Blömeke, Sigrid; Jenßen, Lars; Dunekacke, Simone; Suhl, Ute; Grassmann, Marianne; Wedekind, Hartmut (2015). Leistungstests zur Messung der professionellen Kompetenz frühpädagogischer Fachkräfte. *Zeitschrift für pädagogische Psychologie*, 29, 177-191.

DOI: <http://dx.doi.org/10.1024/1010-0652/a000159>

Post-print version for green Open Access

Running head: LEISTUNGSTESTS FÜR FRÜHPÄDAGOGEN

Leistungstests zur Messung des professionellen Wissens frühpädagogischer Fachkräfte

A test battery for the assessment of early childhood education professionals' knowledge

Sigrid Blömeke (corresponding author), Lars Jenßen

University of Oslo/ Faculty of Education

Centre for Educational Measurement (CEMO)

Niels Henrik Abels hus

N – 0318 Oslo, Norway

Email: sigribl@cemo.uio.no

Tel. ++47 464 18 755

Simone Dunekacke, Ute Suhl, Marianne Grassmann

Humboldt-Universität zu Berlin

Fakultät für Kultur-, Sozial- und Bildungswissenschaften

D – 10099 Berlin, Deutschland

Hartmut Wedekind

Alice-Salomon Hochschule

D – 12627 Berlin, Deutschland

Zusammenfassung

Im Unterschied zur Lehrerforschung ist über das professionelle Wissen frühpädagogischer Fachkräfte kaum etwas bekannt. Der Beitrag stellt eine Testbatterie vor, die auf die Diagnose unterschiedlicher Ausprägungen des pädagogischen, mathematikdidaktischen und mathematischen Wissens angehender Frühpädagogen auf Gruppenebene ausgerichtet ist. In Mehrebenen-Mehrgruppen-Analysen basierend auf einer Vier-Kohorten-Stichprobe ($N = 1\,851$) weisen die Tests konfigurale und metrische sowie annäherungsweise skalare Messinvarianz über die Kohorten, das Geschlecht und die sprachliche Herkunft auf. Die aus der Lehrerforschung übernommene dreidimensionale Wissensstruktur konnte repliziert werden (faktorielle Validität), allerdings weist eine sparsamere zweidimensionale Modellierung eine ebensogute Passung auf. Vorhergesagte Zusammenhänge zwischen Testscores und Mathematik- bzw. Deutschnoten indizieren konvergente und diskriminante Validität. Die Tests sind sensitiv, indem sie vorhergesagte Unterschiede zwischen Fachschülern und Fachhochschülern sowie Ausbildungsbeginn und -ende abbilden. Fachschüler lösen selten Items, die basale Fachsprachlichkeit oder Bezüge auf Theorien erfordern.

Schlüsselwörter: Frühpädagogik, Leistungstest, professionelle Kompetenz, Validität, Messinvarianz, Sensitivität

Abstract

Data about the professional knowledge of Early Childhood Education and Care (ECEC) is scarce. This paper presents a test battery for the assessment of general pedagogical knowledge, mathematical pedagogical content knowledge and mathematical content knowledge of prospective ECEC professionals on the group level. Multi-level multi-group analyses ($N=1.851$) supported configural and metric as well as approximate scalar measurement invariance across cohorts, gender and language background was supported. The 3-dimensional knowledge structure derived from teacher research could be replicated (factorial validity) although the fit of a parsimonious two-dimensional model was as well. Hypothesized correlations between test scores and school grades in German and mathematics supported convergent and discriminant validity. The tests reflected hypothesized differences between vocational school and university students and the beginning and end of their training. Vocational school students have difficulties to solve items which require basic technical terms or theories.

Keywords: Early Childhood Education and Care (ECEC), achievement test, professional competence, validity, measurement invariance, sensitivity

Leistungstests zur Messung der professionellen Kompetenz frühpädagogischer Fachkräfte

1 Einleitung

Der Aufschwung der Lehrer(bildungs)forschung hat zu einem breiten Repertoire an Tests geführt, die eine objektive, reliable, valide, faire und ökonomische Diagnose unterschiedlicher Ausprägungen des fachlichen, fachdidaktischen und pädagogischen Professionswissens angehender und praktizierender Lehrkräfte ermöglichen. Mittlerweile werden zahlreiche Unterrichtsfächer (z.B. Deutschlehrkräfte: Bremerich-Vos & Dämmer, 2013; Physik: Riese et al., 2015; Wirtschaftspädagogik: Bouley et al., 2015) und unterschiedliche Schulstufen abgedeckt (z.B. Mathematiktests für Primarstufenlehrkräfte: Blömeke, Kaiser & Lehmann, 2010; Sekundarstufe I: Baumert et al., 2010; Sekundarstufe II: Buchholtz, Kaiser & Blömeke, 2014). Ebenso existieren Tests, um das pädagogische Wissen (König & Blömeke, 2009) und andere Schlüsselqualifikationen aller Lehrkräfte zu diagnostizieren (Forschungskompetenz: Schladitz, Groß Ophoff & Wirtz, 2015).

Der Bereich der Frühpädagogik hat an diesem Aufschwung bisher nur in Bezug auf die Kompetenz der Fachkräfte teilgenommen, die sprachliche Entwicklung von Kindern zu fördern (Michel, Ofner & Thoma, 2014; Ofner & Thoma, 2014). Über die Struktur und Ausprägung des Wissens von frühpädagogischen Fachkräften in Bezug auf die Förderung der kindlichen Entwicklung in anderen Domänen ist so gut wie nichts bekannt (Anders, 2012). Insbesondere das mathematikbezogene Professionswissen stellt eine gravierende Forschungslücke dar, blickt man auf den engen Zusammenhang zwischen mathematischen Fähigkeiten am Ende des Kindergartens und späteren schulischen Leistungen (Duncan et. al., 2007; Gut, Reimann & Grob, 2012; Krajewski & Schneider, 2009).

Studien machen zugleich deutlich, dass die mathematischen Fähigkeiten von Kindern eng mit der Qualität ihrer Lerngelegenheiten im Kindergarten und dem Professionswissen der frühpädagogischen Fachkräfte zusammenhängen (Kilbanoff, Levine, Huttenlocher, Vasilyeva & Hedges, 2006; Lee, Meadows & Lee, 2003). Letzteres kann wiederum im Rahmen der Ausbildung gefördert werden. So weist Lee (2010) Niveauunterschiede im mathematischen und mathematikdidaktischen Wissen von im Beruf stehenden frühpädagogischen Fachkräften in Abhängigkeit von ihrem Ausbildungsabschluss – Master versus Bachelor – nach. In standardisierter Form wurden bisher allerdings weder dieses Wissen noch die Ausbildungsinhalte erfasst.

Im vorliegenden Beitrag wird diese Forschungslücke angegangen und eine Testbatterie vorgestellt, die auf die Messung und Beschreibung unterschiedlicher Ausprägungen des Professionswissens angehender frühpädagogischer Fachkräfte ausgerichtet ist, um die Stärken und Schwächen ihrer Ausbildung untersuchen zu können. Konkret sind drei Tests zur Erfassung des pädagogischen, mathematikdidaktischen und mathematischen Wissens angehender Fachkräfte entwickelt worden, die nach Beendigung ihrer Ausbildung in Einrichtungen für Kinder im Alter von drei bis sechs Jahren arbeiten können („Kindergarten“ bzw. „Kindertagesstätte“).¹

2 Theoretischer Rahmen und Forschungsstand

2.1 Generelle Prinzipien

Die theoretische Modellierung des Professionswissens frühpädagogischer Fachkräfte orientiert sich zum einen am Stand der pädagogisch-psychologischen Forschung zu Lehrerkompetenzen (Spinath, 2012; Retelsdorf & Südkamp, 2012). Dies gewährleistet eine Anschlussfähigkeit in Bezug auf den Übergang vom Elementar- in den Primarbereich sowie an die internationale Forschung, wo frühpädagogische Fachkräfte vielfach zusammen mit jenen für den Primarbereich ausgebildet werden (Anders, 2012). Bezieht sich das Kompetenzmodell in Grundzügen auf beide Professionen, kann überprüft werden, welche Spezifika im Kompetenzprofil der jeweiligen Berufsgruppe bestehen (Mischo, Wahl, Hendler & Strohmer, 2012). Das Professionswissen wird in der pädagogisch-psychologischen Lehrerforschung üblicherweise in drei Dimensionen ausdifferenziert, und zwar in Fachwissen, fachdidaktisches und pädagogisches Wissen (Weinert, 2001; Shulman, 1986).

Die Spezifika dieses Kompetenzmodells für die Elementarstufe im Unterschied zur Primarstufe wurden unter Bezug auf den entwicklungspsychologischen Forschungsstand zu Lernen, Entwicklung, Diagnose und Förderung von Drei- bis Sechsjährigen herausgearbeitet (Hasselhorn, 2005). Da kognitives, motorisches und sozio-emotionales Lernen in diesem Alter eher implizit und beiläufig stattfindet, können sich Kinder die Welt besonders gut im Spiel aneignen (Oerter, 2008), während Lernen durch instruktional-systematische Maßnahmen in der Regel erst in den folgenden Jahren effizient und nachhaltig gefördert

¹ „Struktur, Niveau und Entwicklung professioneller Kompetenz von Erzieher/innen im Bereich Mathematik (KomMa)“ ist ein Kooperationsprojekt der Humboldt-Universität zu Berlin und der Alice-Salomon-Hochschule Berlin und wird vom Bundesministerium für Bildung und Forschung (FKZ: 01PK11002A) im Rahmen der Förderinitiative „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor (KoKoHs)“ gefördert.

werden kann (Hasselhorn, 2005). Eine wichtige Diagnoseaufgabe ist es aber, früh besondere Begabungen oder mögliche Schwierigkeiten beim Lernen zu identifizieren (Grüßing & Peter-Koop, 2006).

