist, dass er das konzeptualisierte Wissen reliabel und valide erfasst. Studie 1 gibt Hinweise auf die, durch die Konzeptualisierung angenommene, mehrdimensionale Struktur des PPK. In Studie 1 waren nur die Facetten „Wissen über Lehr-Lernziele“ und „Wissen über Individual- und Lernprozessdiagnostik“, für die sich die Frage nach der Allgemeinheit (nicht) stellte, Gegenstand der Analysen. Die geringe Reliabilität bei der Erfassung der Facetten bei Lehrkräften in der EB/WB und die vergleichsweise hohe Anzahl an Aufgaben mit DIF-Effekten im Rahmen von Studie 1 werfen die Frage nach der Dimensionalität der Facetten des Konstrukts auf. Auf der Grundlage dieser Ergebnisse stellte sich nicht nur für diese beiden Facetten, sondern auch für alle anderen Facetten auf einer theoretisch-konzeptuellen Ebene die Frage nach der Dimensionalität der jeweiligen Facetten und die Frage danach, was konkret das auf einer gemeinsamen Dimension liegende Wissen der jeweiligen Facetten sein kann. Diese Frage wurde in Studie 2 explizit für die Facette „Wissen über Lehr-Lernmethoden und -konzepte“ formuliert.

Die Weiterentwicklung des theoretisch-konzeptuellen Verständnisses der Facette „Wissen über Lehr-Lernmethoden und -konzepte“ in Studie 2 vor dem Hintergrund der Frage, welches Wissen das allen Aufgaben dieser Facette gemeinsame ist sowie die empirische Überprüfung der Strukturhypothesen trägt zum Verständnis der Dimensionalität einer Facette des PPK bei. Zudem weist diese eine Möglichkeit auf, welche theoretisch-konzeptuellen

Überlegungen hinsichtlich der Dimensionalität für weitere Facetten bzw. das PPK insgesamt angestellt werden können. Für diese Facette gelang es, gemeinsames Wissen zu konzeptualisieren (Studie 2). Allerdings zeigte sich empirisch nur ein vergleichsweise schwacher Einfluss dieser den allen Aufgaben gemeinsamen Dimension auf die Lösungswahrscheinlichkeit der Aufgaben. Vielmehr scheinen spezifische (Wissens)Dimensionen einen größeren Einfluss auf die Lösungswahrscheinlichkeit zu haben. Interpretiert man alle in Studie 2 und 3 spezifizierten Dimensionen als Wissensdimensionen und nicht als Spezifitäten – für die Möglichkeit dieser Interpretation gibt Studie 3 Hinweise durch plausible Zusammenhangsmuster beruflicher und berufsbiographischer Merkmale mit diesen Dimensionen –, gelingt die Erfassung des Wissens über Lehr-Lernmethoden und -konzepte bei EB/WB-Lehrkräften mit unterschiedlicher erziehungs- und bildungswissenschaftlicher Vorbildung valide und mit guter Reliabilität; allerdings nur dann, wenn man nicht zwischen den verschiedenen Wissensbereichen des Wissens über Lehr-Lernmethoden und -konzepte differenziert.

Mit Hilfe dieses – hinsichtlich der Allgemeinheit der Befunde und der „Gültigkeit“ der Testwerte validierten – Tests können Hinweise auf die Frage nach dem Zusammenhang zwischen Vorbildung und Erfahrung von EB/WB-Lehrkräften mit dem Wissens über Lehr-Lernmethoden und -konzepte gewonnen werden (Studie 3). Für die empirische Lehr-Lernforschung bzw. die Lehrkräfteforschung in der EB/WB kann der Befund zum Tätigkeitsschwerpunkt in allgemeiner Erwachsenenbildung und beruflich-betrieblicher Weiterbildung und dem Wissen über Lehr-Lernmethoden und -konzepte interessant sein. Lehrkräfte in der beruflich-betrieblichen WB verfügten über mehr Wissen als Lehrkräfte in der allgemeinen EB. Dieser Zusammenhang gibt Hinweise darauf, dass es möglicherweise für das Lehren und Lernen sehr grundlegende Unterschiede zwischen Lehrkräften dieser beiden „Bereiche“ gibt, die durch verschiedene Aspekte begründet sein können, z.B. das Eigenengagement der Lehrkräfte begründet durch die Relevanz der Tätigkeit und/oder „bessere“ Fort- und Weiterbildungsmöglichkeiten. Der Befund kann Ausgangspunkt sein für zukünftige die allgemeine EB und die beruflich-betrieblich WB vergleichende Studien. Ein weiterer zentraler Befund ist der Zusammenhang zwischen dem absolvierten Umfang an Fort- und Weiterbildung mit dem erfassten Wissen; vor allem der Hinweis, dass Lehrkräfte mit wenig Lehrerfahrung von Fort- und Weiterbildungen besonders profitieren, ist aus einer praktischen Perspektive interessant. Eine – aus Sicht der Allgemeinen Didaktik bzw. ganz generell aus einer „pädagogisch-erziehungswissenschaftlichen Perspektive“ trivial erscheinende und nicht „empiriebedürftige“ – praktische Implikation dieses Befunds kann sein, dass möglichst zu Beginn bzw. bevor die Lehrkräfte ihre Lehrtätigkeit beginnen, mit deren Ausbildung begonnen werden sollte (ähnlich auch Goeze, 2016; Tietgens, 1988).

Nimmt man den sich zeigenden positiven Zusammenhang zwischen dem Umfang an Fort- und Weiterbildung und dem erfassten Wissen als solchen ernst, dann kann dieser einen Zusammenhang zwischen dem in den Fort- und Weiterbildungen der EB/WB thematisierten Wissen mit dem in dem Test thematisierten Wissen ausdrücken. Das Wissen wurde anhand von Aufgaben ermittelt, die sich aus *empirischen Befunden aus dem Schulkontext* speiste. Dies würde bedeuten, dass das in den Fort- und Weiterbildungen von „EB/WBlern“ für „EB/WBler“ thematisierte Wissen ähnliches Wissen ist, wie es für den Schulkontext thematisiert wird und daher ggf. allgemeines Wissen ist. Dieser Befund kann ein weiterer Hinweis auf die Allgemeinheit (von Facetten) des PPK sein.

9.2 Kritische Betrachtung der berichteten Befunde

Bei der Interpretation der vorgestellten Befunde ist es wichtig, einerseits die Stärken, andererseits aber auch die Einschränkungen zu berücksichtigen. Eine Stärke der vorliegenden Arbeit ist es, dass bei der Modellierung der Dimensionalität von Facetten des PPK theoretisch-konzeptuell begründet verschiedene Messmodelle in Betracht gezogen wurden. In Studie 2 und 3 wurde eine angemessenere Berücksichtigung lokaler stochastischer Abhängigkeiten verglichen zu Studie 1 vorgenommen, durch ein Modell der „Bifactor-Modell-Familie“, die in den vergangenen Jahren (wieder) verstärkt bei der Messmodellierung verwendet werden und eben ihre Stärke in der Berücksichtigung bzw. dem Sichtbarmachen lokaler stochastischer Abhängigkeiten haben (Reise, 2012; Reise et al., 2010).

Zudem wurde über die drei Studien hinweg der Umgang mit fehlenden Werten reflektiert (Lüdtke et al., 2007; Robitzsch, 2016; Rose et al., 2010) und für die jeweilige Studie versucht, die angemessenste Lösung zu finden. Darüber hinaus wurde bei den statistischen Analysen in den drei Studien durch die Berücksichtigung der latenten Variablenstruktur eine Korrektur des Messfehlers vorgenommen (Kline, 2016) und versucht, die Anzahl der zu schätzenden Parameter in einem angemessenen Verhältnis zur Verfügung stehenden Stichprobengröße zu halten (grundlegend Bentler & Chou, 1987).

Eingeschränkt werden die Befunde der drei Studien ganz grundlegend durch die Verwendung von nicht repräsentativen und querschnittlichen sowie kleinen ad-hoc-Stichproben (die auf die Teststärke einen Einfluss haben) und der Verwendung einer nur bedingt die Facetten repräsentierenden Auswahl an Aufgaben.

Eine Unsicherheit bzw. ggf. ganz grundlegende Einschränkung der Befunde kann das Vorgehen sein, einen Test zur Erfassung des allgemeinen PPK von Lehrkräften aus EB/WB und Schule nur auf der Grundlage von Befunden aus dem Schulkontext zu entwickeln. Weiterhin unklar ist, ob und inwiefern diesselben Wirkbeziehungen, für die es empirische Befunde für den Schulkontext gibt, auch für die EB/WB „gelten“. Dies zu überprüfen bzw. für

diese Annahme weitere Befunde zu finden, muss Gegenstand weiterer Studien sein, die in den nachfolgenden Kapiteln skizziert werden.

9.3 Verbleibende Desiderate: Potentiale für und Ausblick auf anschlussfähige Forschungsmöglichkeiten

Der Ausblick auf an diese Dissertationsarbeit *mögliche bzw. denkbare* anschließende Arbeiten zielt in einem ersten Schritt ab auf den Prozess der Aufgaben- und Testentwicklung sowie die Testvalidierung. In einem zweiten Schritt werden weitere Vorgehensmöglichkeiten, unter Verwendung des weiterentwickelten Tests, für die Erforschung der Frage nach der Allgemeinheit des PPK vorgestellt, die gleichzeitig auch einen Beitrag zur empirischen Lehr-Lernforschung in der EB/WB leisten kann.

