

## **Von der Lehrerausbildung in den Beruf – Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht**

**Sigrid Blömeke · Johannes König · Andreas Busse · Ute Suhl ·  
Jessica Benthien · Martina Döhrmann · Gabriele Kaiser**

Eingegangen: 7. Mai 2013 / Angenommen: 11. Juli 2014 / Online publiziert: 27. August 2014  
© Springer Fachmedien Wiesbaden 2014

**Zusammenfassung** Der Berufseinstieg gilt als Schlüsselphase von Lehrerbiographien, doch es fehlen Erkenntnisse zur Entwicklung des am Ende der Ausbildung vorliegenden fachlichen und fachdidaktischen Wissens sowie zu seiner Bedeutung für situationsbezogene und verhaltensnahe Kognitionen. Auf der Basis von Daten aus der internationalen Vergleichsstudie TEDS-M 2008 und seinem 2012 in

---

Prof. Dr. S. Blömeke (✉) · Dr. U. Suhl  
Institut für Erziehungswissenschaften, Humboldt-Universität zu Berlin,  
Unter den Linden 6,  
10099 Berlin, Deutschland  
E-Mail: sigrid.bloemeke@staff.hu-berlin.de

Dr. U. Suhl  
E-Mail: ute.suhl@staff.hu-berlin.de

Prof. Dr. J. König  
Institut für Allgemeine Didaktik und Schulforschung, Universität zu Köln,  
Gronewaldstr. 2,  
50931 Köln, Deutschland  
E-Mail: johannes.koenig@uni-koeln.de

Dr. A. Busse · Prof. Dr. G. Kaiser  
Fakultät für Erziehungswissenschaft, Universität Hamburg,  
Binderstr. 34,  
20146 Hamburg, Deutschland  
E-Mail: andreas.busse@uni-hamburg.de

Prof. Dr. G. Kaiser  
E-Mail: gabriele.kaiser@uni-hamburg.de

J. Benthien · Prof. Dr. M. Döhrmann  
Institut für Didaktik der Mathematik und des Sachunterrichts, Universität Vechta,  
Driverstraße 22,  
49377 Vechta, Deutschland  
E-Mail: jessica.benthien@uni-vechta.de

Prof. Dr. M. Döhrman  
E-Mail: martina.doehrmann@uni-vechta.de

Deutschland durchgeführten Follow-Up wird in diesem Beitrag über einen 4-Jahres-Längsschnitt mit  $n=171$  Mathematiklehrkräften der Sekundarstufe I die Kompetenzentwicklung während des Übergangs von der Lehrerausbildung in den Beruf untersucht. Dabei werden mithilfe von drei Testformaten (Papier-und-Bleistift-, Video-Vignetten- und Geschwindigkeitstest) unterschiedliche Indikatoren herangezogen, die auf das Vorliegen von Lehrerexpertise schließen lassen. Neben der Analyse von Niveauveränderungen im mathematischen (MCK) und mathematikdidaktischen Wissen (MPCK) wird in Pfadmodellen geprüft, inwieweit dieses Wissen ein geeigneter Prädiktor für die Wahrnehmung und Interpretation von Unterrichtssituationen einschließlich der Reaktion darauf sowie das schnelle Erkennen von Schülerfehlern ist. Die Ergebnisse zeigen eher unbedeutende Niveauveränderungen im Gruppenmittel (MCK: leichte lineare Vergessenseffekte, MPCK: Konstanz), aber starke Veränderungen in der MPCK-Rangfolge (d.h., geringe Korrelation), was auf unterschiedliche Lerneffekte aus der Praxiserfahrung hindeutet. Bedeutsam sind zudem Unterschiede in der Vorhersagekraft: MCK ist wichtig für den Geschwindigkeits-Indikator von Lehrerexpertise, MCK *und* MPCK für die Wahrnehmung und Interpretation sowie die Generierung von Handlungsoptionen. Diese Erkenntnisse können zum einen helfen, Entscheidungen hinsichtlich des Unterstützungsbedarfs von Berufsanfängerinnen und -anfängern besser zu gestalten; zum anderen deuten sie auf die Notwendigkeit der frühen Integration von Praxiselementen in die Lehrerausbildung zur Förderung des fachdidaktischen Kompetenzaufbaus hin.

**Schlüsselwörter** Video-Vignette · Geschwindigkeitstest · Junglehrer · Lehrerbildung · Lehrerwissen · Mathematisches Wissen · Mathematikdidaktisches Wissen · Berufseinstieg · Lehrerberuf

### **The transition from teacher training to the teaching profession: subject-specific knowledge as a prerequisite for perception, interpretation and decision-making in the classroom**

**Abstract** The transition from teacher training into the teaching profession is seen to be key in the biography of a teacher. Yet there is a lack of enquiry into the development of the content knowledge (CK) and pedagogical content knowledge (PCK) held by teachers at the end of their education, and into the significance of this knowledge for situational and behavioural cognition. On the basis of data from the international comparative study TEDS-M 2008 and the Germany-based 2012 follow-up, this contribution investigates the knowledge and skills development of teachers during the transition from teacher training to the teaching profession in a four-year longitudinal study with  $n=171$  mathematics teachers. Three test formats (paper-and-pencil, video-vignettes and speed tests) are used to examine various indicators for teacher expertise. In addition to the analysis of changes in mathematical CK and PCK levels, path analysis is used to determine the extent to which this knowledge functions as a predictor for the perception and interpretation of classroom situations, the decision-making in these situations, and the quick identification of student mistakes. The results show only small changes in the group mean (for

MCK there is a small loss, whereas MPCK remains constant), but high fluctuation (i. e. weak correlation) in the MPCK ranking between the two measurement occasions, pointing to differential learning effects gained from teaching practice. There are also significant differences in predictability: MCK is important for the speed indicators of teacher expertise; MCK and MPCK for the perception and interpretation of situations, and for the generation of action strategies. These findings can on the one hand help to facilitate the evaluation of beginning teachers' need for support and, on the other hand, they indicate the necessity for the early integration of practice-based elements into teacher training programmes with the aim of developing PCK and related skills.

**Keywords** Beginning teachers · Pedagogical content knowledge · Mathematical competency · Speed test · Teacher education · Teacher expertise · Teaching profession · Video vignette

## 1 Einleitung

Vor allem unter Bezugnahme auf sozialisatorische Effekte (Fenstermacher 1994; Terhart 1998; Herrmann und Hertrampf 2000) gilt der Berufseinstieg von Lehrkräften als Schlüsselphase von Lehrerbiografien. Auf Selbstberichten beruhende Studien lassen annehmen, dass die eigenen Unterrichtsstunden von Junglehrkräften als *survival stage* (Fuller und Brown 1975) bzw. als „Handeln unter Druck“ (Wahl 1991) erlebt werden, in deren Folge die Berufsanfängerinnen und -anfänger weniger theorieorientierte und empirisch fundierte Unterrichtsansätze, sondern eher einen – möglicherweise noch aus der schulischen Sozialisation bekannten – wenig effektiven Unterrichtsstil implementieren (Müller-Fohrbrodt et al. 1978; Brouwer 2010).

Es fehlen allerdings testbasierte Studien, die unabhängig von Selbstwahrnehmungen fundiert über die Entwicklung der in der Ausbildung erworbenen professionellen Kompetenz von Lehrkräften in der beruflichen Eingangsphase informieren. Zwar kann auf der Basis von phänomenologischen Studien der Expertiseforschung geschlussfolgert werden, dass in den ersten Berufsjahren von Lehrkräften aufgrund der zunehmenden Unterrichtserfahrung Umstrukturierungsprozesse in Bezug auf das in der Ausbildung erworbene Wissen stattfinden (Berliner 1986; Borko und Livingston 1989; Chi 2011). Inwieweit sich das in der Lehrerbildung erworbene professionelle Wissen mit dem Berufseinstieg verändert, inwieweit mittelfristig ein Expertiseprofil entsteht, sind jedoch Fragen, die bislang von der Forschung kaum bearbeitet worden sind. Erkenntnisse hierzu könnten langfristig helfen, pädagogische und personenbezogene Entscheidungen (z. B. hinsichtlich des Unterstützungsbedarfs beim Berufseinstieg oder der Prognosekraft von Abschlussexamina) besser zu gestalten.

Vor diesem Hintergrund zielt der vorliegende Beitrag darauf, die Kompetenzentwicklung von Mathematiklehrkräften beim Übergang von der Lehrerbildung in den Beruf längsschnittlich und mit Blick auf die Entwicklung von Lehrerexpertise als einer besonderen Form von Kompetenz zu modellieren. Neben der Analyse von Veränderungen in Form der Abnahme oder Zunahme des mittleren Niveaus des fach-

bezogenen Wissens sowie in Form von Veränderungen in der Rangfolge (Korrelationen) während dieses Übergangs wird dabei in verschiedenen Pfadmodellen geprüft, inwieweit das Wissen ein geeigneter Prädiktor für situationsbezogene und verhaltensnahe Indikatoren von Lehrerexpertise ist.

## 2 Forschungsstand und theoretischer Rahmen der Untersuchung

Im Anschluss an Weinert (1999) wird die am Ende der Lehrerausbildung vorliegende professionelle Kompetenz von Lehrkräften in kognitive Leistungsdispositionen sowie damit verbundene motivationale, volitionale und soziale Bereitschaften und Fähigkeiten differenziert. Kompetenz als Disposition kann in diesem Verständnis analytisch zum einen vom unterrichtlichen Handeln im Sinne von Performanz und zum anderen von situations- und verhaltensnahen kognitiven Fähigkeiten der Wahrnehmung (*perception*), Interpretation (*interpretation*) und Entscheidung über Handlungsoptionen (*decision-making*) unterschieden werden, die die Transformation von Kompetenz in Performanz vermitteln. Letztere stellen eine spezifische, stärker holistische Form der Wissensrepräsentation dar, indem die zuvor getrennt erworbenen und damit ggf. unabhängig voneinander vorliegenden Kompetenzfacetten nunmehr gut vernetzt und flexibel zugänglich vorliegen und situationsspezifisch angewendet werden können (siehe das von Blömeke, Gustafsson und Shavelson [im Druck](#) entwickelte „P-I-D-Modell zur Transformation von Kompetenz in Performanz“; Weinert et al. 1990, S. 176; Besser und Krauss 2009).

Diese kognitiven Fähigkeiten werden in der Literatur häufig als *Expertise* bezeichnet. Kompetenz stellt damit eine wichtige Basis von Expertise dar, ist aber ggf. nicht hinreichend, weil Umstrukturierungsprozesse nötig sind. Im vorliegenden Beitrag stehen zum einen die Entwicklung der kognitiven Kompetenzfacetten im Berufseinstieg sowie zum anderen das Verhältnis von Kompetenz und Expertise im Mittelpunkt.<sup>1</sup>

### 2.1 Fachliches und fachdidaktisches Wissen als Indikatoren für professionelle Kompetenz

Die Expertiseforschung betont die große Bedeutung einer umfangreichen Wissensbasis für die erfolgreiche Ausübung des Lehrerberufs (Bromme 1992). Daher konzeptualisiert die Kompetenzforschung fachliches und fachdidaktisches Wissen auch als zentrale Elemente professioneller Kompetenz von Lehrkräften (Baumert und Kunter 2006). Kompetenzen werden als latente Leistungsdispositionen angesehen, die situativ durchaus unterschiedlich umgesetzt werden können. Unter Entwicklungsgesichtspunkten muss zugleich von Erweiterungs- und Veränderungs-, aber auch von

<sup>1</sup>Um das Nachvollziehen zu erleichtern, hier ein Abkürzungsverzeichnis zu diesen Indikatoren: MCK=*mathematical content knowledge*, d.h. mathematisches Wissen; MPCK=*mathematics pedagogical content knowledge*, d.h. mathematikdidaktisches Wissen; MSpeed=*speed with respect to student errors in mathematics*, d.h. schnelles Erkennen von Schülerfehlern; M\_PID=*perception, interpretation and decision-making with respect to mathematics instruction*, d.h. die präzise Wahrnehmung von mathematikbezogenen Unterrichtssituationen sowie deren zielangemessene Interpretation und flexible Reaktion darauf.

Prozessen des Vergessens oder Absinkens ausgegangen werden. Kompetenzen werden über die Zeit und Einzelsituationen hinweg als relativ stabile (*trait*) Dispositionen betrachtet, die von dynamischen (*state*) Komponenten begleitet werden können.

Die Modellierung der kognitiven Kompetenzen von angehenden wie berufstätigen Mathematiklehrkräften nimmt ihren Ausgangspunkt typischerweise bei beruflichen Anforderungssituationen. Diesen unterliegt in fachbezogener Hinsicht mathematisches und mathematikdidaktisches Wissen (*mathematics content knowledge*, MCK, bzw. *mathematics pedagogical content knowledge*, MPCK; Shulman 1985; Bromme 1997; Autoren 2002; Baumert und Kunter 2006). Was deren Entwicklung im Berufseinstieg angeht, haben Schoenfeld und Kilpatrick (2008) fallweise untersucht. Sie vermuten danach, dass das mathematikdidaktische Wissen im Berufseinstieg aufgrund der Praxiserfahrungen zunehme und sich das Repertoire an Unterrichtsstrategien ausweitere (siehe auch Llinares und Krainer 2006). Das mathematische Wissen werde dagegen nur in Ausschnitten aktualisiert und erweitert. In einer deutschen Studie wird auf der Basis von Querschnittsdaten dagegen weder für MCK noch für MPCK erwartet, dass von einem Zuwachs allein aufgrund von Praxiserfahrung ausgegangen werden kann (Krauss und Brunner 2011). Quasi-längsschnittliche Vergleiche von Kleickmann et al. (2013, S. 98) schließlich weisen Unterschiede in der Höhe von MCK und MPCK zwischen angehenden Lehrkräften am Ende der Ausbildung und berufstätigen Lehrkräften aus. Da Letztere im Schnitt 47 Jahre alt sind und ca. 21 Jahre Berufserfahrung aufweisen, können Aussagen zur möglichen Veränderung durch den Berufseinstieg allerdings nicht getroffen werden.

## 2.2 Transformation von Kompetenz in Performanz

Angesichts der Veränderungen in der Wissensstruktur von beginnenden Lehrpersonen durch die dauerhafte und berufliche (im Unterschied zur ausbildungsbezogenen) Konfrontation mit der Unterrichtspraxis stellt sich die Frage, welcher Zusammenhang zwischen dem in der Lehrerbildung erworbenen Wissen und der Bewältigung situationsspezifischer Unterrichtsanforderungen besteht. In einem einfachen Modell könnte man annehmen, dass Wissen als Element professioneller Kompetenz eine Disposition darstellt, die unmittelbar und mit hoher Effektstärke mit der erfolgreichen Bewältigung unterrichtlicher Anforderungen zusammenhängt. Auch wenn Weinert selbst ein solch einfaches Modell nie im Sinne hatte, sondern früh auf die notwendigen kognitiven Umstrukturierungs- und situativen Adaptionsprozesse hinwies (Weinert et al. 1990), legt seine vielfach zitierte Definition von Kompetenzen als die „bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert 2001, S. 27 f.) eine entsprechende Interpretation nahe.