Der Kindergarten kann die frühe Entwicklung fördern, indem ein umfangreiches Spektrum von Imitations- und Rollenspielen bis zu Entdeckungs- und Konstruktionsspielen angeboten und entwicklungspsychologische Prinzipien bei der Wahl der Spiele beachtet werden (Oerter, 2008). Kretschmann (2004) arbeitet heraus, dass sich Kindern mathematische Inhalte nicht von selbst erschließen, sondern dass ihre Aufmerksamkeit durch sprachliche Begleitung auf diese gelenkt und Modellbildungsprozesse angeregt werden müssten. Hasemann (2010) ergänzt, dass mathematische Erfahrungen als solche bewusst gemacht werden müssen, um mathematisches Denken vorzubereiten. Inhaltlich bezieht sich dieses auf Zahlen, Mengen und Operationen; Form, Raum und Veränderung; Größen, Messen und Relationen; sowie Daten, Kombinatorik und Zufall. Prozessbezogen bezieht es sich auf Problemlösen, Modellieren, Argumentieren, Darstellen, Kommunizieren sowie Muster und Strukturen (Rathgeb-Schnierer, 2012). Vorgezogenes systematisches Training hat sich im Vergleich zu selbstgesteuerten Aktivitäten als wenig nachhaltig erwiesen (Montie, Xiang & Schweinhart, 2007). Eine wichtige Ausnahme stellen selektive Interventionen zur Kompensation von Entwicklungsrückständen dar (Hasselhorn, 2010).

Die entwicklungspsychologische Operationalisierung des Kompetenzmodells für frühpädagogische Fachkräfte wurde um eine systematische Analyse der Bildungspläne aller 16 Bundesländer für die Kindergärten sowie der Lehrpläne und Modulhandbücher aller Ausbildungsinstitutionen der 16 Bundesländer – Fachschulen sowie Fachhochschulen als den beiden beruflichen Zugangswegen – ergänzt (Dunekacke et al., 2013). Dieses Kategoriensystem samt Deskriptoren bildet den pädagogischen und mathematikbezogenen Kern an staatlich definierten beruflichen Anforderungen und in der Ausbildung gebotenen Lerngelegenheiten ab. Für die Testentwicklung wurden jene Deskriptoren herangezogen, die in der Mehrheit der Bildungspläne und Lehrpläne angesprochen wurden und zugleich der entwicklungspsychologischen Operationalisierung entsprachen (Details siehe Anhang 1).

Mit der Bewältigung von Anforderungen ist die Vorstellung unterschiedlicher Niveaus verbunden. Als Heuristik für die Itemkonstruktion wurde daher ergänzend ein Niveaumodell entwickelt, das beschreibt, mit welchen Schwierigkeiten die Bewältigung der in einem Item präsentierten Anforderung verbunden ist (Harsch & Hartig, 2011). Als Indikatoren wurden die Art des erforderlichen kognitiven Prozesses (Erinnern von Wissen, Verstehen und Anwenden, Handlungsoptionen generieren; Anderson & Krathwohl, 2001), die kognitive

Komplexität (Zahl der gedanklichen Schritte, die zur Bewältigung einer Anforderung durchlaufen werden müssen, Art der Problemrepräsentation; Embretson & Daniel, 2008) sowie das Niveau des Iteminhalts (Elementar- bzw. Primarstufenmathematik vom höheren Standpunkt; Allmendinger, 2011) herangezogen.

2.2 Fachspezifische Operationalisierung (siehe auch Anhang 1)

Pädagogisches Wissen. Das pädagogische Wissen frühpädagogischer Fachkräfte wurde in bildungstheoretische, psychologische und didaktisch-methodische Grundlagen von Erziehungs- und Bildungsprozessen ausdifferenziert (Jenßen et al., im Druck). Die Inhalte beziehen sich auf zentrale generische, nicht-mathematikbezogene Theorien und Forschungserkenntnisse zur Entwicklung und Förderung von Kindern zwischen drei und sechs Jahren mit einem Schwerpunkt auf altersangemessenem implizitem spielerischen Lernen.

Mathematikdidaktisches Wissen. Das mathematikdidaktische Wissen umfasst die Entwicklung und Diagnose mathematischer Fähigkeiten bei Kindern im Alter von drei bis sechs Jahren sowie die spielerische Gestaltung mathematischer Lernumgebungen mit dem Ziel der inzidentellen und impliziten Förderung mathematischer Fähigkeiten (Jenßen et al., im Druck). Ein Schwerpunkt liegt auf der Diagnose und Kompensation von Entwicklungsrückständen.

Mathematisches Wissen. Als Mathematik der Elementarstufe „vom höheren Standpunkt“ geht es um eine verstehensorientierte begriffliche Durchdringung der mathematischen Inhalte (Allmendinger, 2011). Mit Blick auf den Übergang in die Primarstufe ist auch die Mathematik des Anfangsunterrichts vom höheren Standpunkt Bestandteil des Konstrukts. Dieses ist nach Inhaltsbereichen und prozessbezogenen Fähigkeiten ausdifferenziert (Jenßen et al., im Druck).

3 Testkonstruktion und Forschungsstand

Ziel der Testentwicklung war, mithilfe von drei Leistungstests auf Gruppenebene strukturelle Aussagen und Beschreibungen von stärkeren und schwächeren Ausprägungen des mathematischen, mathematikdidaktischen und pädagogischen Wissens angehender frühpädagogischer Fachkräfte zu verschiedenen Zeitpunkten der Ausbildung vornehmen zu können. Hierfür wurden 117 Items – 53 für mathematisches, 42 für mathematikdidaktisches und 22 für pädagogische Wissen – und damit rund doppelt so viele entwickelt, als mit Blick

auf die Testökonomie und die antizipierte Bearbeitungszeit eingesetzt werden können. Individualdiagnostische Aussagen waren nicht angestrebt.

Präpilottierungen mit kleinen Gruppen an angehenden Frühpädagogen, systematische Expertenreviews und zwei Pilotstudien mit 454 Probanden von Fachschulen und Fachhochschulen aus Berlin, Brandenburg, Mecklenburg-Vorpommern und Hamburg sowie 354 angehenden Frühpädagogen von Fachschulen in Berlin und Niedersachsen hatten zum Ziel, diese Itemzahl basierend auf empirischen Kriterien zu reduzieren. Um die Fachkräfte zeitlich nicht zu überfordern, wiesen die Erhebungen ein Multi-Matrix-Design auf, in dem nicht alle Teilnehmenden alle Items bearbeiten, diese über Ankeritems aber auf einer gemeinsamen Skala verortet werden können.

Itemformat. Die drei Itempools enthielten Items im offenen, geschlossenen (Multiple Choice) und gebundenen Format (Zuordnungsaufgaben). Mit wenigen Ausnahmen leitete ein Aufgabenstamm nur ein Item ein, um lokale Abhängigkeiten zu vermeiden. Die Distraktoren der Multiple-Choice-Items wurden unter der Perspektive ihrer Attraktivität und Plausibilität entwickelt, um gleichwertige Antwortoptionen zu erreichen.

Testfairness. Die Itemformulierungen erfolgten geschlechtsneutral, um das Risiko des „stereotype threat“ zu vermeiden, das besteht, wenn sich Sprache bzw. Inhalte von Aufgaben an einer geschlechtsspezifischen Lebenswelt orientieren (Cadinu, Maass, Rosabianca & Kiesner, 2005). Die Items wiesen zudem eine einfache, für die Zielgruppe frühpädagogischer Fachkräfte verständliche Sprache auf, um den Einfluss des Sprachniveaus angesichts der heterogenen Voraussetzungen (Mittlerer Bildungsabschluss und kaum fachbezogene Lerngelegenheiten an den Fachschulen; Abitur und Lehrveranstaltungen in Mathematikdidaktik an den Fachhochschulen) zu minimieren. Lehrende von Fachschulen haben das Einhalten dieser Standards bei allen Items überprüft.

Inhaltsvalidität. Die Items deckten alle Dimensionen, Subdimensionen und Deskriptoren des Kompetenzmodells ab. Sie unterschieden sich zudem in ihrer vorab eingeschätzten Schwierigkeit, um die Variation des Professionswissens in der Zielpopulation abbilden zu können (Jenßen et al., im Druck). Alle Items wurden in Prä-Tests von angehenden frühpädagogischen Fachkräften bearbeitet, um Fehlinterpretationen und Verständnisschwierigkeiten aufzudecken. In einem *cognitive lab* wurden über lautes Denken kognitive Prozesse und Strategien zur Lösung der Items identifiziert.

Da es sich bei frühpädagogischen Fachkräften um eine relativ empirieferne Zielgruppe handelt, wurde auch die Augenscheinvalidität als Auswahlkriterium einbezogen. Items, die in diesen Prä-Tests auf nur geringe Akzeptanz stießen, wurden ausgeschlossen. Im Rahmen

eines standardisierten Expertenreviews mit 15 externen Wissenschaftlern – identifiziert über Forschung und Lehre in der Frühpädagogik bzw. der elementaren mathematischen Bildung und mindestens zehn einschlägige Publikationen, unter Einbezug der unterschiedlichen Paradigmen – sowie neun Praktikern – identifiziert über eine Leitungsfunktion im Kindergarten oder eine mindestens zehnjährige Berufserfahrung – wurde die Inhaltsvalidität der drei Tests auf Item- und Testebene geprüft. Diesem Review zufolge repräsentierten 71 Items das jeweilige Konstrukt uneingeschränkt: 29 Mathematik-, 28 Mathematikdidaktik- und 14 Pädagogik-Items. Weiteren 37 Items wurde unter dem Vorbehalt einer Überarbeitung Inhaltsvalidität bescheinigt (Jenßen, Dunekacke & Blömeke, 2015). Die drei Gesamttests deckten alle relevanten Inhalte in angemessener Gewichtung ab.

Durchführungsobjektivität. Die operativen Untersuchungsprozesse wurden in einem Manual vorab festgelegt. Dies betraf sowohl den Einsatz des Testmaterials als auch die Zeitbegrenzungen und Instruktionen. Die Testleiter wurden einem entsprechenden Schulungsverfahren unterzogen.