9.3.1 Aufgabenweiterentwicklung

Testentwicklung ist ein iterativer Prozess: „Real-world assessment design processes are usually iterative with prototypes and cycles; things learned at later stages of design often prompt the designer to revisit, rethink, and revise work done at the earlier stages” (Almond et al., 2015, S. 24). Das im Rahmen der vorliegenden Arbeit gewählte Vorgehen, durch eine Konzeptualisierung den Rahmen des Aufgabenuniversums zu definieren, erscheint bei der Entwicklung neuer Tests und Fragebögen, wie es für die EB/WB an vielen Stellen der Fall ist/sein wird, empfehlenswert; v.a. eine Expertenbefragung ist zur Rückversicherung der Gültigkeit der Konzeptualisierung hilfreich.

Diesen Schritten folgend ist für die Testweiterentwicklungen empfehlenswert, möglichst früh zu verbalisieren, was ganz konkret das gemeinsame Wissen ist, das bei der Beantwortung der Aufgaben nötig ist, wie in Studie 2 vorgenommen. Anregungen für zukünftige Test(weiter)entwicklungen kann dabei das evidence-centered design geben (Almond et al., 2015), im Rahmen dessen die kausale Beziehung zwischen Konstrukt und Aufgaben sowohl auf einer theoretisch-konzeptuellen als auch auf empirischer Ebene eine große Rolle spielt. Ein solches Vorgehen kann zudem bei der (theoretisch-konzeptuellen) Identifikation von etwaigen ungewollten Mehrfachladungen (within item multidimensionality) einer Aufgabe bzw. bei geschlossenen Aufgaben von deren Antwortalternativen hilfreich sein. Die Vermeidung von etwaigen Abhängigkeiten der Antworten auf eine Aufgabe von mehreren theoretisch-konzeptuellen Konstrukten und damit latenten Variablen ist damit möglich. Dies präzisiert nicht nur die theoretisch-konzeptuelle Grundlage, sondern führt bestenfalls zu Einfachladungen (between item multidimensionality, siehe z.B. bei Wu et al. (2007)) und damit einer Einfachstruktur (simple structure, grundlegend Thurstone (1947)) und verbessert (bestenfalls) die Skalierbarkeit, die Validität und die Reliabilität eines Tests.

Für das PPK, einem sehr breiten und aus vielen unterschiedlichen Wissensbereichen bestehenden Konstrukt, hat dies möglicherweise zur Folge, dass sich die Entwicklung mehrerer Tests, die ggf. nicht auf Facetten- sondern auf Subfacettenebene Bereiche des PPK abbilden, anbietet, wodurch etwaige unplausible und ggf. folgenreiche Vereinfachungen hinsichtlich der Dimensionalität des PPK vermieden werden können (siehe auch Borsboom & Wijsen, 2017). Aus der Perspektive der Forschung sind Tests, die „enge“ Wissensbereiche abdecken, interessant, weil ggf. dann auch präzisere Rückschlüsse getroffen werden können hinsichtlich der Frage, „was wirkt?“, wenn solche Tests in entsprechenden Wirkungsstudien eingesetzt werden (Terhart, 2012). Diese Erkenntnisse können zudem auch in angemessener Form in die Ausbildung der Lehrkräfte eingehen. Darüber hinaus sind im Rahmen von LSA oder anderen Studien der empirischen Lehr-Lernforschung häufig kurze Tests besser einsetzbar. Für die Aufgabenentwicklung bietet sich vor dem Hintergrund der Abhängigkeit der Parameterschätzung von der Stichprobengröße zunächst die Fokussierung auf eine Facette an, auch vor dem Hintergrund der Erfahrungen, die bei der Rekrutierung der nötigen Stichprobengrößen zur Skalierung und Validierung des Tests gemacht wurden, die in der EB/WB immer auch an finanzielle Ressourcen gekoppelt und dadurch begrenzt ist.

Aus einer theoretisch-konzeptuellen und pragmatischen Sicht (da bereits aussichtsreich damit begonnen wurde, vgl. Studie 2) ist besonders die Entwicklung von weiteren Aufgaben zur Erfassung der Facette „Wissen über Lehr-Lernmethoden und -konzepte“ interessant und sinnvoll. Denn das als den Inhalt des Generalfakors dieser Facette deklarierte Wissen der Lehrkraft über die Notwendigkeit einer pädagogisch-didaktisch urteilskräftigen Relationierung von Lehr-Lernziel einerseits und die Wahl der dafür ‚passenden‘ sichtsstrukturellen methodischen Vorgehensweise andererseits, erscheint aus theoretisch-konzeptueller Perspektive zentrales Wissen für die gelingende Gestaltung von Lehr-Lernsituationen zu sein. Nicht nur mit Blick auf die im Rahmen der BDT betonten Inhaltsvalidität, sondern auch hinsichtlich der Verbesserung der bislang erreichten Reliabilität, erscheint bei der Testweiterentwicklung der Fokus auf diese Facette und die Erarbeitung eines umfangreichen Aufgabenpools, und dessen Verwaltung in einer Itembank (Haladyna & Rodriguez, 2013), sinnvoll zu sein. Ein solcher Aufgabenpool kann dann, angepasst an den aktuellsten Forschungsstand, ständig aktualisiert und weiterentwickelt werden. Ein umfangreicher und flexibel verwendbarer Aufgabenpool ist vor dem möglichen Hintergrund der Verwendung des Tests, z.B. auf (weitestgehend frei zugänglichen) Online-Plattformen nötig, da (bestenfalls) viele Testnehmer den Test bearbeiten, wozu unterschiedliche Versionen bzw. Zusammenstellungen nötig sind und nicht nur eine, der Community ggf. bekannte und deswegen ggf. später invalide Version. Hinweise darauf, welche Lehr-Lernmethoden von EB/WB-Lehrkräften besonders häufig (für welche Lehr-Lernziele) eingesetzt werden und welches Wissen daher von besonderer Relevanz für einen Test ist, wird eine geplante

Untersuchung von Bonn, Marx, Goeze und Schrader (in Vorbereitung) voraussichtlich geben können.

Hinsichtlich des Aufgabenformats des Tests stellt sich die Frage, wie die begonnene Testentwicklung weiter fortgeführt werden soll. Den in Kapitel 3 genannten Vorteilen videobasierter Aufgaben stehen einige Nachteile gegenüber. Ein Nachteil videobasierter Aufgaben ist bei deren Entwicklung zu sehen, da für relevante Aufgabenstellungen passende Videosequenzen entweder in vorhandenem Material gesichtet (z.B. Marx et al., 2017) oder nachgestellt (z.B. Beck et al., 2008) werden müssen. Zwei Möglichkeiten zur Entwicklung von Aufgaben bei zur Verfügung stehenden Videosequenzen gibt es: Die Videosequenzen werden auf etwaige für eine Aufgabenstellung passende Abschnitte hin analysiert, z.B. mit Hilfe eines Beobachtungsbogens, dessen Kategoriensystem die in Studie 1 vorgestellte Konzeptualisierung sein kann. Oder eine Aufgabenstellung wird entwickelt und dann eine passende Sequenz im Videomaterial gesucht. Bei beiden Vorgehensweisen besteht letztlich das Problem, empirische Befunde zu finden, die zur Identifikation „richtiger Antworten“ verwendet werden können, da die Aufgabenstellung zumeist sehr stark auf die Videosequenz angepasst werden muss und dadurch sehr spezifisch wird. Beide Vorgehensweisen haben sich als sehr zeitaufwendig erwiesen, ebenso das Nachstellen von Lehr-Lernsituationen, bei denen sich zudem die Frage nach der Authentizität der Darstellung der Situation stellt. Der Zeitaufwand bei der Entwicklung videobasierter Aufgaben war im Rahmen des Projekt Think ein Grund dafür, dass zu einem großen Teil textbasierte Aufgaben entwickelt wurden. In die in Studie 1 und Studie 2 berichteten Analysen gingen aus Mangel an passenden Aufgaben keine videobasierten Aufgaben mit ein.

Ein weiterer bedenkenswerter Punkt ist, dass die Akteure in den Videosequenzen mit der Zeit „aus der Zeit gefallen“ erscheinen werden (z.B. Haarschnitte, Kleidung), was die Verwendung ggf. verunmöglicht oder suboptimal erscheinen lässt, da ggf. die Testnehmer abgelenkt werden. Die Videosequenzen von Zeit zu Zeit zu ersetzen ist sehr zeitaufwendig, da entweder mit Nachstellung und Videodreh oder Suche in vorhandenem Videomaterial verbunden, und damit letztlich kostenintensiv. Unabhängig davon kommt bei videobasierten Aufgaben hinzu, dass in den Videosequenzen vorkommende Charaktere ggf. Assoziationen auslösen und ablenken (wobei der entsprechende Umgang damit auch Ausdruck professioneller Kompetenz sein kann).

