

## Pädagogisches Wissen von angehenden Lehrkräften

### Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerausbildung

Johannes König · Sigrid Blömeke

**Zusammenfassung:** Während zur Testung des fachbezogenen Wissens von angehenden und praktizierenden Lehrkräften mittlerweile Erfahrungen vorliegen, stellt die standardisierte Erfassung ihres fachübergreifenden, pädagogischen Wissens ein neues Feld dar. Der Beitrag stellt vor, wie dieses Wissen unter Bezug auf Standards für die Lehrerausbildung sowie Erkenntnisse aus der Allgemeinen Didaktik und der Unterrichtsforschung definiert und erfasst werden kann. Dabei werden fünf berufliche Anforderungen fokussiert, mit denen sich Lehrpersonen beim Unterrichten konfrontiert sehen: Strukturierung von Unterricht, Motivierung, Umgang mit Heterogenität, Klassenführung und Leistungsbeurteilung. Anschließend erfolgt mittels Rasch-Modellen eine empirische Untersuchung der Struktur des fachübergreifenden Wissens anhand einer Studie mit 802 Lehramtsstudierenden. Erwartungsgemäß spiegelt das Modell, das zwischen den fünf anforderungsbezogenen Wissensgebieten unterscheidet, die Daten deutlich besser wider als ein Modell, das für das pädagogische Wissen eine homogene Struktur annimmt. Damit werden langjährige wissenschaftstheoretische Diskussionen zum pädagogischen Wissen empirisch fundiert. Zudem zeigen sich Zusammenhänge, die typische Lerngelegenheiten in der Lehrerausbildung widerspiegeln.

**Schlüsselwörter:** Pädagogisches Wissen · Lehrerausbildung · Lehrerkompetenzen · Bildungsstandards · Referendariat · Erziehungswissenschaftliches Studium · Mehrdimensionales Rasch-Modell · Leistungstest

### Pedagogic knowledge of future teachers – Capturing and structuring the results of cross-subject teacher education

**Abstract:** Whilst experience of testing subject-specific knowledge of prospective and practicing teachers is generally available, the standardized capture of their cross-subject, pedagogic knowledge remains a new field of inquiry. This contribution will present a way of defining and capturing this knowledge with reference to standards for teacher training and insights from general

---

**Online publiziert:** 07.10.2009  
© VS-Verlag 2009

Dr. J. König (✉) · Univ.-Prof. S. Blömeke  
Philosophische Fakultät IV, Abt. Systematische Didaktik und Unterrichtsforschung  
Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Deutschland  
E-Mail: johannes.Koenig@staff.hu-berlin.de; sigrid.bloemeke@staff.hu-berlin.de

didactic and teaching theory. We focus on five vocational requirements with which teaching staff see themselves confronted: structuring lessons, motivation, dealing with heterogeneity, leading lessons and assessing performance. Using the Rasch-Model, we carried out an empirical study on the structure of cross-subject knowledge based on 802 student teachers. As expected, the model, which differentiates between the five areas of knowledge, reflects the data better than a model, which assumes pedagogic knowledge to have a homogeneous structure. This provides empirical insights into long-going theoretical discussions on pedagogical knowledge. The results also reflect the context of typical learning opportunities in teacher training.

**Keywords:** Educational standards · Multi-dimensional Rasch-Model · Pedagogic knowledge · Performance testing · Teacher competencies · Teacher training

## 1 Einleitung

Die Erfassung der professionellen Kompetenz von berufstätigen wie zukünftigen Lehrerinnen und Lehrern<sup>1</sup> ist für die Entwicklung einer empirisch gestützten Theorie des Lehrerberufs und der Lehrerausbildung sowie für die vergleichende Evaluation der Wirksamkeit verschiedener Modelle der Lehrerausbildung von grundlegender Bedeutung (Allemann-Ghionda & Terhart 2006). Entsprechende Forschung stellte lange Zeit ein gravierendes Defizit dar (Blömeke 2004; Baumert & Kunter 2006). Während Schulleistungsstudien, Vergleichsarbeiten oder zentrale Abschlussprüfungen auf der Schülerebene umfassend Daten bereitstellen, wird mit der standardisierten Erfassung des Wissens und der Überzeugungen ihrer Lehrkräfte gerade erst begonnen. Die Studien *Mathematics Teaching in the 21st Century* (MT21; Blömeke, Kaiser & Lehmann 2008), *Teacher Education and Development: Learning to Teach Mathematics* (TEDS-M; Tatto et al. 2008; Blömeke et al. 2009) und *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz* (COACTIV; Brunner et al. 2006) sind in Deutschland die prominentesten Beispiele hierfür.

Die Erfassung des fachübergreifenden, pädagogischen Wissens stellt eine besondere Herausforderung dar und ist mit Ausnahme der Studie MT21 bislang kaum Gegenstand der Kompetenzmessung gewesen. Dabei stellt es einen bedeutsamen Anteil professioneller Kompetenz von Lehrern dar. Fachübergreifende Lerngelegenheiten – als „erziehungswissenschaftliches Studium“ bzw. „Hauptseminar“ bezeichnet – bilden zudem national wie international einen substanziellen Anteil an institutionellen Lerngelegenheiten angehender Lehrkräfte (Schmidt et al. 2007).

Dieser Beitrag widmet sich daher dem pädagogischen Wissen angehender Lehrer. Es wird zunächst eine theoriegeleitete Konzeption zur Erfassung dieses Wissensbereichs präsentiert. Die Konzeptualisierung mündet in der Annahme, dass pädagogisches Wissen nicht homogen ist, sondern dass zwischen unterschiedlichen Wissensgebieten, die sich auf berufliche Anforderungen beziehen, differenziert werden muss. Diese Annahme wird an einer Gruppe von Lehramtsstudierenden aus Deutschland und Österreich empirisch überprüft.