Der in der Literatur beschriebenen „Wirkungskette“ von *Lehrerbildung* – *Lehrerverhalten* – *Schülerleistung* (zuletzt Terhart 2012) folgend, könnte dies für Mathematiklehrkräfte heißen, dass diejenigen mit höherem fachspezifischen Wissen am Ende der Lehrerbildung einige Jahre später auch angemesseneres Handeln im Mathematikunterricht zeigen und bessere Schülerleistungen in Mathematik errei-

chen. Die Relationen müssten dabei direkt und von hoher Effektstärke sein. Viel Kritik am funktionalen Verständnis des Kompetenzbegriffs (z. B. Nuisl et al. 2002; Pfadenhauer und Kunz 2012) könnte auf einer solch verkürzten Rezeption des Kompetenzbegriffs beruhen.

Wir gehen in unserer Studie dagegen auf die frühen Arbeiten von Weinert zurück und rücken in Ergänzung zu Maßen des fachlichen und fachdidaktischen Wissens von Lehrkräften situations- und verhaltensnahe Maße in den Mittelpunkt der Analyse, um durch eine solche Erweiterung ein zu enges Verständnis der Transformation von Kompetenz in Performanz aufzubrechen. Den neuen Maßen schreiben wir vermittelnde Eigenschaften zu. Um welche Maße es sich handelt, leiten wir aus der Expertiseforschung ab (siehe 2.3). Über videobasierte Tests erfassen wir noch immer „nur“ Kognitionen, die Bewältigung der präsentierten Anforderungen erfordert aber einen schnellen und situationsbezogenen Zugriff auf eine dichte und gut vernetzte Wissensbasis.

### 2.3 Unterrichtswahrnehmung und Geschwindigkeit als Indikatoren für Lehrerexpertise

Unterrichtsbezogene Lehrerexpertise<sup>2</sup> ist durch „den Aufbau und die fortlaufende Aktualisierung situativer mentaler Modelle über den vorhandenen und zu erreichenden Kenntnisstand der Schülerinnen und Schüler, der dafür einzusetzenden Maßnahmen und der dabei zu erwartenden Schwierigkeiten“ (Weinert et al. 1990, S. 191) gekennzeichnet. Ihr Aufbau stellt einen langwierigen Entwicklungsprozess dar, für den rund zehn Jahre umfassender Ausbildung und praktische Erfahrung Voraussetzungen darstellen (Berliner 1986, 2001).

Expertinnen und Experten können Informationen aufgrund höheren Vorwissens, besserer Organisationsleistungen bei der Speicherung und dem Abrufen der Informationen – z. B. durch *chunking* und schnelle Verlagerung ins Langzeitgedächtnis sowie den Besitz von Schemata mit fallbasierten *scripts* – schneller und genauer wahrnehmen, mögliche Probleme antizipieren sowie flexibler reagieren (Calderhead 1984; Gruber 1998). Expertenlehrkräfte strukturieren den Unterricht entsprechend zielangemessener und sie setzen unterrichtliche Anforderungen des Inhalts, der Klassenorganisation und der Schülervoraussetzungen zueinander in Relation, sodass sie über abstraktere Begriffe der Unterrichtskategorisierung verfügen (Bromme 1992). Damit können „während des Unterrichts unter extremem Handlungsdruck vom Lehrer mentale Situationsmodelle gebildet werden (...), die es ihm oder ihr gestatten, aus der Vielzahl der Informationen im Klassenzimmer die relevanten auszuwählen und zu verarbeiten, Entscheidungen verschiedenster Art zu treffen, Probleme zu erkennen, zu lösen und auf diese Weise pädagogisch sensibel, aber auch erfolgreich zu handeln“ (Weinert 1996, S. 149).

Vor diesem Hintergrund betrachten wir im vorliegenden Beitrag die präzise Wahrnehmung von unterschiedlichen mathematikbezogenen Unterrichtssituationen (Clark und Lampert 1986; Carter et al. 1988: „perception accuracy“), deren zielangemessene

<sup>2</sup>Neben Unterrichten müssen Lehrkräfte weitere Anforderungen bewältigen, beispielsweise Erziehen und Schulentwicklung. Diese werden in der vorliegenden Studie nicht betrachtet.

Analyse und Interpretation (Blömeke, Gustafsson und Shavelson [im Druck](#): „interpretation“) sowie die flexible Reaktion darauf (ebd.: „decision-making“) als ersten situationsspezifischen und damit verhaltensnahen Indikator von Lehrerexpertise (kurz M\_PID). In theoretischer Hinsicht lässt sich hier eine Verknüpfung zu dem Konzept des *noticing* von van Es und Sherin (2008) herstellen, in dem in Weiterführungen von Goodwins (1994) Konzept des *professional vision* die beiden Komponenten der „selective attention“ und des „knowledge-based reasoning“ zusammengeführt und um eine handlungsbezogene Komponente erweitert wird: „making connections between specific events and broader principles of teaching and learning“ (Sherin 2007, S. 384; Sherin et al. 2011, S. 80). Fachbezogene Erkenntnisse, die es erlauben, das Wissen von Mathematiklehrkräften mit vignettenbasiert erfassten handlungsnahen Kognitionen zu verknüpfen, würden nicht nur die Natur der Lehrerexpertise weiter aufklären; sie könnten auch helfen, Entscheidungen z. B. hinsichtlich des Unterstützungsbedarfs von Lehrpersonen beim Berufseinstieg besser zu gestalten.

Kersting et al. (2010) haben für das Unterrichten von Bruchrechnung in den USA, gestützt auf die Erkenntnisse der Expertiseforschung (Kersting 2008), ein entsprechendes webbasiertes Videoinstrument entwickelt und validiert. Mathematiklehrkräfte der Grund- und Sekundarschule waren aufgefordert, je nach Studie zu 10 bis 13 online präsentierten Videoclips mit einer Länge zwischen einer und fünf Minuten offene Kommentare zur Lehrer-Schüler-Interaktion im dargestellten realen Mathematikunterricht der fünften und sechsten Klassen abzugeben. Aus den offenen Antworten wurden vier Teilscores zur mathematischen und lernprozessbezogenen Analysefähigkeit (*mathematical content, student thinking*), sowie zur Interpretationsqualität (*depth of interpretation*; entsprechen zusammen in etwa dem Schritt der Interpretation im P-I-D-Modell) und zur Entwicklung von Handlungsoptionen (*suggestions for improvement*; entspricht „decision-making“) gebildet. Eine eigenständige Erfassung von Wahrnehmungsfähigkeiten bleibt ausgespart. Dafür unterzogen sich die Lehrkräfte, die über ein breites Spektrum an Berufserfahrung verfügen, auch einem traditionellen Papier-und-Bleistift-Test zu ihrem mathematischen Wissen (*mathematical knowledge for teaching*).

Die Ergebnisse deuten auf eine signifikante und hohe Korrelation zwischen mathematischem Wissen und den zeitgleich gemessenen verhaltensnahen Kognitionen hin ( $r=0,53-0,62$  je nach Studie). Als Querschnittstudien, die aus interindividuellen Unterschieden auf intraindividuelle Zusammenhänge schließen müssen, unterlagen die Arbeiten allerdings deren typischen methodischen Grenzen, insbesondere im Hinblick auf die Validität dieser Äquivalenzannahme. Auch fehlte – soweit aus den bisherigen Veröffentlichungen erkennbar – ein Instrument, das dezidiert mathematikdidaktisches Wissen erfasst.

Diese Lücke bearbeitet das Video-Assessment von Dunekacke et al. (2013), das ebenfalls auf die Erfassung mathematikbezogener Kompetenz ausgerichtet ist. Mit Hilfe von drei Videos, die Realsituationen aus einer Kindertagesstätte zeigen, werden die Fähigkeiten von Erzieherinnen und Erziehern zur Situationswahrnehmung (entspricht *perception*) und zur Handlungsplanung (entspricht *interpretation* und *decision-making*) erfasst. Die Autoren nehmen an, dass der Situationswahrnehmung eine vermittelnde Rolle in Bezug auf den Zusammenhang von mathematikdidaktischem Wissen sowie Handlungsplanung zukommt. Der latente Zusammenhang zwischen

Wissen und Situationswahrnehmung liegt bei  $\beta=0,60$ , der Gesamtzusammenhang zwischen Wissen und Handlungsplanung (direkter und indirekter Effekt) bei  $\beta=0,58$ .

Weitere Erkenntnisse zum Zusammenhang von Professionswissen und Unterrichtswahrnehmung ergeben sich aus fächerübergreifenden Studien zum *Classroom Management* (Sherin 2007; Stürmer et al. 2012; Gold et al. 2013; König wiedereingereicht). Auch für allgemeindidaktische Planungskompetenzen existieren situationsnahe Vignettenansätze (textbasiert Baer et al. 2007; Blömeke et al. 2008; filmisch basiert Oser et al. 2010). Aus Platzgründen können wir auf diese hier allerdings nicht weiter eingehen.

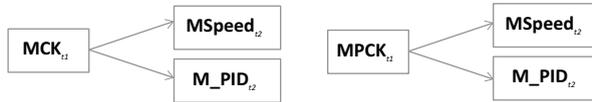
Als zweiten Indikator betrachten wir angesichts der hohen Bedeutung von Geschwindigkeit im Lehrerberuf und den empirischen Hinweisen der Expertiseforschung, dass diese tatsächlich ein geeignetes Maß für die Differenzierung zwischen verschiedenen Expertiseniveaus darstellt, das schnelle Erkennen von mathematikbezogenen Schülerfehlern (*MSpeed*). Krauss und Brunner (2011) haben hierzu ebenfalls einen Reaktionszeittest entwickelt. Ihre Ergebnisse deuten darauf hin, dass fachliche Korrektheit und Bearbeitungszeit unabhängig von der Dauer der Praxiserfahrung variieren. Stattdessen hängt die Fähigkeit, Schülerfehler korrekt und schnell zu erkennen, signifikant vom mathematischen Wissen der Lehrkräfte ab. Als Querschnittstudie war ihre Studie allerdings limitiert in den Aussagen.

### 3 Fragestellungen und Hypothesen

Obwohl Mathematiklehrkräfte die am besten untersuchte Lehrergruppe sind, fehlen Studien, die standardisiert längsschnittlich die intraindividuelle Veränderung des in der Ausbildung erworbenen fachbezogenen Professionswissens im Berufseinstieg sowie dessen Zusammenhang mit situations- und verhaltensnahen kognitiven Fähigkeiten untersuchen. Das Projekt „*Längsschnittliche Entwicklung der Kompetenzen von Junglehrkräften: Follow-Up zur internationalen Vergleichsstudie TEDS-M (TEDS-FU)*“<sup>3</sup> intendiert, diese Forschungslücke in Bezug auf Mathematiklehrkräfte der Sekundarstufe I zu verringern. Indem das in der Ausbildung erworbene Wissen zwei Mal innerhalb von vier Jahren erfasst und in Form einer Mediationsanalyse mit situationsbezogenen und damit verhaltensnahen Maßen verknüpft wird, wird erstmals eine echte Prozessanalyse möglich, die in der Bildungsforschung noch immer eine Ausnahme darstellt (Schmitz 2006), aber für dringend erforderlich gehalten wird, um dynamischen Vorgängen auf die Spur kommen zu können (Judd und Kenny 1981).<sup>3</sup>

Dabei stellt sich zum ersten die Frage, wie sich das MCK- und MPCK-Niveau in den ersten Berufsjahren verändert. Der bisher vorliegende Forschungsstand hierzu ist uneinheitlich (Schoenfeld und Kilpatrick 2008; Krauss und Brunner 2011; Kleickmann et al. 2013). Die Entwicklung des mathematischen und des mathematikdidaktischen Wissens während der ersten Berufsjahre sollte danach aber in jedem Falle durch differenzielle Effekte gekennzeichnet sein. Während das mathematische

<sup>3</sup>TEDS-FU wurde von der Deutschen Forschungsgemeinschaft gefördert (DFG Geschäftszeichen:BL 548/B-1). Die vorliegenden Aussagen sind jene der Autorinnen und Autoren und spiegeln nicht notwendigerweise die Sichtweise der DFG wider.



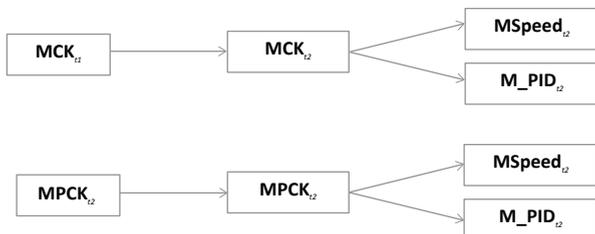
**Abb. 1** Direkte Vorhersage von verhaltensnahen kognitiven Fähigkeiten nach vier Jahren im Beruf durch fachbezogenes Wissen am Ende der Ausbildung (einfaches Kompetenzmodell)

Wissen, das durch praktische Unterrichtserfahrung nur indirekt und immer nur in kleinen Ausschnitten aktualisiert wird, nach den ersten Berufsjahren aufgrund von Vergessenseffekten vermutlich eher geringer ausgeprägt ist als am Ende der Ausbildung (H1a), sollte das mathematikdidaktische Wissen aufgrund der täglichen Auseinandersetzung der Mathematiklehrkräfte mit der Unterrichtspraxis mindestens auf derselben Höhe liegen oder sogar signifikant ansteigen (H1b).

Zum zweiten stellt sich die Frage, wie das am Ende der Ausbildung vorliegende fachbezogene Wissen mit situations- und verhaltensnahen kognitiven Fähigkeiten wie der Wahrnehmung, Interpretation und Entscheidung über Handlungsoptionen (PID) sowie dem schnellen Erkennen von Schülerfehlern zusammenhängt. Der Forschungsstand lässt vermuten, dass MCK und MPCK signifikant  $M\_PID$  vier Jahre später vorhersagen (H2a), während für  $MSpeed$  nach Krauss und Brunner (2011) MCK ggf. relevanter ist als MPCK (H2b). Wegen der dazwischenliegenden kognitiven Umstrukturierungsprozesse nehmen wir für diese direkten Zusammenhänge eher geringe Effektstärken an. Ferner gehen wir davon aus, dass die Residuen der beiden abhängigen Variablen  $MSpeed$  und  $M\_PID$  in den Pfadmodellen nicht signifikant korrelieren, dass ihr Zusammenhang also hinreichend durch MCK bzw. MPCK erklärt werden kann (siehe Abb. 1).

Größere Effektstärken erwarten wir von einem Mediationsmodell, dass die Wissensentwicklung zwischen dem Ende der Ausbildung und dem zweiten Messzeitpunkt vier Jahre später berücksichtigt (siehe Abb. 2). Diese Überlegung ergibt sich aus den oben dargelegten Erkenntnissen der Expertiseforschung zur Veränderungen von Wissensrepräsentationen aufgrund von ausgedehnter Praxiserfahrung und den Ergebnissen von Kersting et al. (2010) sowie Dunekacke et al. (im Druck), die einen starken Zusammenhang zwischen zeitgleich vorliegendem mathematischen bzw. mathematikdidaktischen Wissen von Lehrkräften bzw. Erzieherinnen und ihren handlungsbezogenen Kognitionen, erfasst über einen Video-Vignettest, fanden. Zusammengenommen führt dies zur Hypothese, dass das am Ende der Ausbildung vorliegende mathematische bzw. mathematikdidaktische Wissen mit praktisch bedeutsamen Effektstärken das vier Jahre später vorliegende Wissen vorhersagt und

**Abb. 2** Differenziertes Kompetenzmodell, wonach der Zusammenhang zwischen verhaltensnahen Maßen und Wissen am Ende der Ausbildung durch die Wissensentwicklung vermittelt wird



dass dieses wiederum praktisch bedeutsam vorhersagt, wie gut die Lehrkräfte Unterricht wahrnehmen, interpretieren und Handlungen generieren und ob sie schnell Schülerfehler erkennen (H3).