Ein weiterer Nachteil videobasierter Aufgaben, der sich im Verlauf der Auswertung der Aufgaben zeigte, war, dass die videobasierten Aufgaben verglichen mit den textbasierten Aufgaben sehr einfach zu lösen waren, was sich bei unveröffentlichten Analysen der Pilotstudiendaten und der Hauptstudiendaten zeigte. Grund hierfür war, dass die Erstellung passender Distraktoren schwierig war.

Insgesamt stellt sich die Frage, ob die Nachteile hinsichtlich des Entwicklungsaufwands videobasierter Aufgaben nicht die Vorteile überwiegen. Eine Kompromissmöglichkeit wäre es, textbasierte Aufgaben um videobasierte Aufgaben anzureichern, wobei sich dann bei der Auswertung bzw. Evaluierung des Tests die Frage nach Effekten der Aufgabendarbietung stellt (z.B. Brown, 2006), die nur häufig nur schwer zu beantworten ist bzw. nur mit entsprechenden Untersuchungsdesigns und ausreichender Anzahl an Aufgaben (z.B. Gehrler, 2017).

Für die Neu- und Weiterentwicklung der textbasierten Aufgaben ist das folgende Vorgehen hinsichtlich des Antwortformats denkbar, bei dem das Resultat ein Test bestehend nur aus Aufgaben mit freiem Antwortformat, nur mit geschlossenen (single oder multiple choice) Antwortformat oder eine Kombination dieser Antwortformate sein kann. Eine Kombination verschiedener Antwortformate stellt prinzipiell kein Problem dar, wie verschiedene LSA zeigen (für einen Überblick siehe z.B. Gierl et al., 2017).

In einem ersten Schritt der Testweiterentwicklung ist es denkbar, relevante Aspekte des Wissens über Lehr-Lernmethoden und -konzepte in Aufgabenstellungen zu formulieren, die zunächst dahingehend überprüft werden, ob diesen aus theoretisch-konzeptueller Sicht gemeinsames Wissen und damit Eindimensionalität zugrunde liegt und ob diese eine inhaltsvalide Auswahl darstellen bzw. das Konstrukt aus Sicht der BDT angemessen repräsentieren. Dazu könnte eine Vielzahl derzeit noch ungenutzter Aufgaben und Ideen aus dem Projekt Think verwendet werden.

In einem zweiten Schritt könnten diese dann einer ausreichenden Anzahl an Lehrkräften aus der EB/WB zur Beantwortung im freien Antwortformat vorgelegt werden, mit dem Ziel, die Antworten auf typische Fehler bzw. typische Fehlerkonzepte hin zu untersuchen, was für die Erstellung von Distraktoren für Multiple-Choice Aufgaben ein gängiges Vorgehen ist (Haladyna & Rodriguez, 2013).

In einem dritten Schritt könnte überprüft werden, inwiefern die in Kapitel 3 vorgeschlagene softwarebasierte Auswertung der offenen Aufgaben erfolgsversprechend sein könnte, die eine ähnlich hohe Auswertungsobjektivität verspricht wie bei gebundenen Aufgabenformate. Gelingt dies nicht bzw. sprechen weitere Gründe wie z.B. eine hohe Bearbeitungsdauer gegen ein offenes Antwortformat, können auf der Grundlage der freien Antworten Distraktoren entwickelt werden – bestenfalls auf der Grundlage typischer Fehlerkonzepte. Angestrebtes Ziel bei der Entwicklung von Aufgaben mit geschlossenem Antwortformat wäre es, die Validität des Tests einschränkende Ratewahrscheinlichkeit durch ein entsprechendes Format (z.B. zwei richtige Antwortalternativen bei insgesamt fünf) möglichst gering zu halten (vgl. Bühner, 2011).

Falls (bei manchen Aufgaben) ein geschlossenes Antwortformat verwendet werden würde, könnten die Antwortalternativen (Attraktoren und Distraktoren) erneut auf ihre Ein-

bzw. Mehrdimensionalität hin überprüft werden und in einer weiteren Vorstudie durch die Think-Aloud Technique auf etwaige Missverstehensquellen hin getestet und schließlich dann in einer Hauptstudie an einer ausreichenden Stichprobe an EB/WB-Lehrkräften erprobt werden. Die nötige Stichprobengröße könnte in Abhängigkeit der voraussichtlich zu schätzenden Modellparameter durch Simulationsstudien ermittelt werden, was mit R (R Development Core Team, 2017) oder Mplus (Muthén & Muthén, 1998-2015) umgesetzt werden kann. In die Simulation miteinbezogen werden könnten zudem Konstrukte, die im Rahmen der Testwertvalidierung eine Rolle spielen und Gegenstand des nachfolgenden Kapitels sind. Die Verwendung von Simulationsstudien bei der (Weiter)Entwicklung von Tests als Hilfsmittel wird in jüngerer Zeit vermehrt vorgeschlagen (z.B. Benson, 2018; Morgan, Moore & Floyd, 2018).

9.3.2 Fortsetzung der Testvalidierung

Die faktorielle Validierung der gewonnenen Testwerte könnte in zukünftigen Studien anschließend an die in Studie 2 angesprochene Cattell-Horn-Carroll Theory (Flanagan & Dixon, 2013)) um Maße kognitiver Fähigkeiten ergänzt werden, um die Frage zu beantworten, ob bzw. inwiefern ein Generalfaktor über alle Aufgaben hinweg ggf. durch allgemeine kognitive Fähigkeiten erklärt werden kann bzw. in welchem Ausmaß allgemeine kognitive Fähigkeiten die Wahrscheinlichkeit einer richtigen Antwort erhöhen (siehe dazu auch Blömeke & Jenßen, 2017; Brunner et al., 2008). Welche Maße kognitiver Fähigkeiten dabei letztlich eingesetzt werden könnten, müsste Gegenstand der Planungsprozesse entsprechender Vorhaben sein. Anhaltspunkte bieten die Arbeiten von Voss und Kunter et al. (2011), Blömeke und Jenßen (2017) sowie auch z.B. Schneider und McGrew (2012).

Auch zukünftig erscheint die Berücksichtigung der lokalen stochastischen Unabhängigkeit durch Modelle aus der Familie der bifactor Modelle sinnvoll zu sein, wenn dies aus theoretisch-konzeptueller, methodischer Sicht (hinsichtlich des Aufgabenformats) und empirischer Sicht sinnvoll und nötig erscheint. „Wunschmodell“ für einen zukünftigen Test zur Erfassung des PPK (von Lehrkräften in der EB/WB) wäre aufgrund seiner in Kapitel 4 genannten Vorteile und seiner Unumstrittenheit in der Interpretation (Hooker & Finkelman, 2010; Hooker, Finkelman & Schwartzman, 2009) aber das „herkömmliche“ dichotome Rasch oder 1PL-Modell – wenn sich dieses empirisch so zeigt und aus theoretisch-konzeptueller Sicht nicht das Konstrukt unangemessen stark vereinfacht (Borsboom & Wijsen, 2017).

Neben den in Studie 3 verwendeten Aspekten des nomologischen Netzwerkes des PPK ist vor allem die Erfassung der Unterrichtsqualität und der Lernerfolg der Teilnehmer ein nötiger Schritt weiterer Validierungen, wozu ein Instrument zur Erfassung der Unterrichtsqualität, dessen Entwicklung auf der Grundlage eines Instruments aus dem COACTIV-Projekt (Kunter & Voss, 2011) im Projekt ThinkK begonnen wurde, weiterentwickelt

werden könnte. Zudem kann in einer solchen Studie dann ggf. auch ein Instrument zur Erfassung der Überzeugungen von EB/WB-Lehrkräften sowie ein Instrument zur durch die Lehrkräfte selbsteingeschätzten instruktionalen Praxis eingesetzt werden, dessen Entwicklung ebenfalls im Projekt Think auf der Grundlage von Arbeiten aus dem COACTIV-Projekt begonnen wurde (Marx, Goeze & Schrader, in Vorbereitung). Weitere Möglichkeiten der Validierung bieten gemeinsame Studien mit Tests zur Erfassung von mit dem PPK verwandten Konstrukten, wie der medienpädagogischen Kompetenz (Rohs et al., 2017).

Mit der Überprüfung, ob und inwiefern Facetten des PPK – unter Kontrolle u.a. der im vorangegangenen Abschnitt genannten Variablen, z.B. den Überzeugungen – vermittelt über die selbst- und fremdberichtete Unterrichtsqualität in Zusammenhang mit dem Lernerfolg der Teilnehmer stehen, kann auch (und muss) die im Projekt Think gestellte Frage nach der Allgemeinheit des PPK wieder aufgegriffen werden. Denn selbst wenn die Erfassung des PPK von Lehrkräften in der EB/WB faktoriell valide gelingt und sich auch im Rahmen der nomologischen Validierung entsprechende Hinweise zeigen (vgl. z.B. Studie 3), stellt sich immer noch die Frage, ob das gemessene Wissen „gültiges“ Wissen für die Gestaltung von Lehr-Lernsituationen in der EB/WB ist, da das in den Aufgaben abgefragte Wissen sich auf Wirkungsbeziehungen aus dem Schulkontext bezieht.