Auswertungsobjektivität. Für Items mit offenen Antwortformaten enthielt ein Codebook detaillierte Auswertungsregeln. 20% der Fragebögen aus den Pilotstudien wurden doppelt codiert. Die auf dieser Basis berechneten Übereinstimmungsmaße Cohens Kappa ($min-max = .64-.88$, $.64-.92$ bzw. $.69-.86$; $Md = .76$, $.73$ bzw. $.78$ für Pädagogik, Mathematikdidaktik bzw. Mathematik) und Yules Y ($min-max = .95-1.00$, $.92-1.00$ bzw. $.95-1.00$; $Md = .98$, $.97$ bzw. $.99$) wiesen auf sehr gute Inter-Rater-Reliabilitäten hin.

Reliabilität. Die Skalierung der Pilotdaten erfolgte mithilfe des Raschmodells. Als Fit-Maße für die Passung von Modellen und Daten wurden Weighted-Mean-Square-(*MNSQ*)-Werte verwendet. Als weiteres Kriterium für die Itemauswahl wurden mit Blick auf die Abbildung eines breiten Fähigkeitsspektrums die Testinformationskurven herangezogen. Für die Mathematik konnte eine Skala aus 24 Items bei zufriedenstellender Reliabilität (*expected-a-posteriori estimate*, $EAP = .74$) gebildet werden (Jenßen, Dunekacke & Blömeke, 2015). Für Mathematikdidaktik und Pädagogik wurden dagegen trotz des Ausschlusses schlecht passender Items nur wenig befriedigende Reliabilitäten erreicht ($EAP = .58$ bzw. $.51$ mit jeweils 12 Items), was die Überarbeitung von Items und Neukonstruktionen nach sich zog. In diese wurden Fachschullehrer und Kindergartenleiterinnen einbezogen, da das Wissen der Fachhochschüler reliabler erfasst werden konnte als das der Fachschüler.

Faktorielle Validität. Konfirmatorische Faktorenanalysen stützten die konzeptionelle Ausdifferenzierung des frühpädagogischen Wissens in mathematisches, mathematikdidaktisches und pädagogisches Wissen (Jenßen et al., im Druck). Ein

entsprechendes dreidimensionales Modell bildete die Daten signifikant besser ab als ein eindimensionales Modell. Die latenten Zusammenhänge zwischen den drei Konstrukten variierten zwischen $\Psi = .39$ und $\Psi = .76$.

Kriteriale Validität. Der Unterrichtsforschung zufolge wird die Umsetzung von Wissen in Handeln durch situationsbezogene Wahrnehmungs- und Planungsfähigkeiten vermittelt. Das mathematische und mathematikdidaktische Wissen sagten in einem videobasierten Assessment, das typische Situationen aus dem Kindergartenalltag zeigte, entsprechend signifikant die Situationswahrnehmung und Handlungsplanung angehender frühpädagogischer Fachkräfte voraus ($\beta_{\text{direkt}} = .45$ bzw. $\beta_{\text{indirekt}} = .43$ und $\beta_{\text{direkt}} = .60$ bzw. $\beta_{\text{indirekt}} = .58$; Dunekacke, Jenßen & Blömeke, 2015a, b). Ein weiteres Außenkriterium von mathematischem Wissen ist Mathematikangst, da sie signifikant mit geringeren Leistungen in diesem Bereich einhergeht (Ma, 1999). Latent-State-Trait-Modelle deuteten hypothesenkonform auf moderat negative Zusammenhänge zwischen den Subdimensionen der beiden Konstrukte hin ($[\Psi = -.24; \Psi = -.38]$; Jenßen, Dunekacke, Eid & Blömeke, 2015). Zudem bestätigten sie eine hinreichende Stabilität des mathematischen Wissens über verschiedene Messzeitpunkte, um von einem latenten Konstrukt sprechen zu können.

4 Forschungsfragen und Hypothesen

Ziel des vorliegenden Beitrags ist die systematische Überprüfung zentraler Gütekriterien für die Endversionen der so entwickelten Leistungstests mit einer Vier-Kohorten-Stichprobe angehender frühpädagogischer Fachkräfte im ersten bzw. letzten Jahr ihrer Ausbildung an Fachschulen bzw. Fachhochschulen – als den beiden Ausbildungswegen – aus allen 16 Bundesländern. Im Mittelpunkt stehen die Reliabilität und Messinvarianz sowie faktorielle, konvergente und divergente Validität einschließlich der Sensitivität der Tests mit Blick auf das Ziel, unterschiedlicher Ausprägungen zu diagnostizieren, die im Rahmen der frühpädagogischen Ausbildung in Deutschland erreicht werden können.

Reliabilität. Die erste Hypothese lautet, dass sich drei Skalen hinreichender Reliabilität bilden lassen, mit denen auf Gruppenebene strukturelle Aussagen zum und Beschreibungen von stärkeren und schwächeren Ausprägungen des mathematischen, mathematikdidaktischen und pädagogischen Wissens angehender frühpädagogischer Fachkräfte gemacht werden können (H1).

Messinvarianz. Erst beim Vorliegen konfiguraler Invarianz kann man davon ausgehen, dass ein Konstrukt in verschiedenen Gruppen eine vergleichbare inhaltliche Bedeutung hat

(H2a). Als Gruppierungsmerkmal werden die Ausbildungsinstitution, der Ausbildungsstand, das Geschlecht und die sprachliche Herkunft verwendet. Um auch metrische Invarianz annehmen zu können, müssen die Schwierigkeitsrelationen der Testitems zum unterliegenden Konstrukt in diesen Subgruppen vergleichbar ausfallen. Erst dann erfolgt die Messung auf einer einheitlichen Skala und Zusammenhänge zwischen dem latenten Konstrukt und anderen Merkmalen können valide über die Gruppen hinweg verglichen werden (H2b). Ist auch der Nullpunkt dieser Skala für alle Gruppen vergleichbar, liegen unverzerrte Mittelwerte und damit skalare Messinvarianz vor (H2c).

Faktorielle Validität. H3 lautet, dass sich das Ergebnis der Pilotstudie – Dreidimensionalität des Professionswissens von Frühpädagogen – replizieren lässt und mathematisches, mathematikdidaktisches und pädagogisches Wissen empirisch unterschieden werden kann.

Konvergente Validität. Eine statistische Prädiktion des Abschneidens in standardisierten Leistungstests sollte durch domänenspezifische Schulnoten möglich sein. Eine erste Hypothese lautet, dass die Abschlussnote in Mathematik das mathematische und das mathematikdidaktische Wissen signifikant vorhersagt (H4a, H4b). Eine zweite Hypothese bezieht sich auf die statistische Prädiktion der Pädagogik- und Mathematikdidaktik-Scores durch die Deutschnote (4c, H4d), weil sie trotz der Breite der Kompetenzen, die sie als Bestandteil des Mittleren Bildungsabschlusses bzw. der Allgemeinen Hochschulreife abdeckt, als Proxy für Lesekompetenz und Sprachverständnis – zwei von vier Kompetenzbereichen, die in den jeweiligen Standards von der Kultusministerkonferenz ausgewiesen werden (KMK, 2004, 2012) – verwendet werden kann. Die Forschung zu Textaufgaben verweist darauf, dass – neben domänenspezifischem Wissen – eine sichere Dekordierfähigkeit und ein gutes Sprachverständnis wichtig sind, um die Bedeutung einer Aufgabe identifizieren und ein mentales Modell der Anforderung bilden zu können (Lucangeli, Tressoldi & Cendron, 1998; Riley & Greeno, 1988). Da der Mathematikdidaktik- und der Pädagogiktest große Textanteile haben, kann plausibel angenommen werden, dass in diesen eine bessere Deutschnote von Vorteil ist.

Divergente Validität. Eine statistische Prädiktion des Mathematikwissens durch die Deutschnote ist aufgrund der geringeren Textanteile dagegen nicht zu erwarten (H5a). Ebenso wenig kann begründet erwartet werden, dass die Mathematiknote das Abschneiden im Pädagogiktest signifikant vorhersagt (H5b).

Sensitivität. Die drei Leistungstests sollen Unterschiede zwischen Fachschülern und Fachhochschülern sowie über die Ausbildungszeit hinweg abbilden. Angesichts der

günstigeren Eingangsvoraussetzungen zu Beginn der Ausbildung werden für alle drei Dimensionen signifikante Leistungsvorsprünge der Fachhochschüler gegenüber den Fachschülern vorausgesagt (H6a). Aufgrund der längeren Ausbildung an Fachhochschulen sollten die Unterschiede zwischen Fachhochschülern und Fachschülern am Ende der Ausbildung ebenfalls signifikant sein (H6b). Da die curricularen Analysen für die Mehrheit der Ausbildungsinstitutionen auf Lerngelegenheiten in Mathematikdidaktik und Pädagogik schließen lassen, wenn diese in ihrem Umfang auch stark variieren, lautet eine weitere Hypothese, dass am Ende der Ausbildung an beiden Institutionen ein signifikant höheres pädagogisches und – mit geringerer Effektstärke wegen geringerer Lerngelegenheiten – mathematikdidaktisches Wissen im Vergleich zum Anfang vorliegt (H6c). Angesichts generell weitgehend fehlender Lerngelegenheiten im Bereich Mathematik kann dagegen nicht erwartet werden, dass dieses Wissen signifikant ansteigt (H6d).

5 Methode

5.1 Stichprobe

2014 befanden sich in Deutschland rund 58.000 frühpädagogische Fachkräfte in der Ausbildung (Statistisches Bundesamt, 2014). Rund 90% davon wurden an einer Fachschule, 10% an einer Fachhochschule ausgebildet. In beiden Institutionen lag der Anteil weiblicher Auszubildender bei 82%. Die Stichprobe für die vorliegende Studie wurde über Anschreiben der Fachschulen und Fachhochschulen gewonnen. Wegen der Länderhoheit in Bezug auf die Ausbildungsbedingungen wurden Institutionen in allen Bundesländern kontaktiert. Um die Sensitivität der Tests für stärkere und schwächere Leistungen überprüfen zu können, wurden vier Gruppen in der Stichprobe berücksichtigt: Beginn und Ende der Fachschulausbildung sowie Beginn und Ende der Fachhochschulausbildung. Auf diese Weise können auf Basis einer Gelegenheitsstichprobe erste Hinweise auf Gruppenniveau-Normen geliefert werden (Kolen, 2006). In Bezug auf die Fachhochschüler fand ein Oversampling statt, um eine hinreichende Stichprobengröße zu gewährleisten.