Mit diesen Hypothesen sind weitreichende Aussagen verbunden, insofern ist eine Validierung unserer Modelle notwendig (Kane 2013). Um dies zu erreichen, werden die verwendeten Maße mit Außenkriterien in Verbindung gesetzt, die aus einem nomologischen Netzwerk abgeleitet werden, das heißt aus einem theoretisch und empirisch erwartbaren Zusammenhang. Dies betrifft zum einen die Vorhersage der Mathematikleistung und der Mathematikdidaktikleistung am Ende der Lehrerausbildung durch die Abiturnote (H4a), wofür zahlreiche Belege aus früheren Studien vorliegen (u. a. Blömeke 2009).

Ein weiterer Validitätsbeleg lässt sich über die Note im Zweiten Staatsexamen erbringen. Dieses Examen ist – anders als das Erste Staatsexamen – konzeptionell dezidiert auf *Performanz* im Unterricht ausgerichtet. Diese wird von verschiedenen Expertinnen und Experten (Schulleitung, Fach- und Hauptseminarleitungen, Staatliche Prüfungsämter) anlässlich zahlreicher Unterrichtsbesuche beurteilt. Auch wenn bisher, soweit bekannt, keine empirischen Belege für die Reliabilität und Validität dieser Note vorliegen, lässt ihre Konzeption plausibel erwarten, dass sie – bei allen Einschränkungen, die auch für die Abiturnote gelten (insbesondere unterschiedliche Referenzrahmen der jeweiligen Beurteiler) – auf Gruppenebene die Unterschiede in der Qualität des Unterrichtshandelns hinreichend verlässlich abbildet. Unter dieser Annahme sollten die in unserem Videotest gezeigten Wahrnehmungs-, Interpretations- und Entscheidungsleistungen signifikante Zusammenhänge zur Note im Zweiten Staatsexamen zeigen, da wir beanspruchen, dass jene eine Voraussetzung für Handeln im Unterricht darstellen.

#### 4 Untersuchungsmethode

2008 wurden im Rahmen der „Teacher Education and Development Study: Learning to Teach Mathematics (TEDS-M)“ angehende Primar- und Sekundarstufen-I-Lehrkräfte für das Fach Mathematik im letzten Jahr ihrer Ausbildung aus 15 Ländern getestet und befragt (Tatto et al. 2008, 2012). Eine Lehrberechtigung für die Jahrgangsstufe 8 wurde als Kriterium für die Identifikation von Sekundarstufen-I-Ausbildungsgängen festgelegt, die den Fokus des vorliegenden Beitrags darstellen. In Deutschland umfasste diese Definition angehende Mathematiklehrkräfte mit einer Lehrberechtigung für die Klassenstufen 1 bis 9/10 (Grund-, Haupt- und Realschullehrkräfte; GHR), für die Klassenstufen 5/7 bis 9/10 (Haupt- und Realschullehrkräfte; HR) oder für die Klassenstufen 5/7 bis 12/13 (Gymnasial- und Gesamtschullehrkräfte; GyGS). Die Stichprobenziehung folgte einem mehrstufigen stratifizierten Samplingdesign mit Zufallsziehungen repräsentativer Einheiten sowohl auf der Ebene der Ausbildungsinstitutionen (in Deutschland Studienseminare aller 16 Bundesländer) als auch auf der Ebene der angehenden Mathematiklehrkräfte. Die Ausbildungsgänge bildeten in Deutschland das explizite und die Bundesländer das implizite Stratum.

Um das fachbezogene Wissen der Mathematiklehrkräfte zu erfassen, wurden diese in TEDS-M mit 76 mathematischen und 27 mathematikdidaktischen Items im Mul-

multiple-Choice-, Complex-Multiple-Choice- und Constructed-Response-Format getestet. Im MCK-Test wurden vier Subdimensionen abgedeckt: Arithmetik (27 Items), Algebra (22 Items), Geometrie (23 Items) und Stochastik (4 Items). Als kognitive Anforderungen ging es um das Erinnern von Wissen, um das Anwenden und um das Generieren von Handlungsoptionen bzw. Bewerten. Der Test für das mathematikdidaktische Wissen beinhaltete zwei Subdimensionen: curriculares und planungsbezogenes Wissen (12 Items) sowie interaktionsbezogenes Wissen (15 Items). Die Leistungsmaße wiesen gute bzw. sehr gute Reliabilitäten auf ( $MLE_{MCK} = 0,91$ ;  $MLE_{MPCK} = 0,72$ ; vgl. Tatto et al. 2012).

#### 4.1 Stichprobe der Follow-Up-Studie TEDS-FU im Jahr 2012

Vier Jahre nach der Erhebung der TEDS-M-Daten wurden 2012 171 Sekundarstufen-I-Lehrkräfte mit dem Unterrichtsfach Mathematik untersucht, die sich ungefähr im dritten Berufsjahr befanden. Sie hatten 2008 an TEDS-M teilgenommen und danach den Übergang in den Beruf vollzogen. 2011 waren diejenigen, die sich 2008 mit einer Folgestudie einverstanden erklärt hatten, erneut von uns kontaktiert und zu ihren Berufserfahrungen, ihrer Berufszufriedenheit, ihren Überzeugungen und zu Schulkontext-Merkmalen befragt worden (für Ergebnisse aus dieser Studie siehe Blömeke und Klein 2013). 2012 erfolgte dann nach einer zweiten Kontaktaufnahme die Durchführung der vorliegenden Studie.

Von den 771 angehenden Sekundarstufen-I-Lehrkräfte, die an TEDS-M teilgenommen hatten, gaben rund 62 % (d.h. 477 Personen<sup>4</sup>) ihr Einverständnis, für eine Folgestudie kontaktiert zu werden. Von dieser Gruppe konnten rund 50 % nach drei Jahren wieder aufgefunden werden, also 240 Personen. Hauptgründe für ein Nichtauffinden waren inkorrekt angegebene oder fehlende Codes zur längsschnittlichen Verknüpfung der Daten bzw. nicht mehr funktionierende E-Mail-Adressen. 221 der wieder aufgefundenen 240 Personen (92 %) haben sich bereit erklärt, an einer Befragung zu ihren aktuellen Arbeitsbedingungen teilzunehmen (Blömeke und Klein 2013). Ein Jahr später hat die vorliegende Studie mit 171 Sekundarstufenlehrkräften stattgefunden (71 % der wieder aufgefundenen TEDS-M-Teilnehmenden).

Die TEDS-FU-Stichprobe setzt sich etwa je zur Hälfte aus Gymnasiallehrkräften (55,8 %) sowie Haupt- und Realschullehrkräften zusammen (43,6 %). Die Mathematiklehrpersonen sind im Mittel 32 Jahre alt ( $SD=5,9$ ;  $min=26$ ,  $max=53$ ) und eine Mehrheit ist weiblich (58,9 %). Sie verfügen über vergleichsweise hohes kulturelles Kapital: 53,7 % geben an, zu Hause drei oder mehr Regale mit Büchern bzw. spezialisiert mehr als 200 Bücher zu besitzen. In die Lehrerausbildung ist die Stichprobe mit guten Eingangsvoraussetzungen gegangen: 73,3 % geben an, einen Leistungskurs in Mathematik besucht zu haben; die mittlere Abiturnote lag bei 2,1 ( $SD=0,6$ ;  $min=1,0$ ,  $max=3,9$ ). Entsprechend gut wurden die beiden Staatsexamina absolviert: das erste Examen im Durchschnitt mit 1,9 ( $SD=0,5$ ;  $min=1,0$ ,  $max=3,7$ ), das zweite Examen mit 2,1 ( $SD=0,7$ ;  $min=1,0$ ,  $max=4,1$ ).

<sup>4</sup>Aufgrund von Datenschutzaufgaben war uns untersagt, die stufenspezifische Klassifizierung zu erfassen; diese Zahl ist linear aus dem Gesamtpool an Einverständniserklärungen abgeleitet.

Auch wenn die TEDS-FU-Stichprobe in diesen Hintergrundmerkmalen der repräsentativen Ausgangsstichprobe von TEDS-M in 2008 ähnlich ist, weisen wir darauf hin, dass es sich um eine positiv verzerrte Gelegenheitsstichprobe handelt. Die Teilnahme war freiwillig und die Leistungsmerkmale in Bezug auf das mathematische und mathematikdidaktische Wissen (siehe unten) weisen deutlich darauf hin, dass vorwiegend Lehrpersonen aus dem oberen Kompetenzspektrum an TEDS-FU teilgenommen haben.

## 4.2 Instrumente

### 4.2.1 *Situationsspezifische Unterrichtswahrnehmung, -interpretation und -entscheidung (PID)*

Die präzise Wahrnehmung und Interpretation von mathematikbezogenen Unterrichtssituationen sowie die zielangemessene Reaktion und Fortführung des Unterrichts durch die Mathematiklehrkräfte (*M\_PID*) wurde mithilfe eines videobasierten Tests erfasst, der das Konstrukt in seiner Breite abdeckt. Eine spezifische Erfassung jeder Subdimension durch eigene Teilscores war nicht angestrebt. Die Lehrkräfte wurden informiert, dass es sich – vier Jahre nach TEDS-M – um den zweiten Teil der Erfassung ihrer Kompetenzen und deren Entwicklung im Beruf handelt.

Drei Videovignetten, die zwischen 2,5 und 4 min dauern, zeigen Mathematikunterricht der Klassen 8 bis 10 aus allen Schulformen der Sekundarstufe I, wobei die Unterrichtsszenen gestellt waren („scripted“ lessons). Den Lehrpersonen wurde dies vorab mitgeteilt. Auf diese Weise konnten wir unsere Intention realisieren, einen Eindruck von der gesamten Stunde zu geben, also Eingangs-, Erarbeitungs- und Auswertungsphase, aus denen im Video einzelne Teilsequenzen gezeigt wurden. Des Weiteren sollten gezielt ausgewählte Aspekte von Unterrichtsqualität, die als zentral und häufig vorkommend bekannt sind und u. a. von Helmke (2010) dargestellt werden, in den Vignetten eine Rolle spielen. Im Unterschied beispielsweise zu Kersting et al. (2010) oder Dunekacke et al. (im Druck) haben wir uns daher für gestellten Unterricht entschieden, um eine solche Dichte an Ereignissen zu erreichen.

Die Lehrpersonen erhielten Kontextinformationen zur Zusammensetzung der Klasse, den Vorkenntnissen der Schülerinnen und Schüler sowie den Inhalten der letzten Stunden, damit sie die gezeigten Unterrichtssequenzen angemessen in ihren Kontext einordnen konnten. Es wurden zudem die mathematischen Inhalte erklärt, die Gegenstand der Vignette sein würden, und die Lösungen der thematisierten Schüleraufgaben, um zu vermeiden, dass sich die Versuchspersonen vorrangig mit dem mathematischen Inhalt befassen. Der Fokus ihrer Aufmerksamkeit sollte auf fachdidaktischen und pädagogischen Aspekten liegen. Sie erhielten die Anweisung, sich das folgende Video anzuschauen und anschließend die damit verknüpften Aufgaben zu bearbeiten. Jedes Video konnte nur einmal angesehen werden, um eine Unterrichtssituation so realitätsnah wie möglich nachzustellen. Auch dies wurde den Lehrpersonen vorab bekannt gegeben.

Die Drehbücher wurden mithilfe von freiwillig teilnehmenden Schülerinnen und Schülern umgesetzt. Technisch realisiert wurden die video-basierten Tests – wie das gesamte TEDS-FU-Testpaket – als Online-Tests mit dem CBA ItemBuilder des DIPF (Rölke 2012). Die ausgewählten Unterrichtssituationen gehören inhaltlich zu

den Bereichen Algebra und Geometrie und waren anhand mathematikdidaktischer und pädagogischer Kriterien sowie auf Grundlage von Hinweisen erfahrener Lehrkräfte entwickelt worden, um eine angemessene Relevanz und Häufigkeit des Auftretens der Situationen im Schulunterricht zu sichern. Eine inhaltliche Validierung, inwieweit dies durch die Videoclips tatsächlich erreicht worden waren, erfolgte durch Expertenreviews und Pilotierungen mit Personengruppen auf unterschiedlichem Leistungsniveau (Kane 1992).

Zur Erfassung der Wahrnehmungsgenauigkeit wurden 4-stufige Ratingskalen eingesetzt, die anhand von Clausen et al. (2003) entwickelt worden waren. Eine 26-köpfige Expertengruppe, bestehend aus 13 universitären Mathematikdidaktikern und 13 Mathematikfachleitungen, entschied in einem zweistufigen Prozess, welche Antwort als korrekt akzeptiert werden konnte (für Details siehe Kaiser et al. 2013). Akzeptiert wurde eine Mindestübereinstimmung von 60% auf genau einer der vier Ausprägungen (z. B. „trifft voll und ganz zu“). Items, die diese Übereinstimmung nicht erreichen konnten, bei denen aber mindestens 75% der Expertinnen und Experten in der Tendenz übereinstimmten (Zusammenführung der Kategorien „trifft voll und ganz zu“ und „trifft eher zu“ bzw. „trifft eher nicht zu“ und „trifft überhaupt nicht zu“), wurden dichotom ausgewertet.

Alle übrigen Items wurden überarbeitet und einem erneuten Expertenrating von zwölf Personen – sechs Fachseminarleitungen und sechs Universitätsmitarbeiter – unterzogen. Obwohl solche Einschätzungen angesichts der Komplexität der dargebotenen Situationen typischerweise schwieriger eindeutig zu treffen sind als bei klassischen Papier-und-Bleistift-Tests, die die Komplexität reduzieren (Kane 1992), war die letztlich erreichte Übereinstimmung sehr zufriedenstellend (nach zwei Runden mit Überarbeitungen: 85%).

Über offene Aufgaben wurde die Fähigkeit der Lehrkräfte erfasst, den gezeigten Mathematikunterricht interpretieren und zielangemessen sowie flexibel Entscheidungen bezüglich Handlungsoptionen generieren zu können. Dabei wurde für jede einzelne Aufgabe geprüft, dass für ihre Lösung tatsächlich die Videovignette benötigt wurde und sie nicht allein durch eine Aufgabenanalyse entschieden werden konnte. Konkret bedeutete dies beispielsweise in Bezug auf Abb. 4, dass nicht nur die an der Tafel abgebildete Aufgabenstellung relevant war, sondern vor allem die Phasen des dargebotenen Unterrichts, die sich um die Aufgabe herum entwickelten, wie die Variation der Ausgangsbedingungen, die Umkehrung der Aufgaben, die Entwicklung mehrerer Lösungswege sowie die Begründung des Vorgehens (Ulm 2004). Damit beziehen wir uns auf die konkret gezeigte Unterrichtssituation und sind auf der Ebene der situationsbezogenen Fähigkeiten.