9.3.3 Weitere Bearbeitung der Frage der Allgemeinheit des pädagogisch-psychologischen Wissens

Um sich der Beantwortung der Frage der Allgemeinheit Schritt für Schritt zu nähern, bietet es sich an, wie auch im Schulkontext im Rahmen der PPK-Forschung für verschiedene Fächer praktiziert (Beck et al., 2008; Voss et al., 2011), in der EB/WB Lehrkräfte getrennt nach Subpopulation, z.B. qua „Zugehörigkeit“ zu einem Fachbereich oder getrennt nach allgemeiner Erwachsenenbildung und beruflicher Weiterbildung oder dem Reproduktionskontext, zu untersuchen. Hintergrund dieses Vorgehens ist, dass sich die Lehrkräfte ggf. in relevanten Merkmalen (z.B. intrinsisch oder extrinsisch begründetes Engagement hinsichtlich der eigenen Fort- und Weiterbildung, einschlägige pädagogische und/oder akademische Vorbildung) unterscheiden, die z.B. durch die Anforderungen des Faches (Gesundheitsbildung vs. IT-Kurse) bzw. des Kontexts (z.B. Sprachen für Urlaubsreisen in der EB vs. Sprachen für den Beruf in der WB) bedingt sein können. Diese Merkmale können ggf. auf die Dimensionalität des Wissens einen Einfluss haben, der durch ein entsprechendes Vorgehen kontrolliert werden könnte und es wären für bzw. auf die Lehrkräfte eines Bereichs generalisierbare Aussagen möglich.

Ein sehr großer und wichtiger Bereich der (allgemeinen) EB/WB sind Bildungsveranstaltungen zum (Fremd-)Spracherwerb (Sahlender & Schrader, 2017), die im Zuge der Migrationsbewegungen der vergangenen Jahre noch weiter an Bedeutung

gewonnen haben, weswegen sich die Fortsetzung der Untersuchung der Validität und der Frage der Allgemeinheit des PPK in diesem Bereich lohnen würde. Zudem liegen für die Lehrkräfte der allgemeinen Erwachsenenbildung Studien vor, die die Charakteristika der Population gut beschreiben (Autorengruppe wb-personalmonitor, 2016), was für eine die Population repräsentierende Stichprobenziehung unumgänglich ist (z.B. Särndal, Swensson & Wretman, 2003). Im Zuge einer solchen Studie wäre es zudem interessant, das (allgemeine?) sprachbezogene PCK der Lehrkräfte miteinzubeziehen, das vermutlich hinsichtlich des Lernerfolgs der Teilnehmer das wirksamere Wissen ist, worauf Studien aus dem Schulkontext hinweisen (zusammenfassend z.B. Sorge et al., 2017). Einen solchen Test zu entwickeln, könnte ebenfalls Gegenstand zukünftiger Forschungen sein.

Die Weiterentwicklung der angesprochenen Instrumente zur Erfassung der pädagogischen Überzeugungen der Lehrkräfte, des Wissens über Lehr-Lernmethoden und -konzepte sowie des Instruments zur Einschätzung der Unterrichtsqualität aus Sicht der Lehrkräfte und der Teilnehmenden und ein neu entwickeltes Instrument zur Erfassung des sprachbezogenen allgemeinen PCK würden, bestenfalls mit mindestens zwei Messzeitpunkten der Sprachkompetenz der Teilnehmenden, Antworten auf die Frage, „was wirkt“ – und damit auch Antworten auf die Frage nach der Allgemeinheit des PPK – für den Bereich der EB/WB geben können.

Die Entwicklung eines Tests zur Erfassung des PPK, der sich mit Blick auf die geringe bzw. in weiten Teilen nicht vorhandene curriculare Standardisierung und der hochgradigen Spezifik der Lernziele in der EB/WB (Kuper & Schrader, in Druck) im Rahmen einer auf den Bereich der EB/WB bezogenen empirischen Lehr-Lernforschung breit einsetzen lassen würde bzw. einsetzen lässt, war ein erster und wichtiger Schritt in diese Richtung.

10 Literaturverzeichnis

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29 (1), 67-91.
- Adams, R. J., Wilson, M. & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21 (1), 1-23.
- Aebli, H. (2011). *Zwölf Grundformen des Lehrens. Eine allgemeine Didaktik auf psychologischer Grundlage. Medien und Inhalte didaktischer Kommunikation, der Lernzyklus* (14. Aufl.). Stuttgart: Klett-Cotta.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Hrsg.), *Second international symposium on information theory* (S. 267-281). Budapest: Akademiai Kiado.
- Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D. & Williamson, D. M. (2015). An introduction to evidence-centered design. In R. G. Almond, R. J. Mislevy, L. S. Steinberg, D. Yan & D. M. Williamson (Hrsg.), *Bayesian networks in educational assessment* (Statistics for Social and Behavioral Sciences, S. 19-40). New York, NY: Springer.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, USA: American Educational Research Association.
- Arnold, K.-H., Blömeke, S., Messner, R. & Schlömerkemper, J. (Hrsg.). (2009). *Allgemeine Didaktik und Lehr-Lernforschung. Kontroversen und Entwicklungsperspektiven einer Wissenschaft vom Unterricht*. Bad Heilbrunn: Klinkhardt.
- Arnold, R. (2012). *Ich lerne, also bin ich. Eine systemisch-konstruktivistische Didaktik* (Systemische Pädagogik, 2. Aufl.). Heidelberg: Carl-Auer-Verlag.
- Arnold, R., Krämer-Stürzl, A. & Siebert, H. (2005). *Dozentenleitfaden. Planung und Unterrichtsvorbereitung in Fortbildung und Erwachsenenbildung*. Berlin: Cornelsen.
- Autorengruppe wb-personalmonitor (Hrsg.). (2016). *Personal in der Weiterbildung: Beschäftigungsverhältnisse und Tätigkeiten – wb-personalmonitor*. Bielefeld: Bertelsmann.
- Bastian, H. (1997). *Kursleiterprofile und Angebotsqualität* (Theorie und Praxis der Erwachsenenbildung). Bad Heilbrunn: Klinkhardt.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle*

- Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29-54). Münster: Waxmann.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P. et al. (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.
- Benson, N. F. (2018). Introduction to a special issue on simulation studies as a means of informing psychoeducational testing and assessment. *Journal of Psychoeducational Assessment*, Advanced online publication.
- Bentler, P. M. & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16 (1), 78-117.
- Berk, L. E. (2011). *Entwicklungspsychologie* (5., aktualisierte Aufl.). München: Pearson-Studium.
- Berzbach, F. (2005). *Die Ethikfalle. Pädagogische Theorierezeption am Beispiel des Konstruktivismus*. Bielefeld: Bertelsmann.
- Bilger, F. & Strauß, A. (2015). *Weiterbildungsverhalten in Deutschland 2014. Ergebnisse des Adult Education Survey. AES Trendbericht*. Bonn: BMBF.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Hrsg.), *Statistical theories of mental test scores* (S. 17-20). Reading: Addison-Wesley.
- Blömeke, S. & Jenßen, L. (2017). A question of validity: Clarifying the hierarchical nature of teacher cognition. In J.-E. Gustafsson, M. Rosén, K. Yang-Hansen & U. Wolff (Hrsg.), *Cognitive abilities and educational outcomes. A Festschrift in honour of Jan-Eric Gustafsson* (Methodology of educational measurement and assessment, S. 89-107). Cham: Springer.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2010a). *TEDS-M 2008 - Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2010b). *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich*. Münster: Waxmann.
- Boeren, E. (2018). The methodological underdog. A review of quantitative research in the key adult education journals. *Adult Education Quarterly*, 68, 63-79.
- Bonnes, C. & Hochholdinger, S. (2016). Die berufliche Entwicklung und die Wege des qualifikations- und Kompetenzerwerbs von Lehrenden in der berufsbezogenen

- Weiterbildung. *bwp@ Berufs- und Wirtschaftspädagogik - online* (29), 1-18. Verfügbar unter www.bwpat.de/ausgabe29/bonnes_hochholdinger_bwpat29.pdf.
- Borsboom, D. & Wijsen, L. D. (2017). Psychology's atomic bomb. *Assessment in Education: Principles, Policy & Practice*, 24 (3), 440-446.
- Bosche, B. & Strauch, A. (angenommen). Anerkennung von Kompetenzen Lehrender in der Erwachsenen- und Weiterbildung – erste Ergebnisse des Projekts GRETA. *Hessische Blätter für Volksbildung*.
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52 (3), 345-370.
- Bradlow, E., Wainer, H. & Wang, X. (1999). A bayesian random effects model for testlets. *Psychometrika*, 64 (2), 153-168.
- Breloer, G., Dauber, H. & Tietgens, H. (1980). *Teilnehmerorientierung und Selbststeuerung in der Erwachsenenbildung* (Theorie und Praxis der Erwachsenenbildung, Bd. 204). Braunschweig: Westermann.
- Bromme, R. (2008). Lehrerexpertise. In W. Schneider, M. Hasselhorn & J. Bengel (Hrsg.), *Handbuch der pädagogischen Psychologie* (Handbuch der Psychologie, S. 159-167). Göttingen: Hogrefe.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research* (Methodology in the social sciences). New York: Guilford Press.
- Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A. et al. (2010). The structure of academic self-concepts revisited. The nested Marsh/Shavelson model. *Journal of Educational Psychology*, 102 (4), 964-981.
- Brunner, M., Lüdtke, O. & Trautwein, U. (2008). The internal/external frame of reference model revisited: Incorporating general cognitive ability and general academic self-concept. *Multivariate Behavioral Research*, 43 (1), 137-172.
- Buchholz, J. & Hartig, J. (2017). Comparing attitudes across groups. An IRT-based item-fit statistic for the analysis of measurement invariance. *Applied Psychological Measurement*, Advanced online publication.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. aktualisierte und erweiterte Aufl.). München: Pearson Studium.
- Bundesministerium für Arbeit und Soziales. (2014). *Fortschrittsreport „Altersgerechte Arbeitswelt“*. Ausgabe 4: *Lebenslanges Lernen und betriebliche Weiterbildung*. Bonn: BMAS.