Die Stichprobe umfasste 1851 Personen aus 86 Lehrveranstaltungen von 44 Ausbildungsinstitutionen. Darunter befanden sich 31 der 516 Fachschulen mit 67 Klassen und 1368 angehenden Fachkräften ($M_{Klasse}=20$, $SD=7.9$, $min-max=6-46$) sowie 13 der 80 Fachhochschulen mit 19 Lehrveranstaltungen und 483 Fachkräften ($M_{Lehrv}=25$, $SD=18.8$, $min-max=6-82$). Die demographischen Merkmale der Gruppen spiegeln – soweit bekannt

bzw. plausibel zu begründen – die Merkmale der Zielpopulationen wider (siehe Tabelle 1). In die frühpädagogische Ausbildung kann man nach dem mittleren Bildungsabschluss und anschließender zweijähriger Berufsausbildung oder nach dem Abitur eintreten. Dies reflektierten die Altersminima von 17 bzw. 18 Jahren. Die Ausbildung wird als Option aber auch in späteren Lebensabschnitten gewählt, was die Altersmaxima widerspiegeln. Am Ende der Ausbildung waren die Studienteilnehmer zwei (Fachschule) bzw. vier (Fachhochschule) Jahre älter als zu Beginn. Frauen stellten wie in der Zielpopulation jeweils mehr als 80% der Gruppen.

- hier Tabelle 1 einfügen -

Die Haushaltssprache war jeweils überwiegend deutsch. Groß waren die Unterschiede in der sozialen Herkunft und den kognitiven Eingangsvoraussetzungen. Das kulturelle Kapital der Fachschüler – angezeigt in Form der Anzahl der Bücher im elterlichen Haushalt und des höchsten Bildungsabschlusses der Mutter – war deutlich geringer als das der Fachhochschüler. Die Fachschüler traten mit einem niedrigeren Schulabschluss und weniger Schuljahren mit Mathematikunterricht in die Ausbildung ein als die Fachhochschüler.

5.2 Erhebungsinstrumente

Die Dauer der Testsitzung betrug 90 Minuten, wobei 25 Minuten für die Instruktion, Fragen zur Person sowie zu Überzeugungen und Lerngelegenheiten und 60 Minuten für die Testaufgaben (Pädagogik, Mathematikdidaktik, Mathematik) vorgesehen waren.

Der Pädagogik-Test besteht aus 18 Items, die unter empirischen – anhand der *MNSQ*-Fitwerte und der Testinformationskurve – sowie konzeptionellen Kriterien – Abdeckung aller Dimensionen, Subdimensionen und Deskriptoren – ausgewählt wurden. Drei Items verlangen eine offene Antwort (*open response*, *OR*), 15 Items wurden im Multiple-Choice- (*MC*-) oder Zuordnungs-Format gestellt. Der Mathematikdidaktik-Test besteht aus 28 Items, die anhand derselben empirischen und konzeptionellen Kriterien selektiert wurden. Fünf Items wurden im *OR*-, 23 im *MC*- oder Zuordnungs-Format gestellt. Der Mathematik-Test besteht aus 24 Items; 14 Items weisen ein *OR*-Format auf (einschl. Zeichnungen sowie Ergänzungen von Tabellen und Formeln) neben 10 *MC*-Items. In Anhang 2 findet sich jeweils ein Beispiel-Item für die drei Tests.

Zur Überprüfung konvergenter und divergenter Validität wurde die letzte Schulnote in Deutsch und in Mathematik erfragt, die untereinander nur schwach korrelieren ($r = .14$).

5.3 Erhebungsdesign und Skalierung

Die Testung wurde im Zuge von Lehrveranstaltungen durch geschulte Testleiter durchgeführt. Die Items wurden im Papier-und-Bleistift-Format in sechs Testheften administriert, die ein Multi-Matrix-Design aufwiesen, sodass alle Teilnehmenden nur etwa zwei Drittel aller Items bearbeiten mussten. Die Items jeder Wissensdimension waren hierfür inhalts- und schwierigkeitsbezogen ausgewogen auf fünf Blöcke verteilt worden, die so über die Testhefte rotierten, dass jede Dimension zwei Mal am Anfang, in der Mitte und am Ende der Testung bearbeitet wurde. Für die Kodierung der offenen Items wurde ein weiter verfeinertes Kodierhandbuch herangezogen. Die Doppelkodierung von 20% der Tests ergab eine sehr zufriedenstellende Inter-Rater-Reliabilität ($M_{Kappa} = .76$, $min-max = .64-.92$; $M_{Yules} = .97$, $min-max = .92-1.00$).

Die Reliabilitäten der Pilotstudien hatten darauf hingedeutet, dass eine Skalierung der Daten mit dem Raschmodell zu restriktiv sein könnte. Angesichts der großen Stichprobe wurde für die vorliegende Studie daher das zwei-parametrische logistische (2PL-)Modell herangezogen, in dem neben der Schwierigkeit auch die Trennschärfe der Items variieren kann. Dadurch können mehr Items im Test verbleiben und höhere Reliabilitäten erzielt werden (Andrich, 2004). In die Skalierung gingen alle vier Gruppen (Anfang bzw. Ende Fachschule, Anfang bzw. Ende Fachhochschule) gleichgewichtet ein, um zu vermeiden, dass eine der Gruppen die Ergebnisse dominierte. Zusätzlich wurde für jede Gruppe einzeln geprüft, ob die Skalen angemessene psychometrische Eigenschaften aufwiesen.

Fehlende Werte entstanden designbedingt, weil nicht allen Fachkräften alle Items vorgelegt wurden. Für diese Werte kann begründet angenommen werden (*missing completely at random*), dass sich ihr Fehlen nicht verzerrend auf die Skalierungsergebnisse auswirkt. Weitere fehlende Werte entstanden durch Nicht-Bearbeitung vorgelegter Items. Ihr Anteil lag zwischen unter einem und gut vier Prozent pro Item und sie wurden im Zuge der Kalibrierung und der Schätzung der Personenparameter als „fehlend“ definiert sowie modellbasiert über die Full-Information-Maximum-Likelihood-Methode iterativ imputiert. Dieses Vorgehen nützt alle verfügbaren Informationen aus und ist am wenigsten anfällig für Verzerrungen (Lüdtke, Robitzsch, Köller, Trautwein, 2008). Die resultierenden Personenwerte wurden skalenweise linear auf einen Mittelwert von 50 und eine

Standardabweichung von 10 normiert. In Ergänzung werden als unstandardisierte Werte pro Gruppe das 25., 50. (Median) und 75. Perzentil der Lösungshäufigkeiten auf den bearbeiteten Items angegeben.

5.4 Datenanalysen

Alle Validitäts- und Invarianzprüfungen erfolgten in Form von konfirmatorischen Faktorenanalysen im Mehrebenen-Mehrgruppen-Design (*MG-CFA*; Jöreskog, 1971). Die angehenden frühpädagogischen Fachkräfte stellten die erste und die 86 Lehrveranstaltungen bzw. Klassen die zweite Ebene dar. Studierende im ersten bzw. letzten Jahr an Fachschulen bzw. Fachhochschulen stellten die vier Gruppen dar, für die die Modelle parallel mit Hilfe der Software Mplus 7.3 geschätzt wurden (*twolevel CFA mixture modelling* unter Nutzung der *knownclass*- und *cluster*-Optionen; Muthén & Muthén, 2012). Zur Prüfung der konvergenten und divergenten Validität wurden die Schulnoten in diese Modelle einzeln als Kovariate eingeführt. Alle Modelle wurden mit einem Maximum-Likelihood-Schätzer und robusten Standardfehlern geschätzt; als Algorithmus diente ein Verfahren mit numerischer Integration und zehn Integrationspunkten pro Dimension.

Die Prüfung der Messinvarianz erfolgte schrittweise (Vandenberg & Lance, 2000). Bei der Prüfung konfiguraler Invarianz wurden Intercepts, Faktorladungen und Residualvarianzen frei geschätzt und die Mittelwerte aller Konstrukte in allen Subgruppen auf Null gesetzt. Zur Prüfung metrischer Invarianz wurden die Faktorladungen zwischen den Gruppen gleichgesetzt, bevor schließlich die Intercepts gleichgesetzt wurden, was als skalare Messinvarianz bezeichnet wird. Eine Prüfung der Ergebnisse der Sensitivitätsanalysen erfolgte über die Herstellung näherungsweiser Messinvarianz mithilfe der Alignment-Methode, die ähnlich einer Faktorenanalyse Abweichungen iterativ reduziert, sodass anschließend eine neue Schätzung und Prüfung der Mittelwertunterschiede erfolgen kann (Asparouhov & Muthén, 2014). Simulationsstudien zeigen, dass diese Methode Mittelwertunterschiede sehr verlässlich abbildet (Muthén & Asparouhov, 2013).

Da alle Analysen im Viergruppen-Design vorgenommen wurden, wurde als Signifikanzniveau eine Irrtumswahrscheinlichkeit von 5% festgelegt. Die Bedeutsamkeit von Unterschieden bei Mittelwertvergleichen wurde anhand von Cohens d (1988) eingeschätzt, wonach Differenzen ab $d = 0.2$ als kleine, ab $d = 0.5$ als mittlere und ab $d = 0.8$ als große Effekte bezeichnet werden.

Die Intraklassenkorrelation (ICC) lag, geschätzt nach Eid, Gollwitzer und Schmitt (2010), innerhalb der vier Gruppen für pädagogisches Wissen zwischen .06 und .15, für mathematisches Wissen zwischen .00 und .13 sowie für mathematikdidaktisches Wissen mit einer Ausnahme zwischen .06 und .11. Der Design-Effekt lag entsprechend zwischen 0.5 und 2.6, 1.7 und 3.8 bzw. 1.7 und 2.9. Eine Ausnahme stellte das mathematikdidaktische Wissens zu Beginn der Ausbildung an Fachhochschulen dar. Hier betrug die ICC .29 und der Designeffekt 8.1. Signifikanzprüfungen auf Individualebene wurden zur Kontrolle des Alpha-Fehler-Risikos daher auf Clusterebene repliziert.