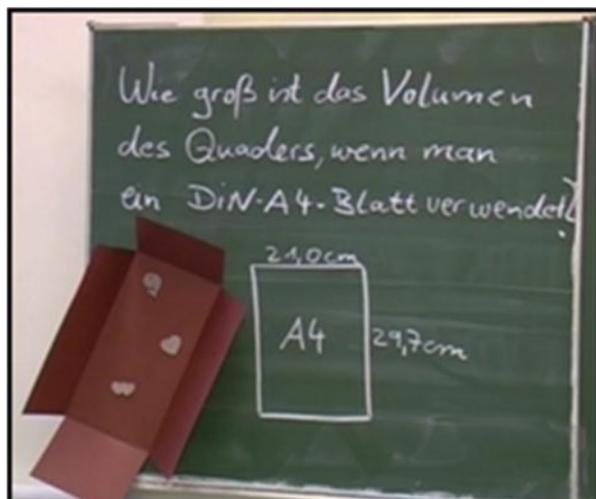
Für die Kodierung der Antworten wurden in einem Wechselspiel von Deduktion (aus unserem theoretischen Rahmen) und Induktion (aus den empirischen Antworten von angehenden Lehrkräften in Pilotstudien, ergänzt um die Antworten aus der Hauptstudie) umfassende Kodierhandbücher mit allgemeinen Kriterien für richtige und falsche Antworten sowie mit Ankerbeispielen und Grenzfällen erstellt. Die Kodes sollten dabei so wenige Interpretationen wie möglich erfordern, also niedriginferent sein. Alle Antworten wurden auf dieser Basis kodiert. Die Ausführlichkeit des Manuals und eine intensive Schulung der Kodierer sicherten eine gute Inter-Koder-Reliabilität, wie unabhängige Doppelkodierungen von 20% der Fälle zeigen.

Wir erreichten ein gutes (Jonsson und Svingby 2007) Cohen's Kappa von im Mittel  $\kappa=0,86$  ( $\min=0,76$ ,  $\max=1,00$ ). In den seltenen Fällen fehlender Übereinstimmung wurde in einer gemeinsamen Diskussion der Rater, ggf. unter Hinzuziehung weiterer Expertinnen und Experten entschieden, wie eine Antwort kodiert werden sollte. Im nächsten Schritt des Scoring wurde entschieden, welche Codes als korrekt und welche als inkorrekt in die Skalierung eingehen sollten.

Im mathematikdidaktischen Videotest bearbeiteten die Lehrkräfte insgesamt 34 Items. Diese setzten sich je zur Hälfte aus Ratingskalen ( $n=16$ ) und offenen Aufgaben ( $n=18$ ) zusammen, die wiederum relativ gleichmäßig über die drei Videovignetten verteilt waren (Vignette 1:  $n=6$  bzw. 5, Vignette 2:  $n=5$  bzw. 8, Vignette 3:  $n=5$  bzw. 5).

Die von uns entwickelte Form der Testung ausgehend von beruflichen Situationen über die kontextualisierte Darbietung zentraler und häufiger kritischer Unterrichtsereignisse in Form von Videos (oder alternativ in anderen Studien über schriftliche Vignetten) hat Ähnlichkeit mit so genannten *Situational Judgment Tests*, die in Europa kaum bekannt sind, in den USA aber häufig bei der Personalauswahl eingesetzt werden und dementsprechend gut untersucht sind (Lievens et al. 2008). Meta-Analysen bescheinigen ihnen eine hohe kriteriale Validität, was die Vorhersage zukünftiger Leistungen im Beruf angeht, große Akzeptanz bei Bewerberinnen und Bewerbern, die diese Tests absolvieren, und inkrementelle Validität über andere Maße wie Wissens- oder Intelligenztests hinaus (McDaniel et al. 2001). Dies gilt, obwohl sich diese Tests im Unterschied zu unserem Assessment mit vielen offenen Aufgaben, in der Regel auf Multiple-Choice-Items beschränken, in denen entweder besonders gute Handlungsoptionen aus einem vorgegebenen Spektrum auszuwählen oder deren Effektivität einzuschätzen ist. Als Probleme mit dieser Art von Testung muss mit Mehrdimensionalität und verbunden damit vergleichsweise geringer Reliabilität der Messungen gerechnet werden.

**Abb. 3** Videotest „Schachtelvignette“ (Mathematikunterricht, 9. Klasse, Gymnasium)



ALLE Aussagen beziehen sich auf die im Video gezeigten Unterrichtsausschnitte. Markieren Sie zu den folgenden Aussagen jeweils den Grad Ihrer Zustimmung.

Bitte treffen Sie pro Zeile EINE Auswahl!

	trifft voll und ganz zu	trifft eher zu	trifft eher nicht zu	trifft überhaupt nicht zu
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Aufgabe enthält einige Merkmale einer offenen Aufgabenstellung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abb. 4 Beispiel-Item zur Erfassung der Wahrnehmungsfähigkeit (M\_PID, *perception*)





1                      2                      3

Im Video wurden drei Paare in ihrem Arbeitsprozess genauer betrachtet. Diese Arbeitsprozesse sollen im Folgenden aus zwei Perspektiven – einer (a) MATHEMATIKDIDAKTISCHEN und einer (b) PÄDAGOGISCHEN – betrachtet werden:

(a) MATHEMATIKDIDAKTISCHE PERSPEKTIVE  
 In jeder der drei gezeigten Herangehensweisen wird die Aufgabe MATHEMATISCH auf eine GANZ EIGENE ART DARGESTELLT UND BEARBEITET.  
 Beschreiben Sie kontrastierend die WESENTLICHEN Aspekte der Herangehensweisen aus mathematikdidaktischer Sicht (in Stichworten).  
 Nennen Sie dabei – falls möglich – auch die dazugehörigen Fachbegriffe.

Den zweiten Teil der Aufgabe finden Sie auf der folgenden Seite.

Abb. 5 Beispiel-Item zur Erfassung der Analysefähigkeit (M\_PID, *interpretation*)

**Beispiel: Schachtelvignette**

Beispielhaft sei die „Schachtelvignette“ beschrieben. In diesem Video werden Szenen aus einer Mathematikstunde in einer leistungsheterogenen 9. Klasse an einem Gymnasium gezeigt. Die Schülerinnen und Schüler führen zu der in Abb. 3 gezeigten Aufgabe eine Partnerarbeit durch. Die Art und Weise, in der die Jugendlichen das Problem bearbeiten, unterscheidet sich u. a. aus mathematikdidaktischer Sicht. Im Anschluss an die Partnerarbeit findet ein Plenumsgespräch zum Austausch der Ergebnisse statt. Auf vierstufigen Likertskalen waren Aussagen zum dargebotenen Video einzuschätzen, die spezifisch die Fähigkeit zur Wahrnehmung erfassten (siehe Abb. 4). Als korrekt wurde im vorliegenden Beispiel „trifft voll und ganz zu“ kodiert.

Zur Erfassung der Interpretationsfähigkeit zeigt das Video drei Schülerpaare in ihrem Arbeitsprozess während der Lösung einer mathematischen Aufgabe. Die in dem Video gezeigten Paare bearbeiten das gegebene mathematische Problem orientiert an den drei Brunerschen Repräsentationsmodi *ikonisch*, *enaktiv* und *symbolisch* (Bruner 1966). Diese Arbeitsprozesse sollten von den Lehrkräften zum Beispiel durch folgende Aufgabenstellung herausgearbeitet werden: „Beschreiben Sie kontrastierend die wesentlichen Aspekte der Herangehensweisen aus mathematikdidaktischer Sicht (in Stichworten). Nennen Sie dabei – falls möglich – auch die dazugehörigen Fachbegriffe.“ Dabei dienten jeweils Bilder der Situationen als Erinnerungshilfe.

In den in Abb. 5 dargestellten Items kommen verschiedene Expertiseaspekte zum Tragen. Unter anderem muss hier Relevantes von Irrelevantem unterschieden werden, es muss wissensbasiert wahrgenommen und interpretiert werden, und es müssen Strukturen erkannt sowie Phänomene klassifiziert werden. Im Folgenden wird – auch exemplarisch für die Kodieranweisungen zu den übrigen Items – stark gekürzt das Kodiermanual zu diesen Items dargelegt und kommentiert.

Erwartet wurde, dass eine Beschreibung *über die unmittelbare Wiedergabe des Gesehenen ohne Hervorhebung des Relevanten hinausgeht*. Das beinhaltet ein minimales Maß an Verallgemeinerung. Dabei kann im konkreten Fall der Abstraktionsgrad der Brunerschen Klassifikation erreicht werden, muss aber nicht. Leitlinien bei der Kodierung waren darüber hinaus in diesem Beispiel, dass die Beschreibung kontrastierend war, das jeweilige Alleinstellungsmerkmal des betrachteten Paares im Vergleich zu den anderen beiden Paaren also deutlich wurde. Und schließlich sollte der Fachausdruck das wesentliche Moment des Gesehenen in einem Wort zusammenfassen. Auch wenn eine Testperson nicht explizit auf die Brunerschen Repräsentationsebenen Bezug nahm, die Beschreibung aber diesen Leitlinie folgte, wurde die Äußerung als korrekt kodiert. Anstelle der Brunerschen Bezeichnung „symbolisch“ konnte z. B. auch „algebraisch“ oder „analytisch“ verwendet werden. Nicht als korrekt gewertet hingegen wurde die Bezeichnung „mathematisch“, weil diese nicht kontrastierend war, da alle drei Paare auf die eine oder andere Weise mathematisch vorgehen.

Die oben genannten allgemeinen Charakterisierungen zur Kodierung der Beschreibungen werden ergänzt durch zahlreiche in-vivo-Antwortbeispiele für *korrekt*, *falsch* und *im Grenzfall korrekt*, die aus der Pilotierung und – bei Bedarf ergänzend – aus

**Tab. 1** Kodierbeispiele für das in Abb. 5 gezeigte Item

	Im Idealfall erwartete Beschreibung	Minimales Maß an Verallgemeinerung (als Grenzfall korrekt)	Unmittelbare Wiedergabe des Gesehenen ohne Hervorhebung des Relevanten (nicht korrekt)
Paar 1	Sie arbeiten mithilfe einer bildhaften Darstellung des Objektes	Sie arbeiten an einem Modell	Sie einigen sich auf eine Seitenlänge
Paar 2	Sie arbeiten am konkreten Objekt	Sie messen am Quader	Sie schneiden die Ecken aus und biegen sie hoch
Paar 3	Sie arbeiten mit abstrakten Repräsentanten wesentlicher Problemaspekte	Sie arbeiten mit einer variablen Länge	Sie rechnen mit $x$ , ziehen $2x$ von $29,7$ ab, ziehen $2x$ von $21,0$ ab und multiplizieren alle drei Faktoren

In den Bildungsstandards der Kultusministerkonferenz werden ALLGEMEINE MATHEMATISCHE KOMPETENZEN formuliert. Nehmen Sie an, Sie müssten die Stunde an der Stelle, an der das Video endet, INHALTLICH SINNVOLL ANKNÜPFEND WEITERFÜHREN: Formulieren Sie in einem Satz wörtlich jeweils einen Arbeitsauftrag, den Sie der Klasse geben würden, ...

... wenn Sie die allgemeine mathematische Kompetenz MATHEMATISCHE DARSTELLUNGEN VERWENDEN betonen wollen.

..., wenn Sie die allgemeine mathematische Kompetenz MATHEMATISCH ARGUMENTIEREN betonen wollen.

Info  
zurück  
weiter

Abb. 6 Beispiel-Item zur Erfassung des Handlungsrepertoires (M\_PID, decision-making)

Abb. 7 Beispiel-Item zur Erfassung der schnellen Erkennung von Schülerfehlern (MSpeed)

Anwendung der p-q-Formel bei der quadratischen Gleichung $x^2 + 8x - 4 = 0$	$x_{1,2} = -4 \pm \sqrt{16 + 4}$	$x_{1,2} = -4 \pm \sqrt{16 - (-4)}$	$x_{1,2} = -4 \pm \sqrt{16 - 4}$
---	----------------------------------	-------------------------------------	----------------------------------

der Hauptstudie gewonnen wurden. Die Kodieranleitungen für die anderen Items des Videotests folgen demselben Muster. Tab. 1 zeigt Beispiele aus dem Kodiermanual.

Das Handlungsrepertoire wurde schließlich mit Aufgaben wie der folgenden erfasst (siehe Abb. 6). Um typischen Problemen performanznaher Assessments entgegenzuwirken, in denen das Fortführen von Situationen häufig in einer Beliebigkeit der Antworten und damit ihrer Unkodierbarkeit resultieren, wurde der Raum möglicher Antworten eingegrenzt. Die Fragen erfolgten also nicht in der Form „Wie würden Sie weitermachen?“, sondern es wurden präzise definierte einschränkende Anforderungen an die Antwort gestellt, indem beispielsweise eine inhaltlich sinnvolle Anknüpfung in Form eines konkret zu formulierenden Arbeitsauftrags unter Berücksichtigung gegebener allgemeiner mathematischer Kompetenzen gefordert wurde.

#### 4.2.2 Fachbezogenes Professionswissen von Mathematiklehrkräften

Das mathematische Professionswissen (MCK) wurde mit einer Kurzversion des TEDS-M-Tests erfasst, die 30 der ursprünglich 76 Items enthielt. Diese deckten mit Arithmetik, Algebra und Geometrie sowie einigen Stochastik-Items die ursprüngliche inhaltliche Breite ebenso ab wie mit Multiple-Choice-Items, komplexen Multiple-Choice-Items und Constructed-Response-Items die verwendeten Item-Formate und mit Wissen, Anwenden und Begründen die verschiedenen kognitive Anforderungsniveau des Ausgangstests. Für weitere Informationen verweisen wir auf die zahlreichen Veröffentlichungen zu TEDS-M, insbesondere auf Blömeke, Kaiser und Lehmann (2010) sowie als zentrale Ergebnisberichte Blömeke et al. (2011, 2012).

Das mathematikdidaktische Professionswissen (*MPCK*) wurde mit dem Originalinstrument aus TEDS-M erfasst, das 27 Items enthielt. Diese erfassen curriculares und planungsbezogenes Unterrichtswissen wie beispielsweise Zusammenhänge zwischen Inhalten des Lehrplans herstellen, didaktische Konzepte zu zentralen mathematischen Themen, typische Fehlvorstellungen oder verschiedene Repräsentationsmodi kennen sowie Wissen zur Lehrer-Schüler-Interaktion im Mathematikunterricht wie beispielsweise Schülerantworten zu diagnostizieren und darauf zu reagieren.