- Carsey, T. M. & Harden, J. J. (2014). *Monte Carlo simulation and resampling. Methods for social science*. Thousand Oaks, California: Sage Publications.
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. Yonkers-on-Hudson, New York: World Book Company.
- Chalmers, R. P. (2012). mirt. A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48 (6), 1-29.
- Chen, F. F., West, S. G. & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41 (2), 189-225.
- Cho, S.-J., Suh, Y. & Lee, W.-y. (2016). After differential item functioning is detected. *Applied Psychological Measurement*, 40 (8), 573-591.
- Collins, J. B. & Pratt, D. D. (2011). The teaching perspectives inventory at 10 years and 100,000 respondents. *Adult Education Quarterly*, 61 (4), 358-375.
- Comenius, J. A. (2007). *Große Didaktik. Die vollständige Kunst, alle Menschen alles zu lehren* (10. Aufl.). Stuttgart: Klett-Cotta.
- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Daley, B. J., Martin, L. G. & Roessger, K. M. (2018). A call for methodological plurality. Reconsidering research approaches in adult education. *Adult Education Quarterly*, Advanced online publication.
- Davier, M. von, Gonzalez, E. J. & Mislevy, R. J. (2009). What are plausible values and why are they useful. In M. von Davier & D. Hastedt (Hrsg.), *Issues and methodologies in large-scale assessments. Volume 2* (IERI monograph series, S. 9-36).
- DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, 13 (4), 354-378.
- Depaepe, F. & König, J. (2018). General pedagogical knowledge, self-efficacy and instructional practice. Disentangling their relationship in pre-service teacher education. *Teaching and Teacher Education*, 69, 177-190.
- Dewe, B. (2010). Wissen - Kontext: Relevanz und Reflexion. In B. Dewe (Hrsg.), *Professionswissen und erwachsenenpädagogisches Handeln. Dokumentation der Jahrestagung 2001 der Sektion Erwachsenenbildung der Deutschen Gesellschaft für*

- Erziehungswissenschaft* (Literatur- und Forschungsreport Weiterbildung, Beiheft zum Report, S. 18-30). Bielefeld: Bertelsmann.
- Dicke, T., Parker, P. D., Holzberger, D., Kunina-Habenicht, O., Kunter, M. & Leutner, D. (2015). Beginning teachers' efficacy and emotional exhaustion. Latent changes, reciprocity, and the influence of professional knowledge. *Contemporary Educational Psychology*, 41, 62-72.
- Dobischat, R., Elias, A. & Rosendahl, A. (Hrsg.). (2018). *Das Personal in der Weiterbildung. Im Spannungsfeld von Professionsanspruch und Beschäftigungsrealität*. Wiesbaden: Springer.
- Downing, S. M. & Haladyna, T. M. (Hrsg.). (2006). *Handbook of test development*. Mahwah, N.J.: L. Erlbaum.
- Edwards, J. R. & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5 (2), 155-174.
- Egetenmeyer, R. & Käßlinger, B. (2011). Professionalisation and quality management. Struggles, boundaries and bridges between two approaches. *European Journal for Research on the Education and Learning of Adults*, 2 (1), 21-35.
- Egetenmeyer, R. & Schäßler, I. (Hrsg.). (2012). *Akademische Professionalisierung in der Erwachsenenbildung / Weiterbildung* (Grundlagen der Berufs- und Erwachsenenbildung, Bd. 70). Baltmannsweiler: Schneider Hohengehren.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden* (1. Aufl.). Weinheim: Beltz.
- Europäische Kommission. (2001). *Mitteilung der Kommission: Einen europäischen Raum des lebenslangen Lernens schaffen*. Verfügbar unter https://www.bibb.de/dokumente/pdf/foko6_neues-aus-europa_04_raum-III.pdf
- Europäische Kommission. (2007). *Mitteilung der Kommission an den Rat, das europäische Parlament, den europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen*. Verfügbar unter <http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=LEGISSUM:c11102>
- Flanagan, D. P. & Dixon, S. G. (2013). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest & E. Fletcher-Janzen (Hrsg.), *Encyclopedia of special education* (S. 368-382). Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Frey, A. (2012). Adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 276-293). Heidelberg: Springer.

- Fricke, R. & Sarges, W. (1986). *Psychologie für die Erwachsenenbildung/Weiterbildung. Ein Handbuch in Grundbegriffen*. Göttingen: Hogrefe.
- Friedrich, H. F. & Mandl, H. (1997). Analyse und Förderung selbstgesteuerten Lernens. In F. E. Weinert & H. Mandl (Hrsg.), *Psychologie der Erwachsenenbildung* (S. 237-293). Göttingen: Hogrefe-Verlag.
- Fuhr, T. (1991). *Kompetenzen und Ausbildung des Erwachsenenbildners. Eine Studie zur Professionalisierung der Erwachsenenbildung*. Bad Heilbrunn: Klinkhardt.
- Fuhr, T., Gruchel, J., Kirchgäßner, U., Klaiber, S., Laros, A. & Michalek, R. (2015). Lehren. In J. Dinkelaker & A. v. Hippel (Hrsg.), *Erwachsenenbildung in Grundbegriffen* (S. 93-100). Stuttgart: Kohlhammer.
- Gehrer, K. (2017). *Der Einfluss von Aufgaben- und Textmerkmalen auf Itemschwierigkeit – Eine experimentelle Studie zur Einschränkung der wiederholten Textsicht bei der Bearbeitung von Lesekompetenztestaufgaben (NEPS Working Paper No. 67)*. Bamberg, Deutschland: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.
- Geißler, K. A. (2016). *Anfangssituationen. Was man tun und besser lassen sollte* (Beltz Weiterbildung, 11., überarbeitete und erweiterte Aufl.). Weinheim: Beltz.
- Gierl, M. J., Bulut, O., Guo, Q. & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education. A comprehensive review. *Review of Educational Research*, 87 (6), 1082-1116.
- Gieseke, W. (2018). Professionalisierung der Erwachsenenbildung/Weiterbildung. Historische Prozesse und strukturelle Herausforderung der Gegenwart. In R. Dobischat, A. Elias & A. Rosendahl (Hrsg.), *Das Personal in der Weiterbildung. Im Spannungsfeld von Professionsanspruch und Beschäftigungsrealität* (S. 57-77). Wiesbaden: Springer.
- Gieseke, W., Tietgens, H., Schäffter, O., Venth, A. & Müller, C. (Hrsg.). (1988). *Professionalität und Professionalisierung*. Bad Heilbrunn.
- Goeze, A. (2016). *Professionalitätsentwicklung von Lehrkräften durch videofallbasiertes Lernen – Voraussetzungen, Prozesse, Wirkungen* (Theorie und Praxis der Erwachsenenbildung). Bielefeld: Bertelsmann.
- Goeze, A., Hetfleisch, P. & Schrader, J. (2013). Wirkungen des Lernens mit Videofällen bei Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 16 (1), 79-113.
- Goeze, A. & Schneider, D. (2014). What creates and regulates access to the adult education profession? Research on recruiting practices. In S. Latke & W. Jütte (Hrsg.), *Professionalisation of adult educators: International and comparative perspectives* (S. 201-213). Frankfurt am Main: Peter Lang.

- Goeze, A., Zottmann, J. M., Vogel, F., Fischer, F. & Schrader, J. (2014). Getting immersed in teacher and student perspectives? Facilitating analytical competence using video cases in teacher education. *Instructional Science*, 42 (1), 91-114.
- Gollwitzer, M. (2012). Latent-Class-Analysis. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 296-323). Heidelberg: Springer.
- Gorges, J. (2015). Warum (nicht) an Weiterbildung teilnehmen? *Zeitschrift für Erziehungswissenschaft*, 18 (1), 9-28.
- Gorges, J., Koch, T., Maehler, D. B. & Offerhaus, J. (2017). Same but different? Measurement invariance of the PIAAC motivation-to-learn scale across key socio-demographic groups. *Large-scale Assessments in Education*, 5 (1), 1-28.
- Gorges, J., Maehler, D. B., Koch, T. & Offerhaus, J. (2016). Who likes to learn new things: Measuring adult motivation to learn with PIAAC data from 21 countries. *Large-scale Assessments in Education*, 4 (1), 1-22.
- Großschedl, J., Harms, U., Kleickmann, T. & Glowinski, I. (2015). Preservice biology teachers' professional knowledge. Structure and learning opportunities. *Journal of Science Teacher Education*, 26 (3), 291-318.
- Gustafsson, J.-E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33 (2), 205-233.
- Haberzeth, E. (2010). *Thematisierungsstrategien im Vermittlungsprozess. Empirische Analysen zum Umgang mit Wissen im Planungsprozess von Weiterbildungsangeboten*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Haladyna, T. M. & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 141-171). Heidelberg: Springer.
- Hartig, J. & Höhler, J. (2010). Projekt MIRT. Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (Zeitschrift für Pädagogik, Bd. 56, S. 189-198). Weinheim: Beltz.
- Hasselhorn, M. & Gold, A. (2013). *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren* (3., vollst. überarb. und erw. Aufl.). Stuttgart: Kohlhammer.

- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Heller, J. (2012). *Experimentelle Psychologie. Eine Einführung*. München: Oldenbourg Verlag.
- Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (5. Aufl.). Seelze-Velber: Kallmeyer.
- Herrle, M. (2013). Classroom Management jenseits des Schulunterrichts. *Zeitschrift für Erziehungswissenschaft*, 16 (3), 599-627.
- Hetfleisch, P., Goeze, A. & Schrader, J. (2014). Implementation eines wissenschaftlich erprobten, didaktischen Konzepts. Der Einfluss pädagogischer Autonomie auf die Wirksamkeit in der Praxis. *Zeitschrift für Erziehungswissenschaft*, 17 (2), 297-322.
- Hippel, A. v. & Tippelt, R. (Hrsg.). (2009). *Fortbildung der Weiterbildner/innen. Eine Analyse der Interessen und Bedarfe aus verschiedenen Perspektiven*. Weinheim: Beltz.
- Hof, C. (2001). *Konzepte des Wissens. Eine empirische Studie zu den wissenstheoretischen Grundlagen des Unterrichtens*. Bielefeld: Bertelsmann.
- Hooker, G. & Finkelman, M. (2010). Paradoxical results and item bundles. *Psychometrika*, 75 (2), 249-271.
- Hooker, G., Finkelman, M. & Schwartzman, A. (2009). Paradoxical results in multidimensional item response theory. *Psychometrika*, 74 (3), 419-442.
- Jameson, M. M. & Fusco, B. R. (2014). Math anxiety, math self-concept, and math self-efficacy in adult learners compared to traditional undergraduate students. *Adult Education Quarterly*, 64 (4), 306-322.
- Jank, W. & Meyer, H. (2008). *Didaktische Modelle* (10. Aufl.). Berlin: Cornelsen Scriptor.
- Jonkisz, E., Moosbrugger, H. & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 28-72). Heidelberg: Springer.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- Kade, J., Nolda, S., Dinkelaker, J. & Herrle, M. (Hrsg.). (2014). *Videographische Kursforschung. Empirie des Lehrens und Lernens Erwachsener*. Stuttgart: Kohlhammer.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.),

- Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung* TIMSS -Befunde für Schule und Forschung (S. 43-58). Bonn: Bundesministerium für Bildung und Forschung.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (Fourth Edition). New York: The Guilford Press.
- Klusmann, U. (2011). Allgemeine berufliche Motivation und Selbstregulation. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 277-294). Münster: Waxmann.
- Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O. & Baumert, J. (2008). Engagement and emotional exhaustion in teachers. Does the school context make a difference? *Applied Psychology, 57* (s1), 127-151.
- Kolb, D. A. (1985). *Learning style inventory*. Boston: McBer and Company.
- König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 12* (3), 499-527.
- König, J. & Blömeke, S. (2012). Future teachers' general pedagogical knowledge from a comparative perspective: Does school experience matter? *ZDM Mathematics Education, 44* (3), 341-354.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A. & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education, 38*, 76-88.
- König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M. et al. (2017). Pädagogisches Wissen versus fachdidaktisches Wissen? Struktur des professionellen Wissens bei angehenden Deutsch-, Englisch- und Mathematiklehrkräften im Studium. *Zeitschrift für Erziehungswissenschaft, 21* (3), 1-28.
- Kraft, S., Seitter, W. & Kollwe, L. M. (2009). *Professionalitätsentwicklung des Weiterbildungspersonals*. Bielefeld: Bertelsmann.
- Kron, F. W. (2014). *Grundwissen Didaktik* (6. Aufl.). München: UTB.
- Kubinger, K. D. (2003). Gütekriterien. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 195-204). Weinheim: Beltz PVU.
- Kubinger, K. D. (2005). Psychological test calibration using the rasch model - some critical suggestions on traditional approaches. *International Journal of Testing, 5* (4), 377-394.

- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55-68). Münster: Waxmann.
- Kunter, M., Kunina-Habenicht, O., Baumert, J., Dicke, T., Holzberger, D., Lohse-Bossenz, H. et al. (2017). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung – Ergebnisse des Projekts BilWiss. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals* (S. 37-54). Wiesbaden: Springer VS.
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85-114). Münster: Waxmann.
- Kuper, H. (in Druck). Bildungsbeteiligung Erwachsener unter besonderer Berücksichtigung individueller und sozialer Bedingungen. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, H. Solga et al. (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale*. Bad Heilbrunn: UTB.
- Kuper, H. & Schrader, J. (in Druck). Wirkungen der Weiterbildung und der Steuerung von Weiterbildung. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, H. Solga et al. (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale*. Bad Heilbrunn: UTB.
- Kuper, H. & Schrader, J. (2013). Stichwort. Weiterbildung im Spiegel empirischer Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 16 (1), 7-28.
- Lattke, S. & Jütte, W. (Hrsg.). (2014). *Professionalisation of adult educators: International and comparative perspectives*. Frankfurt am Main: Peter Lang.
- Lau, P. N. K., Lau, S. H., Hong, K. S. & Usop, H. (2011). Guessing, partial knowledge, and misconceptions in multiple-choice tests. *Educational Technology & Society*, 14 (4), 99-110.
- Lauermann, F. & König, J. (2016). Teachers' professional competence and wellbeing. Understanding the links between general pedagogical knowledge, self-efficacy and burnout. *Learning and Instruction*, 45, 9-19.

- Lencer, S. & Strauch, A. (2016). *Das GRETA-Kompetenzmodell für Lehrende in der Erwachsenen- und Weiterbildung*, Deutsches Institut für Erwachsenenbildung. Verfügbar unter <http://www.die-bonn.de/doks/2016-erwachsenenbildung-02.pdf>
- Lenske, G., Thillmann, H., Wirth, J., Dicke, T. & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften. Evaluation des ProwiN-Tests. *Zeitschrift für Erziehungswissenschaft*, 18 (2), 225-245.
- Lenske, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., Liepertz, S. et al. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht. *Zeitschrift für Erziehungswissenschaft*, 19 (1), 211-233.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lievens, F., Peeters, H. & Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personal Review*, 37 (4), 426-441.
- Linninger, C., Kunina-Habenicht, O., Emmenlauer, S., Dicke, T., Schulze-Stocker, F., Leutner, D. et al. (2015). Assessing teachers' educational knowledge. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47 (2), 72-83.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (Zeitschrift für Pädagogik, Beiheft. 51, S. 47-70). Weinheim u.a: Beltz.
- Lohse-Bossenz, H., Holzberger, D., Kunina-Habenicht, O., Seidel, T. & Kunter, M. (2018). Wie fach(un)abhängig ist bildungswissenschaftliches Wissen? *Zeitschrift für Erziehungswissenschaft*, 84 (2), 433.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58 (2), 103-117.
- Maaz, K., Baumert, J., Gresch, C. & McElvany, N. (Hrsg.). (2010). *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (Bildungsforschung, Bd. 34). Bonn [u.a.]: Bundesministerium für Bildung und Forschung (BMBF), Referat Bildungsforschung.
- Maier-Gutheil, C. (2012). Das Lernen Professioneller. Selbstbeobachtung als konstitutives Merkmal von Professionalitätsentwicklung. In H. v. Felden (Hrsg.), *Erwachsenenbildung*

- und Lernen. Dokumentation der Jahrestagung der Sektion Erwachsenenbildung der Deutschen Gesellschaft für Erziehungswissenschaft vom 22.-24. September 2011 an der Universität Hamburg* (S. 131-142). Baltmannsweiler: Schneider Verlag Hohengehren.
- Mandl, H. & Zumbach, J. (2008). *Pädagogische Psychologie in Theorie und Praxis. Ein fallbasiertes Lehrbuch*. Göttingen: Hogrefe.
- Markus, K. A. & Borsboom, D. (2013). *Frontiers in test validity theory. Measurement, causation and meaning* (First Edition). New York, N.Y.: Psychology Press.
- Marx, C., Goeze, A., Kelava, A. & Schrader, J. (2018). Lehrkräfte in der Erwachsenen- und Weiterbildung – Zusammenhänge zwischen Vorbildung und Erfahrung mit dem Wissen über Lehr-Lernmethoden und -konzepte. *Zeitschrift für Weiterbildungsforschung*, 41 (1), 57-77.
- Marx, C., Goeze, A. & Schrader, J. (in Vorbereitung). Selbstberichtete didaktische Praxis von Lehrkräften aus der Erwachsenenbildung: Abhängig vom pädagogisch-psychologischen Wissen und/oder den pädagogischen Überzeugungen?
- Marx, C., Goeze, A. & Schrader, J. (2014a). Adult education teachers' pedagogical psychological knowledge: Elements and test development. In S. Lattke & W. Jütte (Hrsg.), *Professionalisation of adult educators: International and comparative perspectives* (S. 162-182). Frankfurt am Main: Peter Lang.
- Marx, C., Goeze, A. & Schrader, J. (2014b). Pädagogisch-psychologisches Wissen zur Gestaltung von Lehr-Lernsituationen: (Wie) unterscheidet es sich in Erwachsenenbildung/Weiterbildung und Schule? *Hessische Blätter für Volksbildung*, 64 (3), 238-251.
- Marx, C., Goeze, A., Voss, T., Hoehne, V., Klotz, V. K. & Schrader, J. (2017). Pädagogisch-psychologisches Wissen von Lehrkräften aus Schule und Erwachsenenbildung. Entwicklung und Erprobung eines Testinstruments. *Zeitschrift für Erziehungswissenschaft*, 20 (S1), 165-200.
- Maslach, C., Jackson, S. E. & Leiter, M. P. (1996). *Maslach burnout inventory manual* (Third Edition). Mountain View, California: Consulting Psychologists Press.
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47 (2), 149-174.
- Masur, P. K., Teutsch, D. & Trepte, S. (2017). Entwicklung und Validierung der Online-Privatheitskompetenzskala (OPLIS). *Diagnostica*, 63, 256-268.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