6 Ergebnisse

6.1 Reliabilität (H1)

Die Reliabilitäten der drei Tests waren wie erwartet zufriedenstellend bis gut. Die nach Raykov, Dimitrov und Asparouhov (2010) geschätzte Reliabilität von 2PL-Modellen lag für den Mathematiktest bei $P\gamma = .88$ und für den mathematikdidaktischen Test bei $P\gamma = .87$. Lediglich die Reliabilität des Pädagogiktests fiel mit $P\gamma = .68$ geringer aus.

6.2 Messinvarianz (H2)

Sowohl für die vier Ausbildungsgruppen Anfang bzw. Ende an Fachschulen bzw. Fachhochschulen als auch für weibliche und männliche Studierenden sowie für Studierende mit Deutsch als Haushaltssprache und jene, wo dies nicht immer der Fall ist, lag wie erwartet konfigurale Invarianz vor (H2a; siehe Tabelle 2). Die Anpassung an die Daten verschlechterte sich auch im metrischen Modell nicht signifikant (H2b). Damit können die Ergebnisse von Zusammenhangsanalysen über die Gruppen hinweg verglichen werden. Skalare Invarianz konnte entgegen H2c dagegen nicht bestätigt werden. Die Nullpunkte der Skalen unterschieden sich zwischen den Gruppen.

- Tabelle 2 hier einfügen -

6.3 Faktorielle Validität (H3)

Der Modellvergleich machte deutlich, dass das dreidimensionale Modell wie erwartet besser zu den Daten passte als das eindimensionale (siehe Tabelle 3). Die Spannweite der standardisierten Faktorladungen lag zwischen .33 und .74 ($Md = .51$), .15 und .55 ($Md = .32$) bzw. .12 und .47 ($Md = .35$) für das mathematische, das mathematikdidaktische bzw. das pädagogische Wissen, während sie für das eindimensionale Modell zwischen .08 und .68 lag ($Md = .35$). Wie bei Lehrkräften war die latente Korrelation am niedrigsten, wenn pädagogisches und mathematisches Wissen betrachtet wurde ($\Psi = .62$). Ebenfalls wie bei Lehrkräften bestand ein stärkerer Zusammenhang, wenn mathematisches und mathematikdidaktisches Wissen betrachtet wurde ($\Psi = .73$). Anders als in den Lehrerstudien bestand bei angehenden frühpädagogischen Fachkräften der engste Zusammenhang aber zwischen dem mathematikdidaktischen und dem pädagogischen Wissen, die mit $\Psi = .92$ empirisch kaum zu trennen waren.

- hier Tabelle 3 einfügen -

Ergänzend wurde daher ein zweidimensionales Modell mit Mathematik als einer und Mathematikdidaktik sowie Pädagogik als zweiter Wissensdimension geschätzt. Die Fitwerte dieses Modells fielen nicht schlechter aus als die des dreidimensionalen. Auch blieb die Zahl der Items mit nicht-signifikanter Varianzaufklärung bzw. nicht-signifikanten Faktorladungen gleich. Die Spannweite der standardisierten Faktorladungen war für das mathematische Wissen ähnlich wie im dreidimensionalen Modell ($\lambda = .33-.74$; $Md = .51$) und für den zweiten Faktor ähnlich wie für die einzelnen Faktoren im dreidimensionalen Modell ($\lambda = .11-.54$; $Md = .33$). Ausweislich der latenten Korrelation ($\Psi = .68$) unterschieden sich die Ausprägungen zwischen den beiden Dimensionen deutlich.

6.4 Konvergente und divergente Validität (H4, H5)

Wie erwartet wurden das mathematische und das mathematikdidaktische Wissen signifikant durch die Mathematiknote (H4a und H4b: $\beta = -.21$ bzw. $-.11$) sowie das mathematikdidaktische und das pädagogische Wissen signifikant durch die Deutschnote vorhergesagt (H4c und H4d: $\beta = -.24$ bzw. $-.20$). Die Daten lieferten damit einen ersten Beleg für die konvergente Validität der Testbatterie. In Entsprechung zu H5b wurde das Pädagogikwissen nicht durch die Mathematiknote vorhergesagt ($\beta = -.03$), sodass für diesen Test auch divergente Validität festgehalten werden kann. Entgegen der in H5a formulierten

Annahme wurde das Mathematikwissen dagegen signifikant von der Deutschnote vorhergesagt, wenn auch mit geringer Effektstärke ($\beta = -.08$).

6.5 Sensitivität (H6)

Wie erwartet fanden sich ausweislich der Skalenmittelwerte im pädagogischen und mathematikdidaktischen Wissen signifikante Unterschiede in den Testleistungen zwischen Fachschulen und Fachhochschulen zugunsten Letzterer (siehe Tabelle 4), und zwar sowohl am Anfang (H6a) als auch am Ende der Ausbildung (H6b). Die Effekte fielen am Ende der Ausbildung mit $d = 0.74$ bzw. $d = 0.88$ größer aus als am Anfang ($d = 0.31$ bzw. $d = 0.50$). Entsprechend H6c wiesen die beiden Abschlussgruppen umfangreicheres pädagogisches und mathematikdidaktisches Wissen auf als die Anfängergruppen. Die Unterschiede waren an Fachhochschulen mit Werten um $d = 0.62$ größer als an Fachschulen ($d = 0.21$ bzw. $d = 0.32$). In Bezug auf das mathematische Wissen bestand zu Beginn und am Ende der Ausbildung ein großer Unterschied zwischen den Leistungen der Fachschüler und denen der Fachhochschüler (um $d = 0.73$). Im Vergleich der Anfangs- und Endkohorten ließ sich dagegen in beiden Ausbildungen wie erwartet kein Unterschied feststellen (H6d).

Kriterial gesehen manifestierten sich diese Unterschiede vor allem im mittleren Leistungsbereich. Die schwierigsten Items ($WLE > 1.5$) wurden in allen vier Gruppen nur von 0 bis 5% richtig gelöst. Diese Items sind durch eine konsequente Fachsprachlichkeit gekennzeichnet und fachliche Konzepte werden auf einem abstrakten Niveau beschrieben. Wie im mittleren Schwierigkeitsbereich müssen eigenständig Konzepte auf gegebene Situationen angewendet werden bzw. müssen diese anhand von Konzepten interpretiert werden. Die kognitive Komplexität ist aufgrund von Mehrschrittigkeit relativ hoch. Die leichtesten Items ($WLE < -2.0$) sind in allen drei Dimensionen durch offensichtliche, alltagsintegrierte Aktivitäten gekennzeichnet und es werden keine Fachbegriffe verwendet. Diese Anforderungen sind leicht vorstellbar und können prinzipiell von „informierten Laien“ gelöst werden. In allen Gruppen wurden sie von 98 bis 100% der angehenden Fachkräfte richtig gelöst.

Bei einem WLE-Wert um 0 variierte der Anteil der richtigen Lösungen zwischen 15 und 89%. In diesem mittleren Leistungsbereich müssen basale Fachbegriffe und Konzepte erkannt und angewendet werden. Auch muss professionelles Wissen zum Beispiel über die Entwicklung mathematischer Fähigkeiten präsent sein und angewendet werden. Generell steht der Anwendungsbezug im Vordergrund, indem – wie bei schwierigen Items – Konzepte

auf Situationen angewendet bzw. diese anhand von Konzepten interpretiert werden müssen. Allerdings ist die Problemrepräsentation deutlich einfacher gestaltet. Die kognitive Komplexität ist in der Regel auf einen Schritt begrenzt. Je nach Dimension lösten nur gut 20% bis zu 65% der Fachschüler diese Items am Ende der Ausbildung ($Md = 43$), während die Lösungshäufigkeit bei Fachhochschülern in keiner Dimension unter 50% lag und bis zu 90% reichte ($Md = 74\%$).

- hier Tabelle 4 einfügen -

Mithilfe der Alignment-Methode konnte mit Ausnahme von zwei Items für alle Items der drei Tests annäherungsweise Messinvarianz erreicht werden. Die erneut erfolgte Schätzung der Mittelwerte stützte die in Tabelle 4 dargestellten Rangfolgen. In Bezug auf das mathematische Wissen wiesen die z -Tests dieselben Unterschiede als signifikant bzw. nicht-signifikant aus. Im Bereich des mathematikdidaktischen Wissen fielen die beiden Extremwerte ebenfalls signifikant unterschiedlich aus, während im Bereich des pädagogischen Wissens keiner der Unterschiede signifikant wurde.

Ausweislich der Lösungshäufigkeiten des 50. Perzentils scheint der Mathematikdidaktiktest im Vergleich zu den beiden anderen im Mittel leichter gewesen zu sein (siehe auch Tabelle 4), da im keine der Gruppen weniger als 58% und die stärkste Gruppe 74% dieser Items erfolgreich gelöst hat. Die mittleren Lösungshäufigkeiten des Pädagogik- und des Mathematiktests lagen dagegen zwischen 42 und 56% bzw. 46 und 62%. Die Variabilität der Lösungshäufigkeiten war ausweislich des Interquartilabstands für alle Gruppen und Tests mit einer Ausnahmen ähnlich. Das 75. Perzentil lag in der Regel um 20 Prozentpunkte über dem 25. Perzentil. Eine Ausnahme stellte der Mathematiktest in der Gruppe der Fachschüler dar, wo der Abstand bis zu 28% betrug.