#### 4.2.3 Schnelles Erkennen von Schülerfehlern

Wie rasch die Mathematiklehrkräfte Schülerfehler entdecken konnten, wurde über einen Geschwindigkeitstest erfasst, der sich an das Paradigma des „*experimenter-based testing*“ anlehnt (Davison et al. 2011). Hierbei werden die Testteilnehmer vorab darüber informiert, dass ihnen nur ein eng begrenztes Zeitfenster zur Verfügung steht, um bestimmte Aufgaben zu lösen. Sie bekommen auch Gelegenheit, diese Form des Assessments zu trainieren. Studien zeigen, dass dieses sich besonders gut eignet, um durch den Geschwindigkeitsfaktor zusätzliche Informationen über die Korrektheit von Antworten hinaus zu gewinnen (ebd.).

In Anlehnung daran präsentieren wir in unserem Test 16 Aufgaben, die typische Fehler und Fehlkonzepte der Mittelstufe – beispielsweise bei der Multiplikation von Brüchen – widerspiegeln (siehe u. a. Padberg 1996; Fischer und Malle 2004). Es wurde jeweils zunächst das Thema gezeigt, zu dem anschließend ein typischer Fehler präsentiert wurde. Die Lehrkräfte wurden aufgefordert, beim Lesen solche Fehler zu antizipieren. In dieser Antizipationsphase sollten sich die Versuchsperson die Schar möglicher *typischer* Fehler bezogen auf ein Thema vergegenwärtigen, um dann bei der Begegnung mit drei Schülerantworten sehr schnell – innerhalb von vier Sekunden – auf die typischen kritischen Stellen zu fokussieren. Ein Durchrechnen der Schülerantworten ist aufgrund des starken Zeitdrucks nicht möglich, sodass nur durch eine angemessene Antizipation der Fehler schnell entdeckt werden kann. An dieser Stelle kommt die Lehrerexpertise zum Tragen: Es ist davon auszugehen, dass Expertenlehrkräfte über ein wesentlich reichhaltigeres Wissensrepertoire über typische Schülerfehler verfügen als Novizen. In dieser Hinsicht unterscheidet sich der in TEDS-FU verwendete Geschwindigkeitstest deutlich von dem bei COACTIV benutzten (Krauss und Brunner 2011). Bei Letzterem lag der Fokus nicht auf *typischen* Schülerfehlern, und es gab konsequenterweise auch keine Möglichkeit der Fehlerantizipation.

Wurde keine Taste gedrückt, wurde die Antwort entsprechend dem Vorgehen beim *experimenter-based testing* als falsch kodiert. Wurde eine richtige Lösung als falsch identifiziert, wurde sie ebenfalls als falsch kodiert. Auf diese Weise stellt der resultierende Score eine Kombination von Geschwindigkeit und Korrektheit dar, die sich in anderen Studien auch empirisch abbilden ließ (Davison et al. 2011): Zeitlich wurden die Lehrkräfte unter Druck gesetzt, unter dem dann korrekte Antworten gegeben werden mussten. Wie die Facetten Korrektheit und Geschwindigkeit zusammengeführt werden, ist eine in der Literatur häufig diskutierte Frage bei Geschwindigkeitstests, die allerdings auch anders beantwortet werden kann, indem zum Beispiel die Zeitdauer bis zur Eingabe der Antwort gemessen wird (Phillips und Rabbitt 1995; Arthur

et al. 2003; Krauss und Brunner 2011). Wir haben uns aus konzeptionellen Gründen für obigen Zugang entschieden, da er nah an der unterrichtlichen Realität liegt.

#### 4.2.4 Durchführung der Studie

Alle TEDS-FU-Instrumente wurden online präsentiert. Angesichts der Verteilung der Lehrkräfte über ganz Deutschland mit in der Regel nur einzelnen Lehrpersonen an jeder Schule war dies der einzig gangbare Weg, die Studie durchzuführen. Der Vorteil dieses Vorgehens war, dass die Studie unter forschungsökonomischen Gesichtspunkten kostengünstiger durchgeführt werden konnte als TEDS-M und die Daten zudem zeitnah zur Verfügung standen (OECD 2010). Die video-basierte Testung online durchzuführen hat sich zwischenzeitlich auch durchgesetzt (siehe alle oben dargestellten Studien). Die Instrumente konnten nur am Rechner bearbeitet werden. Postsendungen waren nicht zugelassen. Die einzelnen Item-Blöcke waren zeitlich beschränkt, um eine Kontrolle darüber zu haben, dass zur Beantwortung der Fragen keine Lehrtexte oder Nachschlagewerke eingesetzt wurden.

Damit Müdigkeit nicht zu künstlich erhöhten Item-Schwierigkeiten führt, wurde ein rotiertes Design mit je zwei Item-Blöcken innerhalb der Leistungstests und eine Rotation der drei Videotests verwendet. Angesichts der Neuartigkeit der Instrumente haben wir darauf verzichtet, ein Matrixdesign zu verwenden.

#### 4.3 Skalierung der Daten

Alle Indikatoren wurden mit Hilfe des Raschmodells skaliert. Um eine direkte Vergleichbarkeit mit den TEDS-M-Skalenwerten herzustellen, wurden die beiden fachbezogenen Facetten des Professionswissens zu Beginn des Berufslebens analog zum Vorgehen in TEDS-M skaliert. Hierzu wurden die Item-Parameter aus TEDS-M importiert. Als Personenparameter wurden wie vier Jahre zuvor *Maximum Likelihood Estimates* (MLE) geschätzt. Ferner wurden erneut sowohl übersprungene als auch nicht erreichte Items als falsch kodiert. Für die Skalierung der beiden anderen Indikatoren wurden übersprungene Items als falsch und nicht – wie mittlerweile üblich – erreichte Items als fehlende Werte kodiert. Als Personenparameter wurden *Weighted likelihood estimates* (WLE; Warm 1989) geschätzt.

Die Reliabilität aller vier Skalen war zufriedenstellend bis gut:  $MLE_{MCK} = 0,73$ ,  $MLE_{MPCK} = 0,76$ ,  $WLE_{MSpeed} = 0,73$  und  $WLE_{MPerc} = 0,67$ . Für die Skalierung wurde das Software-Paket *Conquest* (Wu et al. 1997) genutzt. Die Struktur der Skalen – zum einen im Hinblick auf ihre Raschhomogenität und zum anderen auf ihre latenten Zusammenhänge – ist in anderen Zusammenhängen geprüft und bestätigt worden (Blömeke et al. [wiedereingereicht](#)). Aus Tab. 2 können die manifesten Interkorrelationen der Skalenwerte entnommen werden.

#### 4.4 Datenanalysen

Die aufgestellten Hypothesen wurden in Form von Pfadmodellen mit Hilfe des Software-Pakets MPlus (Version 5.1) und der gesamten Stichprobe von  $n=171$  Mathematiklehrkräften geprüft. Zur Schätzung der Parameter wurden robuste

**Tab. 2** Manifeste Interkorrelationen der Skalenwerte (Pearsons  $r$ )

	MCK <sub>t1</sub>	MCK <sub>t2</sub>	MPCK <sub>t1</sub>	MPCK <sub>t2</sub>	MSpeed
MCK <sub>t2</sub>	0,67***	–	–	–	–
MPCK <sub>t1</sub>	0,57***	0,35***	–	–	–
MPCK <sub>t2</sub>	0,50***	0,71***	0,35***	–	–
MSpeed	0,44***	0,46***	0,45***	0,28***	–
M_PID	0,14 (ns)	0,33***	0,05 (ns)	0,39***	0,06 (ns)

*MCK* mathematisches Wissen (*mathematics content knowledge*), *MPCK* mathematikdidaktisches Wissen (*mathematics pedagogical content knowledge*), *MSpeed* schnelles Erkennen von Schülerfehlern (*speed in recognizing student errors*), *M\_PID* situationsbezogene Wahrnehmung, Interpretation und Entscheidung über Handlungsoptionen in Bezug auf Mathematikunterricht (*perception, interpretation and decision-making with respect to mathematics instruction*). *t1* erster Messzeitpunkt im letzten Jahr der Lehrerausbildung 2008 im Rahmen von TEDS-M, *t2* zweiter Messzeitpunkt vier Jahre später im Rahmen von TEDS-FU, *ns* nicht signifikant, \*\*\* $p < 0,001$

Maximum-Likelihood-Algorithmen verwendet (MLR), sodass die vorliegenden nicht-normalen Verteilungen und fehlende Werte gehandhabt werden konnten. Zur Berücksichtigung Letzterer kam das Full-Information-Maximum-Likelihood-Verfahren zum Einsatz. Die Standardfehler wurden mit dem in MPlus implementierten Sandwich-Schätzer ermittelt.

Die Pfadmodelle wurden mit den manifesten Werten spezifiziert. Eine latente Modellierung wäre in unserem Kontext nur möglich gewesen, wenn für jedes Konstrukt IRT-Modelle integriert würden, was in Anbetracht der Itemzahlen in Relation zum Stichprobenumfang nicht machbar ist. Die manifeste Modellierung führt auch nur dazu, dass die Zusammenhänge zwischen den Konstrukten unterschätzt werden, d. h. es handelt sich angesichts unserer Forschungsfragen um ein konservatives Vorgehen.

Die Modellanpassung wurde mit Hilfe von komparativen und absoluten Fit-Indices evaluiert. Werte des komparativen Fit-Index CFI über .95 zeigen eine sehr gute Passung an die Daten an (Fan et al. 1999; Hu und Bentler 1999). Dies gilt auch für Werte des globalen Fit-Index (Root Mean Square Error of Approximation, RMSEA) unter 0,05.

Für den Mittelwertvergleich von MCK und MPCK wurden Varianzanalysen mit Messwiederholungen durchgeführt ( $n=108$ ). Da fehlende Werte in dieser – anders als in Pfadmodellen – nicht kompensiert werden können, können in Varianzanalysen nur jene Personen eingehen, die zu allen Messzeitpunkten und auf allen Indikatoren vollständige Werte aufweisen.

## 5 Ergebnisse

### 5.1 Ergebnisse zur Veränderung des fachbezogenen Wissens (H1a, b)

Varianzanalysen für Messwiederholungen stützen beide Hypothesen (H1a, b) zur Veränderung des mathematischen Wissens (MCK) und des mathematikdidaktischen Wissens (MPCK) im Übergang von der Lehrerausbildung in den Lehrerberuf. Das MCK der Junglehrkräfte geht in den vier Jahren zwischen dem Ende der Ausbildung

**Tab. 3** Varianzaufklärung und Modellanpassung für alle Modelle

Konzept	Modell	Varianzaufklärung		Modellanpassung				
		R <sup>2</sup> MSpeed	R <sup>2</sup> M_PID	CFI	RMSEA	χ <sup>2</sup>	df	χ <sup>2</sup> /df
Einfache Kompetenzmodelle (H2a)	MCK <sub>t1</sub> → verhaltensnahe Indikatoren <sub>t2</sub>	0,18**	0,08 (ns)	0,96	0,07	1,92	1	1,9 (ns)
	MPCK <sub>t1</sub> → verhaltensnahe Indikatoren <sub>t2</sub>	0,18**	0,01 (ns)	0,91	0,11	3,09	1	3,1 <sup>†</sup>
Differenzierte Expertisemodelle (H2b)	MCK <sub>t1</sub> → MCK <sub>t2</sub> → verhaltensnahe Indikatoren <sub>t2</sub>	0,22**	0,16**	0,99	0,04	3,88	3	1,3 (ns)
	MPCK <sub>t1</sub> → MPCK <sub>t2</sub> → verhaltensnahe Indikatoren <sub>t2</sub>	0,09 (ns)	0,22*	0,65	0,17	17,55	3	5,8***
MPCK-Endmodell		0,17**	0,22**	1,00	0,00	2,87	3	1,0 (ns)
Ausbildung vs. Berufseinstieg	MCK <sub>t1</sub> + Veränderung <sub>t2</sub> → verhaltensnahe Indikatoren	0,27***	0,17*	1,00	0,00	0,03	1	0,03 (ns)
	MPCK <sub>t1</sub> + Veränderung <sub>t1,t2</sub> → verhaltensnahe Indikatoren <sub>t2</sub>	0,22**	0,19**	1,00	0,00	0,93	1	0,93 (ns)

R<sup>2</sup> Varianzaufklärung, CFI Comparative Fit Index, RMSEA Root Mean Square Error of Approximation, Chi-Quadrat-Wert, Quotient aus Chi-Quadrat-Wert und Freiheitsgraden, \*\*\**p*<0,001, \*\**p*<0,01, \**p*<0,05, †*p*<0,1

und der Testung im Rahmen von TEDS-FU signifikant zurück ( $F_{(1,107)}=4,1, p<0,05$ ). Die Effektstärke ist allerdings eher gering ( $\eta^2=0,04$ ) und der Rückgang damit nur von mäßiger praktischer Bedeutsamkeit. Zudem ist zu beachten, dass bei positiv selektierten Gruppen aufgrund des Phänomens der Regression zur Mitte bei einer nochmaligen Testung zu erwarten ist, dass die Messwerte dichter an den Mittelwert der Skala heranrücken.

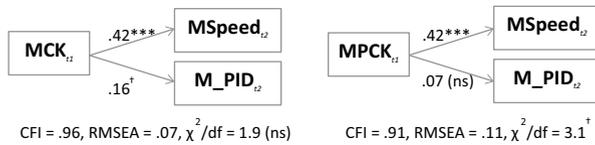
Für MPCK zeigt sich – wie angenommen – ein anderer Effekt. Das MPCK der Lehrkräfte bleibt ungefähr auf einem ähnlichen Leistungsniveau erhalten. Nominell ist sogar ein Zuwachs von fünf Testpunkten in den vier Jahren zwischen TEDS-M und TEDS-FU zu verzeichnen, der allerdings nicht statistisch signifikant wird.

### 5.2 Ergebnisse zur Prüfung der einfachen Kompetenzmodelle (H2a)

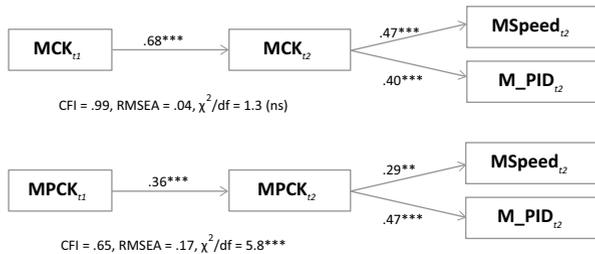
Wie erwartet ist die direkte Wirkung des am Ende der Ausbildung vorliegenden Wissens (t1) auf situationsbezogene und verhaltensnahe Merkmale von Mathematiklehrkräften zu t2 zwar signifikant bzw. marginal signifikant, in jedem Falle aber von geringer Effektstärke sowie unterschiedlich je nachdem, welches Kriterium und welchen Prädiktor man wählt.