- McDonald, R. P. (2003). Behavior domains in theory and practice. *Alberta Journal of Educational Research*, 49, 212-230.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 29, 127-143.
- Messick, S. (1989). Validity. In R. D. Linn (Hrsg.), *Educational measurement* (S. 13-103). Washington, DC: American Council of Education and National Council on Measurement in Education.
- Mietzel, G. (2007). *Pädagogische Psychologie des Lernens und Lehrens* (Lehrbuch, 8., überarb. und erw. Aufl.). Göttingen: Hogrefe.
- Moosbrugger, H. (2012a). Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 228-252). Heidelberg: Springer.
- Moosbrugger, H. (2012b). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 104-117). Heidelberg: Springer.
- Moosbrugger, H. & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 8-25). Heidelberg: Springer.
- Morgan, G. B., Moore, C. A. & Floyd, H. S. (2018). On using simulations to inform decision making during instrument development. *Journal of Psychoeducational Assessment*, 36 (1), 82-94.
- Müsseler, J. & Rieger, M. (Hrsg.). (2017). *Allgemeine Psychologie* (3. Aufl.). Heidelberg: Springer.
- Muthén, L. K. & Muthén, B. O. (1998-2015). *Mplus user's guide*. (Seventh Edition). Los Angeles, CA: Muthén & Muthén.
- Nittel, D. (2000). *Von der Mission zur Profession? Stand und Perspektiven der Verberuflichung in der Erwachsenenbildung* (Theorie und Praxis der Erwachsenenbildung). Bielefeld: Bertelsmann.
- Nittel, D., Schütz, J., Fuchs, S. & Tippelt, R. (2011). Die Orientierungskraft des Lebenslangen Lernens bei Weiterbildnern und Grundschullehrern. Erste Befunde aus dem Forschungsprojekt PAELL. In W. Helsper & R. Tippelt (Hrsg.), *Pädagogische Professionalität* (Zeitschrift für Pädagogik. Beiheft, Bd. 57, S. 167-183). Weinheim: Beltz.

- Nittel, D., Schütz, J. & Tippelt, R. (Hrsg.). (2014). *Pädagogische Arbeit im System lebenslangen Lernens - Ergebnisse komparativer Berufsgruppenforschung*. Weinheim: Beltz.
- OECD. (2001). *Lernen für das Leben. Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000*. Paris: OECD.
- Oser, F. & Baeriswyl, F. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Hrsg.), *Handbook of research on teaching* (Fourth Edition, S. 1031-1065). Washington, DC: American Educational Research Association.
- Osterlind, S. J. & Everson, H. T. (2009). *Differential item functioning* (Second Edition). Los Angeles: Sage.
- Pachner, A. (2013). Selbstreflexionskompetenz. Voraussetzung für Lernen und Veränderung in der Erwachsenenbildung? *Magazin erwachsenenbildung.at* (20), 1-7.
- Pätzold, H. (2011). *Learning and teaching in adult education. Contemporary theories* (Study guides in adult education). Leverkusen: Budrich.
- Peterßen, W. H. (2009). *Kleines Methoden-Lexikon* (2., aktualisierte Aufl., 1. Dr.). München: Oldenbourg.
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, 13 (1), 121-148.
- Pohl, S., Gräfe, L. & Rose, N. (2014). Dealing with omitted and not-reached items in competence tests. *Educational and Psychological Measurement*, 74 (3), 423-452.
- R Development Core Team. (2017). *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, Hrsg.), Vienna, Austria. Verfügbar unter <http://www.R-project.org>
- Rasch, G. (1960). *Probabilistic models for some intelligent and attainment tests*. Copenhagen: Paedagogiske Institut.
- Rauch, D. & Hartig, J. (2012). Interpretation von Testwerten in der IRT. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Aufl., S. 253-274). Heidelberg: Springer.
- Reckase, M. (2009). *Multidimensional item response theory* (Statistics for social and behavioral sciences). New York: Springer.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47 (5), 667-696.

- Reise, S. P., Moore, T. M. & Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *Journal of personality assessment*, 92 (6), 544-559.
- Reise, S. P., Widaman, K. F. & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.
- Research voor Beleid. (2008). *ALPINE - Adult learning professions in Europe. A study of the current situation, trends and issues. Final report*. Zoetermeer.
- Revelle, W. & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74 (1), 145-154.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73 (2), 417-458.
- Robitzsch, A. (2016). Zu nichtignorierbaren Konsequenzen des (partiellen) Ignorierens fehlender Item Responses im Large-Scale Assessment. In B. Suchan, C. Wallner-Paschon & C. Schreiner (Hrsg.), *Die Kompetenzen in Lesen, Mathematik und Naturwissenschaften am Ende der Volksschule. Österreichischer Expertenbericht* (S. 55-64). Graz: Leykam.
- Robitzsch, A., Kiefer, T. & Wu, M. L. (2017). *TAM: Test analysis modules. R package version 2.0-37*. Verfügbar unter <https://CRAN.R-project.org/package=TAM>
- Rohs, M., Rott, K. J., Schmidt-Hertha, B. & Bolten, R. (2017). Medienpädagogische Kompetenzen von ErwachsenenbildnerInnen. *Magazin erwachsenenbildung.at*, 30, 1-14.
- Rose, N., Davier, M. von & Xu, X. (2010). *Modeling nonignorable missing data with item response theory (IRT)*. Research Report No. RR-10-11. Princeton: Educational Testing Service.
- Rösler, F. (2011). *Psychophysiologie der Kognition. Eine Einführung in die Kognitive Neurowissenschaft*. Heidelberg: Spektrum Akademischer Verlag.
- Rost, J. (2004). *Testtheorie – Testkonstruktion* (2., vollst. überarb. und erw. Aufl.). Bern: Huber.
- Rost, J. (2014). Reliabilitäts-Validitätsdilemma. In M. A. Wirtz (Hrsg.), *Dorsch – Lexikon der Psychologie* (17. Aufl., S. 1321). Bern: Verlag Hans Huber.
- Roths, A., Lemos, M. S. & Gonçalves, T. (2017). Motivational profiles of adult learners. *Adult Education Quarterly*, 67 (1), 3-29.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

- Sahlender, M. & Schrader, J. (2017). Spracherwerb bei Zugewanderten und funktionalen Analphabeten. *Zeitschrift für Weiterbildungsforschung*, 40 (3), 275-293.
- Santelices, M. V. & Wilson, M. (2012). On the relationship between differential item functioning and item difficulty. An issue of methods? Item response theory approach to differential item functioning. *Educational and Psychological Measurement*, 72 (1), 5-36.
- Särndal, C.-E., Swensson, B. & Wretman, J. (2003). *Model assisted survey sampling*. New York: Springer.
- Schläfli, A. & Sgier, I. (2008). *Porträt Weiterbildung Schweiz* (Länderporträts Weiterbildung). Bielefeld: Bertelsmann.
- Schneider, W. & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. Harrison (Hrsg.), *Contemporary intellectual assessment: theories, tests, and issues* (S. 99-144). New York: Guilford.
- Schöb, S., Rohs, M., Biel, C. & Scholze, T. (2016). Professionalisierung von Lehrhandeln in einer digital geprägten Lernkultur – Bedarf und Ansatzpunkte der Entwicklung einer Online-Lernumgebung. In O. Dörner, C. Iller, H. Pätzold & S. Robak (Hrsg.), *Differente Lernkulturen – regional, national, transnational* (S. 171-183). Opladen: Barbara Budrich.
- Schrader, J. (in Druck). Institutionelle Rahmenbedingungen, Anbieter, Angebote und Lehr-Lernprozesse der Erwachsenen- und Weiterbildung. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, H. Solga et al. (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale*. Bad Heilbrunn: UTB.
- Schrader, J. (2008). *Lerntypen bei Erwachsenen: Empirische Analysen zum Lernen und Lehren in der beruflichen Weiterbildung* (2. aktualisierte Aufl.). Bad Heilbrunn: Klinkhardt.
- Schrader, J. (2010a). Fortbildung von Lehrenden in der Erwachsenenbildung: Notwendig? Sinnvoll? Möglich? Bedarf und Angebote im Überblick. In J. Schrader, R. Hohmann & S. Hartz (Hrsg.), *Mediengestützte Fallarbeit. Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern* (S. 25-70). Bielefeld: Bertelsmann.
- Schrader, J. (2010b). Reproduktionskontexte der Weiterbildung. *Zeitschrift für Pädagogik*, 56 (2), 267-284.
- Schrader, J., Anders, Y. & Richter, D. (in Druck). Fortbildung des pädagogischen Personals in der frühen Bildung, der Schule und der Erwachsenen- und Weiterbildung. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, H. Solga et al. (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale*. Bad Heilbrunn: UTB.