7 Diskussion

Ziel des vorliegenden Beitrags war die Überprüfung zentraler Gütekriterien einer neu entwickelten Testbatterie, die auf Gruppenebene strukturelle Aussagen und Diagnosen von unterschiedlichen Ausprägungen des pädagogischen, mathematikdidaktischen und mathematischen Wissens angehender frühpädagogischer Fachkräfte und dessen Relation zu Lerngelegenheiten erlaubt. Hier bestand eine empfindliche Forschungslücke. Die Ergebnisse liefern ermutigende Belege, dass die Testscores das Wissen für diesen Zweck weitgehend

reliabel, faktoriell, konvergent und divergent valide sowie hinreichend messinvariant und sensitiv in Bezug auf die Rangfolge vorhergesagter Leistungsunterschiede abbilden. Damit können die Wissensdimensionen in ihren Kernbestandteilen erstmals standardisiert und nicht lediglich anhand distaler und wenig reliabler Indikatoren in Form von Abschlüssen oder Selbstberichten erfasst und Effekte der Ausbildung oder individueller Voraussetzungen untersucht werden.

Dieser globale Befund muss allerdings in Teilen eingeschränkt werden. So handelte es sich zwar um eine große und über alle 16 Bundesländer verteilte Stichprobe, die zudem zwei unabhängig davon gezogenen Pilotstichproben folgte. Dennoch haben alle drei Stichproben nur den Charakter einer Gelegenheitsstichprobe, womit Grenzen der Verallgemeinerbarkeit verbunden sind. Mit Blick auf die Sensitivitätsanalysen konnte zwar für fast alle Items annäherungsweise Messinvarianz erreicht werden und stützte die Replikation der Mittelwertschätzungen auch alle Rangfolgen. Allerdings konnten nicht mehr alle Unterschiede als signifikant abgesichert werden. Zudem lässt die Höhe der Faktorladungen in Bezug auf die Pädagogik- und Mathematikdidaktiktests zu wünschen übrig, was vermutlich auf die Heterogenität der zugrundeliegenden Konstrukte und die Variabilität der gebotenen Lerngelegenheiten zurückzuführen ist. Und schließlich konnte divergente Validität nicht in Bezug auf die Relation von Mathematikwissen und Deutschnote bestätigt werden. Der Effekt war zwar gering, aber signifikant. Dekordierfähigkeit und Sprachverständnis scheinen als domänenübergreifende Fähigkeit die Güte der Bearbeitung aller Tests zu beeinflussen (Riley & Greeno, 1988). Weitere Validierungsstudien zum Beispiel mit der Note im Fach Pädagogik sollten sich daher anschließen.

In faktorieller Hinsicht stützten die Daten im ersten Schritt die aus der Lehrerforschung übernommene dreidimensionale Wissensstruktur und replizierten das Ergebnis unserer Pilotstudie (Jenßen et al., im Druck). Danach wäre nicht nur eine konzeptionelle, sondern auch eine empirische Differenzierung zwischen mathematischem, mathematikdidaktischem und pädagogischem Wissen bei angehenden Frühpädagogen sinnvoll. Allerdings war das mathematikdidaktische sehr stark mit pädagogischem Wissen assoziiert, sodass sich bei empirischen Analysen die Verwendung des ergänzend untersuchten zweidimensionalen Modells anbietet. Dieses Ergebnis spiegelt vermutlich das Spezifikum der Elementarpädagogik, dass die Arbeit mit jüngeren Kindern stärker durch pädagogische Aspekte beeinflusst wird als durch fachliche (Anders, 2012), obwohl die Förderung domänenspezifischer Fähigkeiten bei Kindern ein wichtiges Bildungsziel aller Bundesländer ist. Da entwicklungspsychologisch gesehen das Spiel im Elementarbereich das zentrale

Vermittlungsmedium darstellt, könnte ein entsprechend differenziertes pädagogisches Wissen so stark mit mathematikbezogenen Anwendungen verbunden sein, dass sich eine mathematikdidaktische Dimension – anders als im Schulbereich – davon nicht unterscheiden lässt. Die Natur der Mathematikdidaktik speist sich in der fröhpädagogischen im Vergleich zur schulischen Mathematikdidaktik entsprechend stärker aus elementarpädagogischen Bezügen als aus mathematischen (Gasteiger, 2010).

Insgesamt fallen alle latenten Korrelationen zwischen den Wissensdimensionen relativ hoch aus. Erkenntnissen der Expertiseforschung zufolge könnte dies ein Indikator eines eher niedrigen Wissensniveaus sein (Livingston & Borko, 1998). Damit würde indirekt auf die in der Tat geringen Lerngelegenheiten im Bereich der Mathematik und der Mathematikdidaktik hingedeutet. Eine alternative Annahme könnte lauten, dass generelle kognitive Fähigkeiten, insbesondere in Form kristalliner Intelligenz diesen Zusammenhang verursachen (Horn & Cattel, 1966). Eine Zusatzstudie wurde zwischenzeitlich mit den vorliegenden Instrumenten für das mathematische Wissen durchgeführt, die in diese Richtung weist. Die Frage stellt sich angesichts der Variabilität des realisierten Curriculums an den verschiedenen Ausbildungsinstitutionen (Dunekacke et al., 2013) und dem vergleichsweise geringen Wissensunterschied zwischen dem Beginn und dem Ende der Ausbildung aber auch für das pädagogische Wissen. Eine wichtige Aufgabe zukünftiger Forschung besteht daher in der weiteren Aufklärung dieser Ursache-Wirkungszusammenhänge.

Auch wenn die Validierung der Testinstrumente Hauptanliegen der vorliegenden Studie war, liefert sie erste Befunde zum Wissenstand in den verschiedenen Gruppen. Besorgniserregend ist der mittlere Wissenstand der Fachschüler im Vergleich zu den Fachhochschülern am Ende der Ausbildung. Die Fachschüler haben substanzielle Problem mit Items, die basale Fachsprachlichkeit oder Bezüge auf Theorien erfordern, und dies in pädagogischer *und* mathematikbezogener Hinsicht. Laut den Ausbildungsordnungen gehört das jeweilige Wissen aber zum Kerncurriculum der Fachkräfte. Hier deutet sich an, dass die Fachschulen aufgrund geringerer Eingangsvoraussetzungen und weniger Lernangeboten als an Fachhochschulen möglicherweise Schwierigkeiten haben, ihre Aufgaben umfänglich zu erfüllen. Mit Blick auf die Bildungsqualität im Kindergarten würde dies ein Problem darstellen. Die Variabilität der Leistungen ist dabei besonders groß im Bereich des Mathematikwissens. Die vorliegenden drei Tests stellen zur vertieften Untersuchung dieser Problematik erstmals geprüfte Instrumente zur Verfügung.

Literaturverzeichnis

- Allmendinger, H. (2011). Elementarmathematik vom höheren Standpunkt: Eine Begriffsanalyse in Abgrenzung zu Felix Klein. In Haug, R. & Hölzapfel, L. (Hrsg.), *Beiträge zum Mathematikunterricht 2011* (S. 51–54). Münster: WTM-Verlag.
- Anders, Y. (2012). *Modelle professioneller Kompetenzen für frühpädagogische Fachkräfte: Aktueller Stand und ihr Bezug zur Professionalisierung*. München: Knoblingesign.
- Anderson, L. W. & Krathwohl, D. R. (Hrsg.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care*, 42, 1–16.
- Asparouhov, T. & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 495–508.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47, 371–406.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich*. Münster: Waxmann.
- BMFSFJ (Bundesministerium für Familien, Senioren, Frauen und Jugend) (2010). *Männliche Fachkräfte in Kindertagesstätten: Eine Studie zur Situation von Männern in Kindertagesstätten und in der Ausbildung zum Erzieher*. Berlin: Sinus Sociovision.
- Bouley, F., Berger, S., Fritsch, S., Wuttke, E., Seifried, J., Schnick-Vollmer, K. & Schmitz, B. (2015). Der Einfluss von universitären und außeruniversitären Lerngelegenheiten auf das Fachwissen und fachdidaktische Wissen von angehenden Lehrkräften an kaufmännisch-berufsbildenden Schulen. *Zeitschrift für Pädagogik*. 61. Beiheft, 100–115.
- Bremerich-Vos, A. & Dämmer, J. (2013). Professionelles Wissen im Studienverlauf: Lehramt Deutsch. In Blömeke, S. (Hrsg.), *Professionelle Kompetenzen im Studienverlauf: Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerbildung aus TEDS-LT* (S. 47–76). Münster: Waxmann.
- Buchholtz, N., Kaiser, G. & Blömeke, S. (2014). Die Erhebung mathematikdidaktischen Wissens: Konzeptualisierung einer komplexen Domäne. *Journal für Mathematikdidaktik*, 35, 101–128.
- Cadinu, M., Maass, A., Rosabianca, A. & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, 16, 572–578.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Lawrence Erlbaum Associates.