MCK 2008 klärt signifikant Varianz im Geschwindigkeitsindikator MSpeed auf, ist also relevant dafür, wie gut Mathematiklehrkräfte vier Jahre später im Jahr 2012 unter Zeitdruck Schülerfehler erkennen können (siehe Tab. 3, Zeile „einfache Kompetenzmodelle – MCK“, Spalte „R<sup>2c</sup>“). Dabei ist in Einklang mit den COACTIV-Ergebnissen die praktische Bedeutsamkeit des Effekts gegeben (siehe Abb. 8, linkes

**Abb. 8** Pfadmodelle zur Prüfung von direkten Wirkungen von Wissen auf situationsbezogene und verhaltensnahe Merkmale (*links* MCK, *rechts* MPCK)



**Abb. 9** Pfadmodelle zur Prüfung der differenzierten Expertise-Modelle (oben MCK, unten MPCK)



Modell:  $\beta=0,42^{***}$ ). Für die Wahrnehmung und Interpretation von Mathematikunterricht sowie die Reaktion darauf (M\_PID) wird das am Ende der Ausbildung vorliegende mathematische Wissen dagegen nur marginal signifikant und die praktische Bedeutsamkeit des Effekts ist gering ( $\beta=0,16^\dagger$ ). Eine signifikante Varianzaufklärung von M\_PID durch MCK erfolgt nicht (Tab. 3, Spalte „ $R^{2c}$ “). Entsprechend zeigt das MCK-Modell zwar eine gute, aber keine sehr gute Anpassung an die Daten (Spalten „CFI“ bis „ $\chi^2/df^c$ “).

Dass ein einfaches Kompetenzmodell dem Verhältnis von Lehrerwissen am Ende der Ausbildung und handlungsnahen Kognitionen vier Jahre später nur begrenzt gerecht wird, zeigt sich insbesondere und wie erwartet in Bezug auf das mathematikdidaktische Wissen der Lehrkräfte. Das in Abb. 8, rechte Seite, dokumentierte MPCK-Modell zeigt nur eine befriedigende Anpassung an die Daten (siehe Tab. 3, Zeile „einfache Kompetenzmodelle – MPCK“, Spalten „CFI“ bis „ $\chi^2/df^c$ “). Die Varianz in MSpeed wird mit diesem Modell zwar signifikant und praktisch bedeutsam aufgeklärt (Spalte „ $R^{2c}$ “), sodass man schlussfolgern kann, dass auch mathematikdidaktisches Wissen dafür wichtig ist, wie schnell Lehrkräfte Schülerfehler erkennen. Keine Bedeutsamkeit kommt MPCK aber für den verhaltensnahen Indikator M\_PID zu, dessen Varianz nicht signifikant aufgeklärt wird (Spalte „ $R^{2c}$ “).

### 5.3 Ergebnisse zur Prüfung der differenzierten Expertise-Modelle (H2b)

Im Unterschied zu diesen zwar erwarteten, wegen ihrer geringen Erklärungskraft aber dennoch unbefriedigenden Ergebnissen der einfachen Modelle, stützen die Daten deutlich die Annahmen der differenzierten Expertise-Modelle. Wie erwartet sind auch in diesen differenzielle Effekte zu beobachten je nachdem, welches Kriterium und welcher Prädiktor herangezogen wird.

Das mathematische Wissen zum zweiten Messzeitpunkt wird hochsignifikant durch das am Ende der Ausbildung vorliegende MCK vorhergesagt (siehe Abb. 9, oberes Modell). Der Effekt fällt mit  $\beta=0,68^{***}$  stark aus und ist damit von sehr großer praktischer Bedeutsamkeit, da er die Rangreihe der Mathematiklehrkräfte als vergleichsweise stabil ausweist: Wer am Ende der Mathematiklehrerausbildung über mehr mathematisches Wissen verfügt, schneidet mit hoher Wahrscheinlichkeit auch

vier Jahre später besser in unserem Test ab. Das zum zweiten Messzeitpunkt vorliegende MCK sagt wiederum signifikant und mit ebenfalls deutlichem Effekt die Fähigkeit voraus, schnell Schülerfehler zu entdecken (MSpeed:  $\beta=0,47^{***}$ ) sowie das Geschehen in der Klasse wahrzunehmen, zu interpretieren und Handlungsoptionen zu generieren (M\_PID:  $\beta=0,40^{***}$ ). In diesem Modell wird erstmals auch signifikant Varianz in *beiden* Kriterien aufgeklärt und nicht nur im Geschwindigkeitsindikator (siehe Tab. 3, Zeile „differenzierte Expertisemodelle – MCK“, Spalten „R<sup>2</sup> MSpeed“ und „R<sup>2</sup> M\_PID“).

Für beide Indikatoren von Lehrere Expertise ist MCK damit praktisch bedeutsam. Die Anpassung dieses differenzierten MCK-Modells an die Daten fällt entsprechend sehr gut aus (Spalten „CFI“ bis „ $\chi^2/df$ “). Unsere Hypothese H2b kann in Bezug auf MCK damit als durch die Daten gestützt angesehen werden. Das Verhältnis von mathematischem Wissen am Ende der Ausbildung und situationsbezogenen Kognitionen vier Jahre später lässt sich zwar direkt modellieren. Die Aussagekraft ist aber größer, wenn man die Entwicklung des mathematischen Wissensniveaus beim Übergang in den Beruf einbezieht und damit ein indirektes Verhältnis annimmt, das über diesen Prozess vermittelt wird.

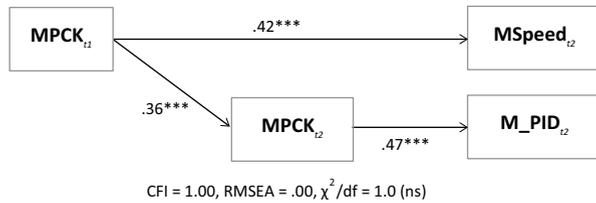
Wie im Falle von MCK wird das MPCK zum zweiten Messzeitpunkt durch das in TEDS-M am Ende der Ausbildung erhobene mathematikdidaktische Wissen signifikant vorhergesagt (Abb. 9, unteres Modell). Der Effekt von  $\beta=0,36^{***}$  ist von praktischer Bedeutsamkeit, fällt allerdings entsprechend des Signifikanztests nach Meng et al. (1992) geringer aus als im Falle von MCK. Dieses Ergebnis deutet – nicht überraschend – daraufhin, dass bei MPCK durch die Konfrontation mit der Schulpraxis stärkere Veränderungen erfolgen als bei MCK. Die Rangreihe der Mathematiklehrkräfte am Ende der Ausbildung und vier Jahre später ist im Falle des mathematikdidaktischen Wissens deutlich weniger stabil als im Falle des mathematischen Wissens.

Die Wahrnehmung, Analyse und Reaktion auf die gezeigten Unterrichtssituationen wird dann wie erwartet signifikant durch das zum zweiten Messzeitpunkt vorliegende MPCK vorhergesagt, wobei der Effekt mit  $\beta=0,47^{***}$  praktisch bedeutsam ist und ähnlich hoch ausfällt wie für MCK. Wie rasch Mathematiklehrkräfte Schülerfehler erkennen können, hängt wie erwartet ebenfalls signifikant vom MPCK zum zweiten Messzeitpunkt ab. Der Effekt ist in diesem Falle mit  $\beta=0,29^{**}$  aber signifikant schwächer als für M\_PID und auch als für den Prädiktor MCK (Meng et al. 1992).

Obwohl die Hypothese H2b durch diese Ergebnisse auch für MPCK gestützt wird, macht die schlechte Passung des MPCK-Gesamtmodells an die Daten (siehe Tab. 3, Zeile „differenzierte Expertisemodelle – MPCK“, Spalten „CFI“ bis „ $\chi^2/df$ “) darauf aufmerksam, dass es sich noch nicht um das optimale Modell handelt. Auch wird nur Varianz im Wahrnehmungsindikator signifikant aufgeklärt (Spalte „R<sup>2</sup> M\_PID“), aber nicht im Geschwindigkeitsindikator (Spalte „R<sup>2</sup> MSpeed“). Letzteres stellte sich im Modell, das H2a zugrunde lag, anders dar, sodass sich die Frage stellt, ob nicht für MPCK eine Kombination der beiden Modelle das Optimalmodell darstellt.

In einem solchen, ergänzend geschätzten Modell (siehe Abb. 10) zeigt sich dann tatsächlich, dass das differenzierte Expertise-Modell zwar für M\_PID gilt, dass für MSpeed die Annahme des einfachen Kompetenzmodells aber hinreichend ist. Für die Geschwindigkeit, mit der Schülerfehler entdeckt werden, besteht ein direkter Effekt des mathematikdidaktischen Wissens, das am Ende der Lehrerausbildung vorliegt,

**Abb. 10** Differenziertes Expertise-Modell für M\_PID und einfaches Kompetenzmodell für MSpeed (nur MCPK)



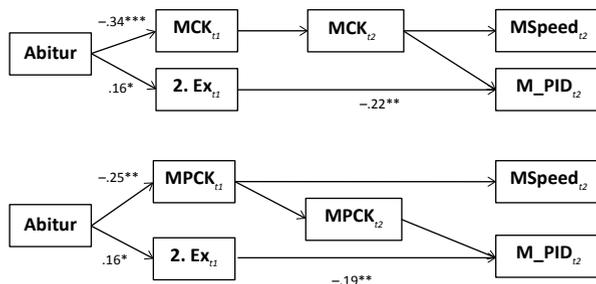
der von hoher praktischer Bedeutsamkeit ist ( $\beta = 0,42^{***}$ ). Dieser direkte Effekt deutet entweder auf spezifische Lerngelegenheiten in Bezug auf typische Schülerfehler hin, die in der mathematikdidaktischen Lehrerbildung bestanden haben, auf eine Kovariation des mathematikdidaktischen mit dem mathematischen Wissen, sodass sich hier indirekt die Nähe von MSpeed zu MCK spiegelt, oder auf eine Differenzierung des mathematikdidaktischen Wissens in verschiedene Wissensfacetten, wobei hier insbesondere eine stoffbezogene mathematikdidaktische Wissensdimension zum Tragen kommen dürfte (zur Differenzierung der verschiedenen Wissensdimensionen mathematikdidaktischen Wissens siehe Buchholtz et al. 2014). Das in Abb. 10 dokumentierte endgültige MPCK-Modell zeigt eine sehr gute Anpassung an die Daten (siehe Tab. 3, Zeile „MPCK-Endmodell“, Spalten „CFI“ bis „ $\chi^2/df$ “).

5.4 Ergebnisse zur Validierung der Modelle mithilfe von Außenkriterien

Damit bleibt die Aufgabe der Validierung des mathematischen und des mathematikdidaktischen Expertise-Modells, um zu klären, ob die von uns gewählten Indikatoren in ihrer methodischen Umsetzung als Papier-und-Bleistift-, Video-Vignetten- bzw. Geschwindigkeitstests sowie die daraus abgeleiteten Modelle tragen. Hierfür werden diese mit Außenkriterien verknüpft (siehe Abb. 1), die aus einem nomologischen Netzwerk abgeleitet, also aufgrund des Forschungsstands in einen überzeugenden Zusammenhang gebracht werden.

Der Forschungsstand lässt eine signifikante Vorhersage von MCK und MPCK durch die Abiturnote sowie von M\_PID durch das Ergebnis im Zweiten Staatsexamen erwarten. Diese Zusammenhangsstruktur wird in der Tat durch unsere Daten gestützt. Der Pfadkoeffizient von der Note des Zweiten Staatsexamens zu M\_PID ist in beiden Modellen signifikant, liegt – angesichts der Breite der mit dieser Note erfassten Merkmale und der heterogenen Standards über die verschiedenen Bundesländer und Studienseminare hinweg nicht erstaunlich – aber nur im unteren Bereich ( $\beta = -0,22^{**}$  im MCK- bzw.  $\beta = -0,19^*$  im MPCK-Modell).

**Abb. 11** Validierung der Expertise-Modelle anhand von Außenkriterien (MCK oben, MPCK unten)



Wie aus vielen Studien bekannt, fällt der Koeffizient des Pfades von der Abiturnote auf MCK trotz des größeren zeitlichen Abstands höher aus ( $\beta = -0,34^{***}$ ). Hier macht sich vermutlich die Kumulativität von Bildungsgängen bemerkbar. MPCK wird ebenfalls signifikant, aber als im Lehramt neu einsetzender tertiärer Bildungsbereich entsprechend weniger stark als MCK durch die Abiturnote vorhergesagt ( $\beta = -0,25^{**}$ ). Die Note des Zweiten Staatsexamens wird, wie zu erwarten war, ebenfalls signifikant durch die Abiturnote vorhergesagt werden ( $\beta = 0,16^*$  in beiden Modellen).

Das in Abb. 11 dokumentierte MCK-Modell passt sehr gut auf die Daten und kann damit als Validitätsbeleg herangezogen werden: CFI=0,99, RMSEA=0,02,  $\chi^2/df=1,1$  (ns). Selbiges gilt für das MPCK-Modell, dessen Passung an die Daten gut ist: CFI=0,93, RMSEA=0,05,  $\chi^2/df=1,5$  (ns).

## 6 Diskussion

### 6.1 Zusammenfassung und Einordnung der Ergebnisse

Ziel des vorliegenden Beitrags war, die Entwicklung des am Ende der Sekundarstufen-I-Lehrerbildung vorliegenden mathematischen und mathematikdidaktischen Wissens – als wichtigen kognitiven Facetten professioneller Kompetenz (Weinert 1999) – während der ersten Berufsjahre und seine Vorhersagekraft für situations- und verhaltensnahe Indikatoren zu untersuchen, die auf das Vorliegen von Lehrerexpertise schließen lassen (Bromme 1992; Chi 2011).

In Bezug auf das mittlere Wissensniveau unserer Stichprobe kann festgehalten werden, dass das MCK der Junglehrkräfte in den vier Jahren zwischen der Testung 2008 am Ende der Ausbildung im Rahmen von TEDS-M und der Testung 2012 im Rahmen von TEDS-FU signifikant zurückgegangen ist. Allerdings ist dieser Rückgang praktisch nicht bedeutsam. Ihr mathematikdidaktisches Wissen bleibt auf etwa demselben Leistungsniveau erhalten. Dieses letzte Ergebnis ist bemerkenswerter als das erste, da aufgrund der bei Messwiederholungen typischerweise auftretenden Regression zur Mitte in unserer positiv selektierten Stichprobe ebenfalls eher mit einem Rückgang hätte gerechnet werden müssen. Insofern stellt dieses Ergebnis ein erstes Indiz für die Bedeutsamkeit der Praxiserfahrung als Lerngelegenheit für die Weiterentwicklung von MPCK hin, wie sie von der Expertiseforschung bereits seit langem mit Blick auf andere Berufe postuliert worden ist (Gobet 2005).

Diese Perspektive wird nochmals gestärkt, wenn man die Veränderungen in der leistungsbezogenen Rangreihe der Mathematiklehrkräfte in den vier Jahren zwischen TEDS-M und TEDS-FU betrachtet. Für MCK gilt, dass diese praktisch unverändert bleibt, da das mathematische Wissen zum zweiten Messzeitpunkt mit hoher Effektstärke durch das am Ende der Ausbildung vorliegende Wissen vorhergesagt wird. Wer am Ende der Mathematiklehrerbildung über mehr MCK verfügte, schneidet in unserem Test auch vier Jahre später besser ab. Anders stellt sich die Situation in Bezug auf das MPCK dar. Zwar wird – wie im Falle von MCK – auch das MPCK zum zweiten Messzeitpunkt durch das in TEDS-M am Ende der Ausbildung erhobene mathematikdidaktische Wissen signifikant vorhergesagt. Der Effekt fällt

allerdings signifikant geringer aus als im Falle von MCK, womit die Rangreihe der Mathematiklehrkräfte in dieser Wissensfacette deutlich weniger stabil ist.