- Schrader, J., Hohmann, R. & Hartz, S. (Hrsg.). (2010). *Mediengestützte Fallarbeit. Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern*. Bielefeld: Bertelsmann.
- Schulenberg, W. (1957). *Ansatz und Wirksamkeit der Erwachsenenbildung*. Stuttgart: Enke.
- Schulenberg, W., Loeber, H. D., Loeber-Pautsch, U. & Pühler, S. (1978). *Soziale Lage und Weiterbildung*. Stuttgart: Klett-Cotta.
- Schütz, J. & Nittel, D. (2012). Von der Heterogenität zur Vielfalt. Akademische Professionalisierung im Blick einer komparativen pädagogischen Berufsgruppenforschung. In R. Egetenmeyer & I. Schüßler (Hrsg.), *Akademische Professionalisierung in der Erwachsenenbildung / Weiterbildung* (Grundlagen der Berufs- und Erwachsenenbildung, Bd. 70, S. 229–244). Baltmannsweiler: Schneider Hohengehren.
- Schwarz, G. (1978). Estimation the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Seidel, T. & Krapp, A. (Hrsg.). (2014). *Pädagogische Psychologie*. Weinheim: Beltz.
- Seidel, T. & Reiss, K. (2014). Lerngelegenheiten im Unterricht. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (S. 253-276). Weinheim: Beltz.
- Seifert, A., Hilligus, A. H. & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstrumentes zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2 (1), 82-103.
- Sgier, I. & Lattke, S. (Hrsg.). (2012). *Professionalisierungsstrategien der Erwachsenenbildung in Europa. Entwicklungen und Ergebnisse aus Forschungsprojekten*. Bielefeld: Bertelsmann.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4-14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review*, 57, 1-22.
- Siebert, H. (2004). Weiterbildungsbeteiligung und Lernmotivation. *Report - Zeitschrift für Weiterbildungsforschung*, 27 (3), 9-14.
- Siebert, H. (2012). *Didaktisches Handeln in der Erwachsenenbildung. Didaktik aus konstruktivistischer Sicht* (Grundlagen der Weiterbildung, 7., überarbeitete Aufl.). Augsburg: ZIEL.

- Sorge, S., Kröger, J., Petersen, S. & Neumann, K. (2017). Structure and development of pre-service physics teachers' professional knowledge. *International Journal of Science Education*, Advanced online publication.
- Stanik, T. (2016). Mikrodidaktische Planungen von Lehrenden in der Erwachsenenbildung. *Zeitschrift für Weiterbildungsforschung – Report*, 39 (3), 317-330.
- Statistisches Bundesamt. (2015a). *Bildung und Kultur: Allgemeinbildende Schulen. Schuljahr 2014/2015*, Statistisches Bundesamt. Fachserie 11: Reihe 1. Zugriff am 10.05.2016. Verfügbar unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100157004.pdf?__blob=publicationFile
- Statistisches Bundesamt. (2015b). *Bildung und Kultur: Berufliche Schulen. Schuljahr 2014/2015*, Statistisches Bundesamt. Fachserie 11: Reihe 2. Zugriff am 10.05.2016. Verfügbar unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/BeruflicheSchulen2110200127004.pdf?__blob=publicationFile
- Statistisches Bundesamt. (2016). *Bildung und Kultur: Personal an Hochschulen 2015.*, Statistisches Bundesamt. Verfügbar unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PersonalHochschulen2110440157004.pdf?__blob=publicationFile
- Steiner, P. (2010). Die Weiterbildungsakademie Österreich (wba). Erfahrungen und Zugänge zu Beratung und Kompetenzanerkennung für ErwachsenenbildnerInnen. *Magazin erwachsenenbildung.at* (14), 1-6.
- Strijbos, J.-W. & Müller, A. (2014). Personale Faktoren im Feedbackprozess. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 83-134). Münster: Waxmann.
- Strobl, C. (2012). *Das Rasch-Modell. Eine verständliche Einführung für Studium und Praxis* (Sozialwissenschaftliche Forschungsmethoden, Bd. 2, 2., erw. Aufl.). München: Rainer Hampp Verlag.
- Strzelewicz, W., Raapke, H. D. & Schulenberg, W. (1966). *Bildung und gesellschaftliches Bewußtsein. Eine mehrstufige soziologische Untersuchung in Westdeutschland*. Stuttgart: Enke.
- Terhart, E. (2002). Fremde Schwestern. Zum Verhältnis von Allgemeiner Didaktik und empirischer Lehr-Lern-Forschung. *Zeitschrift für Pädagogische Psychologie*, 16 (2), 77-86.

- Terhart, E. (2005). *Lehr-Lern-Methoden. Eine Einführung in Probleme der methodischen Organisation von Lehren und Lernen* (Grundlagentexte Pädagogik, 3. Aufl.). Weinheim: Juventa-Verlag.
- Terhart, E. (2009). *Didaktik. Eine Einführung*. Stuttgart: Reclam.
- Terhart, E. (2012). Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung*, 2 (1), 3-21.
- Terhart, E. (2013). Neuansätze in der Allgemeinen Didaktik: Ein Kommentar. *Jahrbuch für Allgemeine Didaktik*, 3, 219-230.
- Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.
- Tietgens, H. (1967). *Schule und Erwachsenenbildung*. Braunschweig: Georg Westermann Verlag.
- Tietgens, H. (1988). Professionalität in der Erwachsenenbildung. In W. Gieseke, H. Tietgens, O. Schäffter, A. Venth & C. Müller (Hrsg.), *Professionalität und Professionalisierung* (S. 28-75). Bad Heilbrunn.
- Tsai, C.-Y., Li, Y.-Y. & Cheng, Y.-Y. (2017). The relationships among adult affective factors, engagement in science, and scientific competencies. *Adult Education Quarterly*, 67 (1), 30-47.
- Vinepac-Project. (2008). *Handbook for the use of Validpack for the validation of psychopedagogical adult educators' competences*. Timisoara: Editura Mirton.
- Voss, T., Goeze, A., Marx, C., Hoehne, V., Klotz, V. K. & Schrader, J. (2017). Using digital media to assess and promote school and adult education teacher competence. In J. Buder & F. W. Hesse (Hrsg.), *Informational Environments* (S. 125-148). Cham: Springer International Publishing AG.
- Voss, T., Kleickmann, T., Kunter, M. & Hachfeld, A. (2011). Überzeugungen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 235-258). Münster: Waxmann.
- 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.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103 (4), 952-969.

- Voss, T., Kunter, M., Seiz, J., Hoehne, V. & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik*, 60 (2), 184-201.
- Wainer, H. & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24 (3), 185-201.
- Wang, W.-C. & Wilson, M. (2005a). Assessment of differential item functioning in testlet-based items using the Rasch testlet model. *Educational and Psychological Measurement*, 65 (4), 549-576.
- Wang, W.-C. & Wilson, M. (2005b). The Rasch testlet model. *Applied Psychological Measurement*, 29 (2), 126-149.
- Wecker, C. & Fischer, F. (2014). Lernen in Gruppen. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (S. 277-296). Weinheim: Beltz.
- Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Berlin: Springer Gabler.
- Weidenmann, B. (2008). *Handbuch active Training. Die besten Methoden für lebendige Seminare* (2. Aufl.). Weinheim: Beltz.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92 (4), 548-573.
- Weinert, F. E. & Mandl, H. (Hrsg.). (1997). *Psychologie der Erwachsenenbildung*. Göttingen: Hogrefe-Verlag.
- Wild, E. & Möller, J. (Hrsg.). (2015). *Pädagogische Psychologie* (Springer-Lehrbuch, 2. vollständig überarbeitete u. aktualisierte Aufl.). Berlin: Springer.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.
- Wilson, M., Boeck, P. de & Carstensen, C. H. (2008). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 83-110). Toronto: Hogrefe & Huber Publishers.
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8 (3), 370.
- WSF Wirtschafts- und Sozialforschung. (2005). *Erhebung zur beruflichen und sozialen Lage von Lehrenden in Weiterbildungseinrichtungen*. Kerpen: Bundesministerium für Bildung und Forschung.

-
- Wu, M. L., Adam, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest version 2.0. Generalised item response modelling software*. Camberwell, Vic.: ACER Press.
- Ziep, K.-D. (1990). *Der Dozent in der Weiterbildung. Professionalisierung und Handlungskompetenzen*. Weinheim: Deutscher Studien Verlag.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4 (2), 223-233.