- Common Core State Standards Initiative (2014). *Common Core State Standards for Mathematics*. Abgerufen von http://www.corestandards.org/wp-content/uploads/Math_Standards.pdf.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P. ... Japel, C. (2006). School readiness and later achievement. *Developmental Psychology*, 43, 1428–1462.
- Dunekacke, S., Jenßen, L., Baack, W., Tengler, M., Wedekind, H., Grassmann, M. & Blömeke, S. (2013). Was zeichnet eine kompetente pädagogische Fachkraft im Bereich Mathematik aus? Modellierung professioneller Kompetenz für den Elementarbereich. In Greefrath, G., Käpnick, F. & Stein, M. (Hrsg.), *Beiträge zum Mathematikunterricht 2013* (S. 280–283). WTM: Münster.
- Dunekacke, S., Jenßen, L. & Blömeke, S. (2015a). Effects of mathematics content knowledge on pre-school teachers' performance: A video-based assessment of perception and planning abilities in informal learning situations. *International Journal of Science and Mathematics Education*, 13, 267–286.
- Dunekacke, S., Jenßen, L. & Blömeke, S. (2015b). Mathematikdidaktische Kompetenz von Erzieherinnen und Erziehern: Validierung des KomMa-Leistungstests durch die videogestützte Erhebung von Performanz. *Zeitschrift für Pädagogik*, 61. Beiheft, 80–99.
- Embretson, S. E. & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science Quarterly*, 50, 328–344.
- Gasteiger, H. (2010). *Elementare mathematische Bildung im Alltag der Kindertagesstätte: Grundlegung und Evaluation eines kompetenzorientierten Förderansatzes*. Münster: Waxmann.
- Grüßing, M. & Peter-Koop, A. (Hrsg.). *Die Entwicklung mathematischen Denkens in Kindergarten und Grundschule: Beobachten – Fördern – Dokumentieren* (3. Aufl., S. 5–9). Offenburg: Mildenberger.
- Gut, J., Reimann, G. & Grob, A. (2012). Kognitive, sprachliche, mathematische und sozial-emotionale Kompetenzen als Prädiktoren späterer schulischer Leistungen. *Zeitschrift für Pädagogische Psychologie*, 26, 213–220.
- Harsch, C. & Hartig, J. (2011). Modellbasierte Definition von fremdsprachlichen Kompetenzniveaus am Beispiel der Bildungsstandards Englisch. *Zeitschrift für Interkulturelle Fremdsprachenforschung*, 6–17.
- Hasemann, K. (2010). Mathematische Einsichten von Kindern im Vorschulalter. In M. Grüßing & A. Peter-Koop (Hrsg.), *Die Entwicklung mathematischen Denkens in Kindergarten und Grundschule: Beobachten – Fördern – Dokumentieren* (3. Aufl., S. 67–79). Offenburg: Mildenberger.
- Hasselhorn, M. (2005). Lernen im Altersbereich zwischen 4 und 8 Jahren: Individuelle Voraussetzungen, Entwicklung, Diagnostik und Förderung. In T. Guldemann (Hrsg.), *Bildung 4-bis 8-jähriger Kinder* (S. 77–88). Münster: Waxmann.
- Hasselhorn, M. (2010). Möglichkeiten und Grenzen der Frühförderung aus entwicklungspsychologischer Sicht. *Zeitschrift für Pädagogik*, 56, 168–177.

- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253–270.
- Jenßen, L., Dunekacke, S. & Blömeke, S. (2015a). Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. *Zeitschrift für Pädagogik*; 61. Beiheft, 11–31.
- Jenßen, L., Dunekacke, S., Eid, M. & Blömeke, S. (2015b). The relationship of mathematical competence and mathematics anxiety: An application of latent state-trait theory. *Zeitschrift für Psychologie*, 223, 31–38.
- Jenßen, L., Dunekacke, S., Baack, W., Tengler, M., Koinzer, T., Schmude, C., ... Blömeke, S. (im Druck). KomMa: Kompetenzmodellierung und Kompetenzmessung bei frühpädagogischen Fachkräften im Bereich Mathematik. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttko (Hrsg.), *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte*.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.
- Klibanoff, R.S., Levine, S.C., Huttenlocher, J., Vasilyeva, M. & Hedges, L.V. (2006). Preschool children's mathematical knowledge: The effect of teacher "Math Talk". *Developmental Psychology*, 42, 56–69.
- Kultusministerkonferenz (KMK) (2004). *Bildungsstandards im Fach Mathematik für den Primarbereich: Beschluss vom 15.10.2004*. München: Wolters Kluwer.
- König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerbildung. *Zeitschrift für Erziehungswissenschaft*, 12, 499–527.
- Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Hrsg.), *Educational measurement* (4. Aufl., S. 155–186). Westport, CT: American Council on Education.
- Krajewski, K. & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction*, 19, 513–526.
- Kretschmann, R. (2004). Lesen, Schreiben, Rechnen – schon im Kindergarten? In I. Wehrmann (Hrsg.), *Kindergärten und ihre Zukunft* (S. 220–234). Weinheim: Beltz.
- Lee, J. (2010). Exploring kindergarten teachers' pedagogical content knowledge of mathematics. *International Journal of Early Childhood*, 47, 27–41.
- Lee, J., Meadows, M. & Lee, J. O. (2003). *What causes teachers to implement high quality mathematics education more frequently: Focusing on teachers' pedagogical content knowledge*. Washington, D.C: ERIC.
- Liegle, L. (2008). Erziehung als Aufforderung zur Bildung: Aufgaben der Fachkräfte in Tageseinrichtungen für Kinder in der Perspektive der frühpädagogischen Didaktik. In W. Thole,

- H.-G. Roßbach, M. Fölling-Albers & R. Tippelt (Hrsg.), *Bildung und Kindheit: Pädagogik der Frühen Kindheit in Wissenschaft und Lehre* (S. 85–113). Opladen: Budrich.
- Livingston, C. & Borko, H. (1998). Expert-novice differences in teaching: A cognitive analysis and implications for teacher education. *Journal of Teacher Education*, 40, 36–42.
- Lucangeli, D., Tressoldi, P. E. & Cendron, M. (1998). Cognitive and metacognitive abilities involved in the solution of mathematical word problems: Validation of a comprehensive model. *Contemporary Educational Psychology*, 23, 257–275.
- Lüdtke, O., Robitzsch, A., Trautwein, A. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58, 103–117.
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30, 520–540.
- Michel, M., Ofner, D. & Thoma, D. (2014). Early childhood educators' competences for supporting children's academic language skills: Empirical evidence from a German context. *Language Awareness*, 23, 138–156.
- Mischo, C., Wahl, S., Hendler, J. & Strohmmer, J. (2012). Warum in einer Kindertagesstätte arbeiten? *Zeitschrift für Pädagogische Psychologie*, 26, 167–181.
- Montie, J. E., Xiang, Z. & Schweinhart, L. J. (2007). *The role of preschool experience in children's development: Longitudinal findings from 10 countries*. Ypsilanti, MI: High/Scope Press.
- Muthén, B. & Asparouhov, T. (2013). *New methods for the study of measurement invariance with many groups: Technical report*. Abgerufen von <http://statmodel2.com/download/PolAn.pdf>.
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus user's guide* (6. Aufl.). Los Angeles, CA: Muthén & Muthén.
- Oerter, R. (2008). Kindheit, Spiel und kindliche Entwicklung. In Oerter, R. & Montada, L. (Hrsg.), *Entwicklungspsychologie* (6. Aufl., S. 236–249). Weinheim: Beltz.
- Ofner, D. & Thoma, D. (2014). Early childhood educators' abilities in planning language learning environments increase with pedagogical content knowledge relevant for language training. *European Journal of Applied Linguistics*, 2, 121–143.
- Rathgeb-Schnierer, E. (2012). Mathematische Bildung (S. 50–85). In Kucharz, D. (Hrsg.), *Elementarbildung*. Weinheim: Beltz.
- Raykov, T., Dimitrov, D. M. & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 17, 265–279.
- Retelsdorf, J. & Südkamp, A. (2012). Entwicklung professioneller Kompetenz von Lehrerinnen und Lehrern. *Zeitschrift für Pädagogische Psychologie*, 4, 227–231.
- Riese, J., Borowski, A., Fischer, H. E., Gramzow, Y., Kulgemeyer, Ch., Reinhold, P. ... Zander, S. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. *Zeitschrift für Pädagogik*. 61. Beiheft, 55–79.

- Riley, M. S. & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction*, 5, 49–101.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Huber.
- Schladitz, S., Groß Ophoff, J. & Wirtz, M. (2015). Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz. *Zeitschrift für Pädagogik*. 61. Beiheft, 167–184.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4–14.
- Spinath, B. (2012). Beiträge der Pädagogischen Psychologie zur Professionalisierung von Lehrerinnen und Lehrern: Diskussion zum Themenschwerpunkt. *Zeitschrift für Pädagogische Psychologie*, 26, 307–312.
- Statistisches Bundesamt (2014). *Bildung und Kultur: Berufliche Schulen Schuljahr 2013/2014* (=Fachserie 11; Reihe 2). Wiesbaden: Statistisches Bundesamt.
- Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–69.
- Weinert, F. E. (2001). Concept of competence: A conceptual classification (S. 45–66). In D. S. Rychen & L. Hersh Salganik (Hrsg.), *Defining and selecting key competencies*. Göttingen: Hogrefe.

Tabelle 1.

Deskriptive Merkmale der Stichprobe (N=1 851)

	Anfang Fachschule (n = 594)	Ende Fachschule (n = 774)	Anfang Fachhochschule (n = 287)	Ende Fachhochschule (n = 196)
Alter (Mittelwert, <i>min-max</i>)	22 (17-53)	24 (18-54)	22 (18-47)	26 (19-53)
Geschlecht (Anteil weiblich)	85%	83%	90%	90%
Haushaltssprache Deutsch (Anteil immer)	88%	89%	83%	86%
Bücher Eltern (Anteil >200)	23%	24%	41%	44%
Abschluss Mutter (Anteil Fach-/Abitur)	17%	16%	32%	27%
eigener Abschluss (Anteil Fach-/Abitur)	36%	44%	99%	99%
Schuljahre mit Mathematikunterricht (Anteil max. 10 Schuljahre)	47%	48%	2%	5%

Tabelle 2.

Prüfungen auf Messinvarianz über verschiedene Subgruppen hinweg

Modell	Anzahl Parameter			Log likelihood			Modellvergleich								
							Chi-Quadrat			df			p-Wert		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Konfigural	575	287	287	-41 894.0	-40 485.8	-40 528.0	---	---	---	---	---	---	---	---	---
Metrisch	374	220	220	-41 999.2	-40 514.9	-40 567.8	200.7	50.5	48.2	201	67	67	ns	ns	ns
Skalar	173	153	153	-42 216.5	-40 584.4	-40 617.1	381.2	138.1	97.9	201	67	67	<.05	<.05	<.05

Anmerkungen. df = Freiheitsgrade, ns = nicht signifikant, Subgruppen: Modell 1 = vier Gruppen (Anfang bzw. Ende der Ausbildung an Fachschulen bzw. Fachhochschulen), Modell 2 = zwei Gruppen (Männer, Frauen), Modell 3 = zwei Gruppen (Sprache immer Deutsch, nicht immer Deutsch), Modellvergleiche: metrisch gegen konfigural, skalar gegen metrisch.

Tabelle 3.