In Übereinstimmung mit der bislang vor allem auf andere Berufe ausgerichteten und auf qualitativen Einzelfallstudien beruhenden Expertiseforschung können wir für Mathematiklehrkräfte und basierend auf standardisiert gewonnenen Daten daher in aller Vorsicht schlussfolgern, dass deren mathematische und mathematikdidaktische Wissensbasis zu Beginn ihres Berufslebens flexibel ist und die Praxiserfahrung eine Lerngelegenheit darstellt, die Einfluss auf die Weiterentwicklung der beiden Wissensfacetten nimmt. Dabei deuten unsere Ergebnisse daraufhin, dass – nicht überraschend, wenn man die Natur der beiden Facetten betrachtet (Buchholtz et al., 2014) – durch die Konfrontation mit der Schulpraxis bei MPCK individuell stärkere Veränderungen erfolgen als bei MCK.

Diese Veränderungen im mathematikdidaktischen Wissen können zum einen durch Unterschiede in der beruflichen Praxis bedingt sein, mit der die Berufsanfänger an ihren Schulen konfrontiert sind und die damit unterschiedliche Lerngelegenheiten bieten. Möglicherweise ist hierfür aber auch der unterschiedliche Umgang der Mathematiklehrkräfte mit der täglichen Unterrichtspraxis ausschlaggebend. Aus der Expertiseforschung liegen auch Hinweise dazu vor, dass für Expertise nicht allein der Umfang oder die Qualität an praktischer Erfahrung entscheidend sind, sondern dass der Grad an „deliberate practice“ (Ericsson et al. 1993) wichtig ist. Erst die Kombination von umfangreichen Praxiserfahrungen mit einem *zielgerichteten, intensiven und angeleiteten* Bemühen um Verbesserungen führt zu Expertise.

Beide Erklärungsansätze stellen plausible Interpretationen unserer Ergebnisse dar. Diese bedeuten auch, dass die Entwicklung der Lehrkräfte durch eine entsprechend gestaltete Praxiserfahrung gefördert werden kann. Auch wenn hier noch viel Forschungsbedarf auszumachen ist, wie die Gestaltung konkret aussehen sollte, können Lehrerbildner, Kultusministerien sowie Schulleitungen aus unseren Ergebnissen die Anregung ziehen, dass es sich um einen kontinuierlichen und unterstützten Reflexionsprozess handeln sollte, der intensiv an den individuellen Stärken und Schwächen der Berufseinsteiger ansetzt und der notgedrungen zeitlichen Freiraum benötigt. Es ist schwer vorstellbar, dass eine reflektierte Praxis in den ersten Berufsjahren neben der vollen Unterrichtsbelastung entwickelt werden kann.

In Bezug auf die Klärung des Zusammenhangs von Wissen und verhaltensnahen Kognitionen konnten wir zeigen, dass – wie erwartet – ein einfaches Kompetenzmodell nur begrenzt zutrifft. Zwar erweisen sich mathematisches und mathematikdidaktisches Wissen als direkt relevant dafür, wie gut Mathematiklehrkräfte unter Zeitdruck Schülerfehler erkennen können. Hierbei handelt es sich um eine wichtige Anforderung, mit denen Lehrkräfte im Mathematikunterricht konfrontiert sind. Das Ergebnis spricht also dafür, eine solide Wissensbasis in der Ausbildung aufzubauen (Bromme 1992).

Für die Wahrnehmung und Interpretation von Mathematikunterricht sowie die Reaktion auf Unterrichtssituationen (M\_PID) sind MCK und MPCK direkt dagegen nur wenig bis gar nicht bedeutsam. Erst eine komplexere Modellierung, die entsprechend des P-I-D-Modells (Blömeke, Gustafsson und Shavelson *im Druck*) differenziert die oben angesprochene Veränderung im Wissensniveau während des Berufseinstiegs aufgreift und die vermutlich ein erstes Abbild der kognitiven

Umstrukturierungsprozesse darstellt, die stattgefunden haben, führt zu einer sehr guten Erklärung des Verhältnisses. Diesem differenzierten Expertisemodell zufolge sagen das in TEDS-FU erfasste MCK und MPCK signifikant die Fähigkeit voraus, schnell Schülerfehler zu entdecken sowie das Geschehen in der Klasse präzise wahrzunehmen, zu interpretieren, Probleme zu antizipieren und Handlungsoptionen zu generieren. Dabei hängt das nach drei Jahren Berufspraxis entwickelte Wissensniveau in Bezug auf MCK als auch auf MPCK substanziell vom Wissensniveau am Ende der Lehrerausbildung ab.

Während für M\_PID beide Wissensfacetten von hoher Bedeutung sind, ist für das schnelle Erkennen von Schülerfehlern das mathematische Wissen relevanter als das mathematikdidaktische. Einmal mehr zeigt sich hier der angesprochene konzeptionelle Unterschied in der Natur der beiden fachbezogenen Wissensfacetten (Buchholtz et al., 2014). Und einmal mehr wird deutlich, wie zentral es ist, in der Ausbildung eine solide Wissensbasis zu legen, da sie erst die Weiterentwicklung professioneller Kompetenz in der Praxis ermöglicht. Für das Erzielen einer hohen Unterrichtsqualität und letztlich höherer Schülerleistungen stellt diese insofern eine Schlüsselstelle dar, wie Studien zum Zusammenhang der verschiedenen Kompetenz- und Expertise-Indikatoren mit diesen zeigen (Kersting et al. 2010).

Für MSpeed ergeben sich dabei geringere Änderungen im Anteil der erklärten Varianz im Vergleich zwischen einfachem und komplexem Expertisemodell. Dies deutet darauf hin, dass MSpeed zwar situationsbezogene Elemente aufweist, offensichtlich aber stark wissensbasiert ist. Insofern ermöglicht bei MSpeed bereits die Ausgangslage am Ende der Lehrerbildung eine gute Vorhersage, ob eine Mathematiklehrkraft in der Lage sein wird, schnell Schülerfehler zu erkennen. Bei M\_PID sind die Unterschiede zwischen dem einfachen und dem komplexen Expertisemodell deutlich größer; hier ist also die Veränderung durch die in der Praxis gewonnene Berufserfahrung von höherer Relevanz. Konzeptionell lässt sich dennoch rechtfertigen, die von uns vorgelegte Operationalisierung des Geschwindigkeitsindikators weiterhin als Teil von Expertise zu betrachten. Das schnelle Erkennen von Schülerfehlern ist eine zentrale Anforderung der unterrichtlichen Praxis an Lehrkräfte. Im Unterschied zu COACTIV (Krauss und Brunner 2011) konzentrieren wir uns dabei explizit auf das Erkennen *typischer Schülerfehler*, was letztlich eine Sache der Berufserfahrung ist, und nicht auf das schnelle Erkennen einer richtigen Lösung, was stärker als Wissenselement betrachtet werden kann.

## 6.2 Limitierungen des methodischen Vorgehens

Die externe Validität unserer Maße konnte über eine Vorhersage von MCK und MPCK durch die Abiturnote sowie von M\_PID durch das Ergebnis im Zweiten Staatsexamen belegt werden. Unser Testformat reiht sich damit in die guten Validitätskennwerte für Situational Judgment Tests ein, obwohl unser Vorgehen deutlich offener ist als üblicherweise. Bevor wir zu Schlussfolgerungen für die weitere Forschung kommen, seien methodische Einschränkungen unserer Studie diskutiert. Obwohl uns eine Validierung gelungen ist, ist darauf hinzuweisen, dass die Stichprobe vergleichsweise homogen ist, da sie im Vergleich zur Gesamtgruppe an berufstätigen Lehrkräften nur Lehrkräfte mit annähernd demselben Umfang an Berufserfahrung umfasst.

Weissensteiner et al. (2008) machen anhand ihres Videotests im Bereich Sport darauf aufmerksam, dass die Homogenität von Probanden mit ähnlich langer Berufserfahrung deutlich größer ist als in den üblichen Experten-Novizen-Vergleichen. Allerdings macht diese Einschränkung das Aufdecken von Effekten eher schwieriger, sie wirkt also konservativ.

In Bezug auf die eher generischen Merkmale der Abiturnote und des Staatsexamens stellt unsere Stichprobe ein gutes Abbild der zufällig gezogenen repräsentativen TEDS-M-Stichprobe dar. In fachbezogener Hinsicht ist sie allerdings positiv selektiert, was erneut mit einer Varianzeinschränkung einhergeht (Williams und Ericsson 2008). Während die Mittelwerte der Zufallsstichprobe in TEDS-M für das mathematische Wissen bei 537 und für das mathematikdidaktische Wissen bei 549 Testpunkten lagen, liegen dieselben Werte für die TEDS-FU-Stichprobe bei 573 bzw. 567 Punkten. Diese Einschränkungen führen jedoch erneut eher zu einer konservativen Unterschätzung der Effekte als zu einer Überschätzung.

Als mögliche Kritik ist festzuhalten, dass die Videotestung in Form von Unterrichtssequenzen die Komplexität von Unterricht einschränkt. Der Unterrichtsgegenstand und Fokus im Klassenraum sind vorgegeben. Dies lässt sich notgedrungen kaum vermeiden, wenn große Stichproben standardisiert getestet werden sollen, und es ist zum Teil auch explizites Ziel solcher Assessment-Formate (wie zum Beispiel auch im Falle von Situational Judgment Tests), um forschungsökonomisch handhabbar zu bleiben. Solange die Validität hinreichend belegt werden kann, sehen wir diese Einschränkung als wenig gravierend an.

Problematischer ist in dieser Hinsicht eher, dass wir Lehrer-Schüler-Interaktionen ohne Bezug zum Unterrichtsgegenstand nicht aufgenommen haben, obwohl diese selbstverständlich häufig vorkommen. Weiterhin ist anzumerken, dass das schnelle und reziproke Wechselspiel zwischen Wahrnehmung, Interpretation, Handlung und erneuter Interpretation, welches charakteristisch für das Unterrichtshandeln ist, in Form von Videovignetten nicht erfassbar ist. Hier stellt sich eine Aufgabe für weitere Forschungen, noch realitätsnähere Instrumente zu entwickeln. Es lässt sich damit direkt der Bezug zur methodischen Diskussion um den Einsatz unterschiedlicher Assessmentformate und weitere Forschungsbedarfe herstellen.

### 6.3 Ausblick auf zukünftige Forschung

Die vorliegende Studie regt zahlreiche weiterführende Forschungsfragen an. Eine entscheidende Frage ist, welche Faktoren konkret auf die Veränderung zwischen den beiden Messzeitpunkten Einfluss genommen haben. Aufgrund erster Studien ist anzunehmen, dass für die Entwicklung in den ersten Berufsjahren die Überzeugungen der Lehrkräfte, ihre Berufsmotivation und Selbstregulationsfähigkeit sowie der Schulkontext eine Rolle spielen könnten (Blömeke und Klein 2013). Zudem stellt sich die Frage, welcher der von uns eingesetzten Kompetenz- und Expertise-Indikatoren näher an den Merkmalen von Unterrichtsqualität und Schülerleistungen liegt. Vermutet werden kann hierzu auf der Basis neuerer Studien, dass verhaltensnähere Indikatoren, wie sie in M\_PID abgebildet sind, stärkere Effekte zeigen als rein wissensbasierte Indikatoren wie MCK (siehe insbesondere Kersting et al. 2013).

Unter Generalisierungsgesichtspunkten wäre auf der einen Seite zu prüfen, inwieweit sich die hier dokumentierten Zusammenhänge zwischen Wissen und Expertise auf Lehrpersonen mit mehr Berufserfahrung übertragen lassen, wie sich die Expertise der Mathematiklehrkräfte also in den Folgejahren nach dem Berufseinstieg weiterentwickelt. Auf der anderen Seite stellt sich die Frage, wie die Zusammenhänge vor dem Abschluss der Mathematiklehrausbildung aussehen. TEDS-M setzte am Ende des Referendariats an. In diesem haben die angehenden Lehrkräfte bereits Praxiserfahrung sammeln können. Dagegen fanden keine Mathematikveranstaltungen mehr statt. Möglicherweise haben wichtige Entwicklungsprozesse also bereits zuvor stattgefunden, während im Laufe des späteren Berufslebens eher wenige Veränderungen stattfinden.

Entscheidend ist aus unserer Sicht, das Verhältnis von Wissen, Kompetenz, Expertise und unterrichtlicher Performanz mithilfe unterschiedlicher Instrumente weiter aufzuklären, wozu die vorliegende Studie nur einen kleinen Beitrag leisten konnte. Zwischen diesen Merkmalen finden zahlreiche vermittelnde Prozesse statt. Wass et al. (2001) und Albino et al. (2008) haben daher – Millers Kompetenzpyramide weiterführend – ein vierschrittiges Modell mit entsprechenden Assessmentformaten vorgeschlagen, das sequenziell Transformationsprozesse wie kognitive Umstrukturierungen und unterschiedliche Rahmenbedingungen berücksichtigt.

Miller (1990) hatte, aufbauend auf Blooms Taxonomie kognitiver Prozesse, dem Grad der Authentizität nach methodologisch zwischen vier Assessment-Formaten unterschieden, mit denen die Kognitionen von Medizinerinnen untersucht werden können: 1) Wissenstests, in denen die angehenden Mediziner das Vorhandensein von deklarativem Wissen (*factual knowledge*) demonstrieren (*know*); 2) Kompetenztests im engen Sinne, in denen das Vorhandensein prozeduralisierter anwendungsbezogener Wissens (*applied knowledge*) demonstriert werden muss (*know how*); 3) Performanztests, in denen die angemessene Umsetzung des Wissens situiert in repräsentativen berufsbezogenen Situationen (*performance assessments in vitro*) demonstriert werden muss (*show how*); und 4) handlungsbezogene Tests (*performance assessment in vivo*), in denen die angemessene Umsetzung des Wissens unter den Bedingungen des beruflichen Alltags gezeigt werden muss (*does*).

Unsere Papier-und-Bleistift-Tests erfassen wie Baer et al. (2007) oder Kleickmann et al. (2013) die ersten beiden Stufen, während wir mit dem Video-Vignettest als Assessment unter kontrollierten Bedingungen wie Seidel et al. (2010) oder Sherin (2007) einen Schritt näher an das tatsächlich zu erwartende Handeln gekommen sind. Nur wenige Studien existieren, in denen diese Maße zusammen und dann auch noch in Relation zum tatsächlichen Handeln im beruflichen Alltag untersucht werden. Für eine solche holistische Betrachtung dessen, was Lehrkräfte wirklich können, würden beispielsweise Videoaufnahmen nötig sein, die standardisiert analysiert würden (als Beispiel im Lehrerkontext siehe Vogelsang und Reinhold 2012). Hier stellen sich weitreichende Aufgaben, denen wir uns in zukünftigen Projekten stellen werden.