Vergleich ein-, zwei- und dreidimensionaler Modelle des Professionswissens frühpädagogischer Fachkräfte

Modell	# Par	<i>LL</i>	<i>SCF</i>	<i>AIC</i>	<i>BIC_{adj}</i>	ns Ladung	ns Var.- aufklärung
1-dim	143	-42535.4	1.92	85356.9	85692.4	1 Item	7 Items
3-dim	146	-42398.1	1.90	85088.3	85430.8	1 Item	7 Items
2-dim	144	-42403.3	1.90	85094.7	85432.6	1 Item	7 Items

Anmerkung. # Par. = Anzahl Parameter, LL = Loglikelihood, SCF = Scaling Correction Factor, AIC = Akaike's Information Criterion, *BIC_{adj}* = adjusted Bayesian Information Criterion.

Tabelle 4.

Gruppennormwerte zum pädagogischen (1), mathematikdidaktischen (2) und mathematischen (3) Wissen

		Fachschule			Fachhochschule			<i>t</i> -Werte			Cohens <i>d</i>		
		1	2	3	1	2	3	1	2	3	1	2	3
An-	<i>M</i> (<i>SD</i>)	48 (9.8)	47 (9.9)	48 (9.5)	51 (9.9)	52 (10.0)	55 (9.3)	4.6*	6.8*	10.4*	0.31	0.50	0.74
fang	<i>Q</i> _{.25} ; <i>Q</i> _{.50} ; <i>Q</i> _{.75}	.31; .42; .50	.47; .58 ;.69	.36; .46; .62	.36; .46; .57	.53; .65; .76	.50; .62; .71						
Ende	<i>M</i> (<i>SD</i>)	50 (9.5)	50 (9.2)	48 (9.6)	57 (9.2)	58 (8.7)	55 (9.5)	9.7*	11.1*	8.8*	0.74	0.88	0.73
	<i>Q</i> _{.25} ; <i>Q</i> _{.50} ; <i>Q</i> _{.75}	.36; .46; .57	.53; .63; .73	.36; .49; .64	.44; .56; .64	.67; .74; .87	.50; .62; .71						
Ver-	<i>t</i> -Werte	4.4*	5.7*	0.8	7.2*	6.9*	0.2						
gleich	Cohens <i>d</i>	0.21	0.32	0.00	0.62	0.63	0.00						

Anmerkung. *M* = Mittelwert und *SD* = Standardabweichung der Personenfähigkeitswerte aus der 2PL-IRT-Skalierung, die skalenweise auf einen Mittelwert von 50 und eine Standardabweichung von 10 transformiert wurden. *Q*_{.25}, *Q*_{.50} und *Q*_{.75} = mittlere Lösungshäufigkeiten des 25., 50. und 75. Perzentils. *T*-tests auf Mittelwertunterschiede erfolgten paarweise zunächst auf Individualebene ohne Berücksichtigung der genesteten Datenstruktur. Zur Kontrolle des Alpha-Fehler-Risikos wurden die *t*-Tests erfolgreich auf der Klassenebene repliziert. * $p < .05$.

Anhang 1.

Dimensionen, Subdimensionen und Deskriptoren des Professionswissens frühpädagogischer Fachkräfte (in Klammern Zahl der Testitems)

Pädagogisches Wissen (18)	Mathematikdidaktisches Wissen (28)	Mathematisches Wissen (24)
Bildungstheoretische Grundlagen (5) - Grundbegriffe und kindgemäße Bildungsziele auswählen (2) - Erziehungsstile und pädagogische Ansätze anwenden (2) - formelle und informelle Lerngelegenheiten unterscheiden (1)	Entwicklung und Diagnose mathematischer Fähigkeiten bei Kindern (17) - Psychologische Grundlagen der Entwicklung mathematischer Fähigkeiten kennen (2) - Entwicklungsstand im Bereich Zahlen, Mengen und Operationen anhand von Kinderaussagen diagnostizieren (5) - Entwicklungsstand im Bereich Form, Raum und Veränderung anhand von Kinderaussagen diagnostizieren (2) - Offene und standardisierte Diagnoseverfahren analysieren und bewerten (2) - Alltagssituationen im Bereich Zahlen, Mengen und Operationen identifizieren und begründen (3) - Alltagssituationen im Bereich Form, Raum und Veränderung identifizieren und begründen (1) - Alltagssituationen im Bereich Größen, Messen und Relationen identifizieren und begründen (2)	Zahlen, Mengen und Operationen (6) - Zahlbereiche kennen (1) - Grundoperationen ausführen (2) - Zählprinzipien anwenden (2) - Mengenverständnis zeigen (1) Form, Raum und Veränderung (6) - Formeln erkennen und anwenden (2) - Eigenschaften geometrischer Formen erkennen und begründen (2) - Räumliche Orientierung nachweisen (1) - Geometrische Körper konstruieren (1) Daten, Kombinatorik und Zufall (6) - Tabellen und Strichlisten erstellen und interpretieren (2) - Anzahl an Möglichkeiten ermitteln (1) - Wahrscheinlichkeiten bei Spielen einschätzen und begründen (3) Größen, Messen und Relationen (6) - Relationen von Geschwindigkeiten und Zeiten herstellen (2) - Verbale Aussagen mathematisieren und Mathematisches verbalisieren (2) - Muster erkennen und mathematisieren (2)
Psychologische Grundlagen (6) - Motivations- und Attributionsprozesse kennen (2) - Fachunabhängige Lern- und Entwicklungsprozesse diagnostizieren (2) - Strategien zur Verhaltensänderung und Angstbekämpfung entwickeln (2)	Gestaltung mathematischer Lernumgebungen (11) - Förderprinzipien kennen und anwenden (einschl. besonderer Angebote für Risikokinder) (3) - Spielerische Erfahrungen zu Zahlen, Mengen und Operationen ermöglichen (2) - Spielerische Erfahrungen zu Form, Raum und Veränderung ermöglichen (4) - Spielerische Erfahrungen zu Daten, Kombinatorik und Zufall ermöglichen (1) - Spielerische Erfahrungen zu Größen, Messen und Relationen ermöglichen (1)	
Didaktisch-methodische Grundlagen (7) - Kommunikations- und Kooperationsansätze kennen und anwenden (2) - Ansätze zur Förderung heterogener Kindergruppen kennen und anwenden (3) - Inklusionsprinzipien kennen und anwenden (2)		

Anhang 2.

Beispielitems zum mathematischen, mathematikdidaktischen und pädagogischen Wissen (das vollständige Instrument ist beim Autorenteam erhältlich)

Wissensdimension	Subdimension	Richtige Antwort
Mathematisches Fachwissen	Daten, Kombinatorik und Zufall	3

Chris hat einen blauen, einen grünen, einen roten und einen gelben Würfel. Chris will durch Übereinanderstapeln aller 4 Würfel einen Turm bauen.

Welcher Rechenausdruck gibt die Anzahl der Möglichkeiten verschiedener Türme an?

Kreuzen Sie bitte ein Kästchen an.

- ☐ 1 $4 + 4 + 4 + 4$
☐ 2 $4 \cdot 4 \cdot 4 \cdot 4$
☐ 3 $4 \cdot 3 \cdot 2$
☐ 4 $4 + 3 + 2$

Wissensdimension	Subdimension	Richtige Antwort
Mathematikdidaktisches Wissen	Gestaltung mathematischer Lernumgebungen	siehe Coding (Auszug)

Sie spielen mit drei Kindern ein Würfelspiel.

Begründen Sie bitte kurz, warum die folgenden mathematischen Inhalte damit gefördert werden.

a) Zahlen und Operationen (z.B. „Rechnen“)	
--	--

Code Antwort

- 1 *Richtige Antworten*
- 11 *Mengen (Mengenvergleich, Mengenverständnis):* Antworten, die sich auf den konkreten Umgang mit Mengen beziehen, zB „Die Würfel haben „Ziffern“ von 1-6, die Kinder lernen Mengenunterschiede der Punkte.“
- 12 *Simultanes erfassen:* Simultanerfassung als Teil der Zahlbegriffsentwicklung. Kann auch umschrieben werden z.B. mit dem Verb „erkennen“, „Sie müssen die Augenzahl auf dem Würfel erfassen.“
- 13 *Eins-zu-eins-Zuordnungen (durch Setzen der Spielfigur):* Alle Aussagen, bei denen es hauptsächlich darum geht, dass durch das Setzen der Spielfigur eine Zuordnung zu der gewürfelten Zahl hergestellt werden muss, zB „Die Kinder erkennen eine Zahl auf dem Würfel und setzen diese mit ihrer Figur auf dem Spielbrett um.“
- 5 *Falsche Antworten*
- 50 *Mathematisch unangemessene Antworten:* Würfelaugen sollten simultan erkannt bzw. gezählt werden, zB „Jeder Punkt steht für die 1 also kann ich 1+1+1 zum Beispiel rechnen.“
- 51 *Unvollständige Antworten:* Es ist nicht eindeutig ersichtlich, dass vom Würfeln mit zwei Würfeln ausgegangen wird, und ausschließlich oder hauptsächlich auf Zusammenzählen/Addieren argumentiert wird, zB „Augenzahlen zusammenzählen“.

- 52 *Dem Entwicklungsstand unangemessen, zB „Augenzahlen können addiert werden, bis jedes Kind 20 oder mehr Augenzahlen erreicht hat.“*
- 53 *Allgemein-pädagogische Aspekte, zB „Durch das Zählen und Lernen wieviel Augen der Würfel hat, kommen die Kinder schneller mit Spaß ans Rechnen.“*
Andere Antworten

Wissensdimension	Subdimension	Richtige Antwort
Pädagogisches Wissen	Psychologische Grundlagen	2

Welche der folgenden Aussagen trifft für Kinder im Alter von 4 bis 6 Jahren im Gegensatz zu Grundschulkindern zu?

Kreuzen Sie bitte ein Kästchen an.

- ☐ 1 Die Kinder sind motivierter für Lernprozesse.
- ☐ 2 Die Kinder lernen eher implizit.
- ☐ 3 Die Kinder lernen eher explizit.
- ☐ 4 Die Kinder lernen durch „inneres Nachsprechen“