## Literatur

- Albino, J. E., Young, S. K., Neumann, L. M., Kramer, G. A., Andrieu, S. C., Henson, L., Horn, B., & Hendricson, W. D. (2008). Assessing dental students' competence: Best practice recommendations in the performance assessment literature and investigation of current practices in predoctoral dental education. *Journal of Dental Education*, 72, 1405–1435.
- Arthur, W., Doverspike, D., & Bell, S. T. (2003). Information processing tests. In M. Hersen & J. C. Thomas (Hrsg.), *The comprehensive handbook of psychological assessment: Vol. 4. Industrial/organizational assessment* (S. 56–74). New York: Wiley.
- Baer, M., Dörr, G., Fraefel, U., Kocher, M., Küster, O., Larcher, S., Müller, P., Sempert, W., & Wyss, C. (2007). Werden angehende Lehrpersonen durch das Studium kompetenter? *Unterrichtswissenschaft*, 35(1), 15–47.
- Baer, M., Kocher, M., Wyss, C., Guldemann, T., Larcher, S., & Dörr, G. (2011). Lehrerbildung und Praxiserfahrung im ersten Berufsjahr und ihre Wirkung auf die Unterrichtskompetenzen von Studierenden und jungen Lehrpersonen im Berufseinstieg. *Zeitschrift für Erziehungswissenschaft*, 14(1), 85–117.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Berliner, D. (1986). In pursuit of the expert pedagogue. *Educational Researcher*, 7, 5–13.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35, 463–482.
- Besser, M., & Krauss, S. (2009). Zur Professionalität als Expertise. In R. Mulder, O. Zlatkin-Troitschanskaja, K. Beck, R. Nickolaus, & D. Sembill (Hrsg.), *Professionalität von Lehrenden – Zum Stand der Forschung* (S. 71–82). Weinheim: Beltz.
- Blömeke, S. (2009). Ausbildungs- und Berufserfolg im Lehramtsstudium im Vergleich zum Diplomstudium: Zur prognostischen Validität kognitiver und psycho-motivationaler Auswahlkriterien. *Zeitschrift für Erziehungswissenschaft*, 12(1), 82–110.
- Blömeke, S., Busse, A., Kaiser, G., Klein, P., König, J., & Suhl, U. (wiederingereicht). On the nature of teacher expertise: Modeling the relationship of knowledge, perceptual accuracy and speed. *American educational Research Journal*.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (im Druck). Beyond dichotomies: Viewing competence as a continuum. *Zeitschrift für Psychologie*.
- Blömeke, S., Kaiser, G., & Döhrmann, M. (2011). Bedingungsfaktoren des fachbezogenen Kompetenzerwerbs von Lehrkräften: Zum Einfluss von Ausbildungs-, Persönlichkeits- und Kompositionsmerkmalen in der Mathematiklehrerausbildung für die Sekundarstufe I. *Zeitschrift für Pädagogik*, 57 (Beiheft), 77–103.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S., & Klein, P. (2013). When is a school environment perceived as supportive by beginning mathematics teachers? Effects of leadership, trust, autonomy and appraisal on teaching quality. *International Journal of Science and Mathematics Education*, 11, 1029–1048.
- Blömeke, S., Paine, L., Houang, R. T., Hsieh, F.-J., Schmidt, W. H., Tatto, M. T., Bankov, K., Cedillo, T., Cogan, L., Han, S. I., Santillan, M. N., & Schwillie, J. (2008). Future teachers' competence to plan a lesson. First results of a six-country study of the efficiency of teacher education. *ZDM – The International Journal on Mathematics Education*, 40, 749–762.
- Blömeke, S., Suhl, U., & Döhrmann, M. (2012). Zusammenfügen was zusammengehört: Kompetenzprofile am Ende der Lehrerausbildung im internationalen Vergleich. *Zeitschrift für Pädagogik*, 58(4), 422–440.
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instructed by expert and novice teachers. *American Educational Research Journal*, 26, 413–498.
- Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des Professionellen Wissens*. Bern: Huber.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule*. Göttingen: Hogrefe.
- Brouwer, C. N. (2010). Determining long term effects of teacher education. In P. Peterson, E. Baker, & B. McGaw (Hrsg.), *International encyclopedia of education* (7. Aufl. S. 503–510). Amsterdam: Elsevier.
- Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge: Harvard University Press.

- Buchholtz, N., Kaiser, G., & Blömeke, S. (2014). Die Erhebung mathematikdidaktischen Wissens: Konzeptualisierung einer komplexen Domäne. *Journal für Mathematikdidaktik*, 35, 101–128.
- Calderhead, J. (1984). *Teachers' classroom decision-making*. London: Holt.
- Calderhead, J. (1991). The nature and growth of knowledge in student teaching. *Teaching and Teacher Education*, 7, 531–535.
- Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. C. (1988). Expert-novice differences in perceiving and processing visual information. *Journal of Teacher Education*, 39, 25–31.
- Chi, M. T. H. (2011). Theoretical Perspectives, Methodological Approaches, and Trends in the Study of Expertise. In Y. Li & G. Kaiser (Hrsg.), *Expertise in mathematics instruction: An international perspective* (S. 17–39). Springer.
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In Sternberg, R. S. (Hrsg.), *Advances in the Psychology of Human Intelligence* (Bd. 1, S. 1–75). Hillsdale: Erlbaum.
- Clark, C., & Lampert M. (1986). The study of teacher thinking: Implications for teacher education. *Journal of Teacher Education*, 37, 27–31.
- Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen. *Unterrichtswissenschaft*, 31, 122–141.
- Crespo, S., Oslund, J. A., & Parks, A. N. (2011). Imagining mathematics teaching practice: prospective teachers generate representations of a class discussion. *ZDM – The International Journal on Mathematics Education*, 43, 119–131.
- Davison, M. L., Semmes, R., Huang, L., & Close, C. N. (2011). On the reliability and validity of a numerical reasoning speed dimension derived from response times collected in computerized testing. *Educational and Psychological Measurement*, 72, 245–263.
- 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 G. Greefrath, F. Käpnick, & M. Stein (Hrsg.), *Beiträge zum Mathematikunterricht*. Münster.
- Dunekacke, S., Jenßen, L. & Blömeke, S. (im Druck). Validierung eines Leistungstests zur Erfassung mathematikdidaktischer Kompetenz angehender frühpädagogischer Fachkräfte durch die videogestützte Erhebung von Performanz. *Zeitschrift für Pädagogik*, 61(Beiheft).
- Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- van Es, E. & Sherin, M. G. (2008). Mathematics teachers' „learning to notice“ in the context of a video club. *Teaching and Teacher Education*, 24, 244–276.
- Fan, X., Thompson, B., & Wang, L. (1999). The effects of sample size, estimation methods, and model specification on SEM fit indices. *Structural Equation Modeling*, 6, 56–83.
- Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. In L. Darling-Hammond (Hrsg.), *Review of Research in Education* (Bd. 20, S. 3–56). Washington
- Fischer, R. & Malle, G. (2004). *Mensch und Mathematik: Eine Einführung in didaktisches Denken und Handeln* (= Klagenfurter Beiträge zur Didaktik der Mathematik; Bd. 5). München: Profil.
- Fuller, F., & Brown, O. (1975). Becoming a teacher. In K. Ryan (Hrsg.), *Teacher Education (74th Yearbook of the National Society for the Study of Education; Bd. 2, S. 25–52)*. Chicago: University of Chicago Press.
- Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology*, 19, 183–204.
- Gold, B., Förster, St., & Holodyski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie*, 27, 141–155.
- Goodwin, Ch. (1994). Professional vision. *American Anthropologist*, 96, 606–633.
- Gruber, H. (1998). Expertise. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 25–52). Weinheim: Psychologie Verlags Union.
- Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität*. Seelze-Velber: Kallmeyer/ Klett.
- Herrmann, U., & Hertrampf, H. (2000). Der Berufsanfang des Lehrers – der Anfang von welchem Ende? *Die Deutsche Schule*, 92, 54–65.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144.

- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602–619.
- Kaiser, G., Benthien, J., Döhrmann, M., König, J., & Blömeke, S. (2013). Expert Ratings as an Instrument for Validating Results of Video-Based Testing. In A. Lindmeier & A. Heinze (Hrsg.). Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education. (Bd. 1). Kiel, Germany: PME.
- Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527–535.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement Spring*, 50, 1–73.
- Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement*, 68, 845–861.
- Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teacher's analysis of classroom video predicts student learning of mathematics: Further exploration of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61, 172–181.
- Kersting, N. B., Givvin, K. B., Thompson, B. W., Santagata, R., & Stigler, J. W. (2013). Measuring usable knowledge: Teachers' analyses of mathematics classroom Videos predict teaching quality and student learning. *American Educational Research Journal*, 49, 568–589.
- KHK (2004)=Kultusministerkonferenz. (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss, Beschluss vom 4.12.2003*. München: Luchterhand.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106.
- 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.
- Krauss, S., & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematikdidaktik*, 32, 233–251.
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78, 75–95.
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37, 426–441.
- Llinares, S., & Krainer, K. (2006). Mathematics (student) teachers and teacher educators as learners. In A. Gutiérrez & P. Boero (Hrsg.), *Handbook of Research on the psychology of mathematics education: Past, present and future* (S. 429–459). Rotterdam: Sense Publisher.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., & Campion, M. A. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730–740.
- Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172–175.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine. Journal of the Association of American Medical Colleges*, 65, 63–67.
- Müller-Fohrbrodt, G., Cloetta, B., & Dann, H. D. (1978). *Der Praxisschock bei jungen Lehrern*. Stuttgart: Klett.
- Nuissl, E., Schiersmann, Ch., & Siebert, H. (Hrsg.), (2002). *Kompetenzentwicklung statt Bildungsziele?* (= Literatur- und Forschungsreport Weiterbildung; 49).
- OECD (2010) = Organisation for Economic Cooperation and Development. (2010). *TALIS: Technical report*. Paris: OECD.
- Oser, F. K., & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Hrsg.), *Handbook of research on teaching* (4. Aufl., S. 1031–1065). Washington: American Educational Research Association.
- Oser, F., Heinzer, S., & Salzmann, P. (2010). Die Messung der qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. Chancen und Grenzen des advokatorischen Ansatzes. *Unterrichtswissenschaft*, 38(1), 5–28.
- Padberg, F. (1996). *Didaktik der Arithmetik*. Heidelberg: Spektrum.
- Pfadenhauer, M., & Kunz, A. M. (Hrsg.). *Kompetenzen in der Kompetenzerfassung: Ansätze und Auswirkungen der Vermessung von Bildung*. Weinheim: Beltz.

- Phillips, L. H., & Rabbitt, P. M. A. (1995). Impulsivity and speedaccuracy strategies in intelligence-test performance. *Intelligence*, *21*, 13–29.
- Rölke, H. (2012). The itembuilder: A graphical authoring system for complex item development. In Association for the Advancement of Computers in Education (AACE) (Hrsg.). *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (ELEARN)* (S. 344–353).
- de Saussure, F. (1916/1967). *Grundfragen der allgemeinen Sprachwissenschaft*. Berlin: De Gruyter.
- Schmitz, B. (2006). Advantages of studying processes in educational research. *Learning and Instruction*, *16*, 433–449.
- Schoenfeld, A. H., & Kilpatrick, J. (2008). Toward a Theory of Proficiency in Teaching Mathematics. In D. Tirosh & T. Wood (Hrsg.), *International handbook of mathematics teacher education. Bd. 2: Tools and processes in mathematics teacher education* (S. 321–354). Rotterdam: Sense Publishers.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. *Zeitschrift für Pädagogik*, *56*, 296–306.
- Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Hrsg.), *Video research in the learning sciences* (S. 383–395). Mahwah: Lawrence Erlbaum.
- Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Hrsg.) (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.
- Shulman, L. (1985). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (3. Aufl., S. 3–36). New York: Macmillan.
- Shulman, L. S. (1991). Von einer Sache etwas verstehen: Wissensentwicklung bei Lehrern. In Terhart, E. (Hrsg.), *Unterrichten als Beruf: Neuere amerikanische und englische Arbeiten zur Berufskultur und Berufsbiographie von Lehrern und Lehrerinnen* (= Studien und Dokumentationen zur vergleichenden Bildungsforschung, 50; S. 145–160). Köln und Wien: Böhlau.
- Stürmer, K., Könings, K. D., & Seidel, T. (2012). Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *British Journal of Educational Psychology*, *20*.
- Swanson, H. L., O'Connor, J. E., & Cooney, J. B. (1990). An information processing analysis of expert and novice teachers' problem solving. *American Educational Research Journal*, *27*, 533–556.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodriguez, M., & Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Findings from the IEA teacher and development study in mathematics (TEDS-M)*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Terhart, E. (1998). Lehrerberuf: Arbeitsplatz, Biographie, Profession. In *Handbuch der Schulentwicklung* (S. 560–585). Innsbruck: Österreichischer Schulbuch Verlag.
- Terhart, E. (2012). Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung*, *3*, 3–21.
- Ulm, V. (2004). *Mathematikunterricht für individuelle Lernwege öffnen*. Seelze: Kallmeyer.
- Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, *18*, 211–222.
- Vogelsang, C. & Reinhold, P. (2012). Gemessene Kompetenz und Unterrichtsqualität. In Höttecke, D. (Hrsg.), *Konzepte fachdidaktischer Strukturierung für den Unterricht: Gesellschaft für Didaktik der Chemie und Physik – Jahrestagung in Oldenburg 2011* (S. 286–288). Münster: LIT-Verlag.
- Wahl, D. (1991). *Handeln unter Druck. Der weite Weg vom Wissen zum Handeln bei Lehrern, Hochschullehrern und Erwachsenenbildnern*. Weinheim: Deutscher Studien Verlag.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*, 427–450.
- Wass, V., Van der Vlugten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *Lancet*, *357*, 945–949.
- Weinert, F. E. (1996). „Der gute Lehrer“, „die gute Lehrerin“ im Spiegel der Wissenschaft: Was macht Lehrende wirksam und was führt zu ihrer Wirksamkeit? *Beiträge zur Lehrerbildung*, *14*, 141–151.

- Weinert, F. E. (1999). *Konzepte der Kompetenz. Gutachten zum OECD-Projekt „Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo)“*. Neuchatel: Bundesamt für Statistik.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–66). Göttingen: Hogrefe.
- Weinert, F. E., Schrader, F.-W., & Helmke, A. (1990). Unterrichtsexpertise: Ein Konzept zur Verringerung der Kluft zwischen zwei theoretischen Paradigmen. In L.-M. Alisch, J. Baumert,, & K. Beck (Hrsg.), *Professionswissen und professionalisierung: sonderband in zusammenarbeit mit der zeitschrift empirische pädagogik* (= Braunschweiger Studien zur Erziehungs- und Sozialarbeit, 28; S. 173–206). Braunschweig: Copy-Center Colmsee.
- Weissensteiner, J., Abernethy, B., Farrow, D., & Müller, S. (2008). The development of anticipation: a cross-sectional examination of the practice experiences contributing to skill in cricket batting. *Journal of Sport and Exercise Psychology*, 30, 663–84.
- Williams, A. M. & Ericsson, K. A. (2008). From the guest editors: How do experts learn? *Journal of Sport and Exercise Psychology*, 30, 653–662.
- Wu, M. L., Adams, R. J., & Wilson, M. (1997). *ConQuest: Generalised item response modelling software*. Hawthorn: ACER.