## 1.1 Standards für die Lehrerbildung und professionelle Kompetenz

Mit den „Standards für die Lehrerbildung“ der Kultusministerkonferenz (KMK 2004) haben die Bundesländer ein neues Steuerungsinstrument erhalten, das die globale Entwicklung von der *input*- zur *output*-Orientierung in Bildungssystemen auf den Bereich der Lehrerbildung ausweitet. Zwar wurden die Standards dahingehend kritisiert, dass sie die Kriterien qualitativ hochwertiger Bildungsstandards, wie sie etwa von Klieme et al. (2003) genannt werden, nicht erfüllen (vgl. dazu im Detail Blömeke 2006). Jedoch liefern sie für Studien zur professionellen Kompetenz von Lehrpersonen einen wichtigen Ansatzpunkt, indem sie die beruflichen Aufgaben von Lehrern definieren: Lehrkräfte sind „Fachleute für das Lehren und Lernen“, deren „Kernaufgabe ... die gezielte und nach wissenschaftlichen Erkenntnissen gestaltete Planung, Organisation und Reflexion von Lehr- und Lernprozessen sowie ihre individuelle Bewertung und systemische Evaluation“ ist (KMK 2004, S. 3). Eine Verknüpfung dieser staatlichen Festlegung mit dem pädagogisch-psychologischen Ansatz der Kompetenzorientierung, wie ihn Weinert (2001) ausgearbeitet und Bromme (1992, 1997; Bromme & Haag 2004) für den Lehrerberuf konkretisiert hat, führt zu einem Rahmenkonzept für die detaillierte Ausarbeitung der Anforderungen, die an Lehrkräfte in ihrem Beruf gestellt werden. In diesem Sinne wird auch in der vorliegenden Studie „professionelle Kompetenz“ von Lehrkräften als die erfolgreiche Bewältigung zentraler Anforderungen definiert, die typisch für den Lehrerberuf sind.

In Anlehnung an die Topologie von Wissensdomänen nach Shulman (1986) umfasst „professionelle Kompetenz“ verschiedene, für den Lehrerberuf spezifizierte Wissensbereiche: fachliches Wissen (*content knowledge*), fachdidaktisches Wissen (*pedagogical content knowledge*) und fachübergreifendes, pädagogisches Wissen (*general pedagogical knowledge*). Während zum fachbezogenen Wissen eine bescheidene Anzahl von empirischen Studien vorliegt (MT21, TEDS-M, COACTIV), ist bislang für den Bereich des pädagogischen Wissens theoretisch nicht vollständig und empirisch so gut wie gar nicht geklärt, was genau unter diesem Wissensbereich verstanden werden kann und wie er strukturiert ist. Zwar setzen hier einzelne Studien an (z.B. Baer et al. 2007; Schulte, Bögeholz & Watermann 2008), sie beschränken sich aber stets auf einzelne Standorte (z.B. Hochschulen; vgl. zu dieser Problematik Terhart 2002) und weisen nur sehr geringe Fallzahlen auf; empirisch gestützte Aussagen zur Struktur des pädagogischen Wissens angehender Lehrer lassen sich ihnen nicht entnehmen. Zudem erfasst die Mehrheit der Studien Kompetenzen lediglich über Selbstberichte (z.B. Oser & Oelkers 2001; Abs et al. 2005; Schubarth & Pohlentz 2006). Selbstbeurteilungsverfahren können zwar prinzipiell eine diagnostische Funktion übernehmen. In Hinblick auf die standardisierte Erfassung von Wissen werden sie jedoch kritisch diskutiert und testdiagnostische *paper and pencil*-Verfahren gefordert (vgl. Schaefers 2002; Terhart 2002).

## 1.2 Zur Struktur des pädagogischen Wissens von Lehrkräften

Eine wichtige Ausnahme stellt die Erfassung des fachübergreifenden Wissens in der Studie *Mathematics Teaching in the 21st Century* (MT21) dar, einer internationalen Vergleichsstudie zur Mathematiklehrerbildung in sechs Ländern. Dort wird diese

Wissenskomponente von angehenden Mathematiklehrkräften der Sekundarstufe I in Bulgarien, Deutschland, Mexiko, Südkorea, Taiwan und den USA mit offenen Aufgaben zur Unterrichtsplanung, zur Lernzielkontrolle und zum Umgang mit sozialer Ungleichheit bei 2.628 Testpersonen standardisiert erfasst. In Deutschland wurden vier Ausbildungsregionen in die Studie einbezogen, die wichtige Strukturmerkmale der hiesigen Lehrerbildung widerspiegeln. In diesen vier Regionen wurde eine Vollerhebung an allen Universitäten und Studienseminaren durchgeführt (Blömeke, Kaiser & Lehmann 2008; Schmidt et al. 2007).

Die in MT21 fokussierte Aufgabe, Unterricht zu planen, bildet einen Kern der KMK-Anforderung „Unterrichten“; die darüber hinausgehende Aufgabe, zu kontrollieren, inwieweit Lernziele erreicht wurden, stellt einen Kern der KMK-Anforderung „Beurteilen“ dar (KMK 2004). Das ergänzend aufgenommene „Umgehen mit sozialer Ungleichheit“ ist ein Thema, das viele Länder zunehmend beunruhigt (Kozol 2006; Wiggan 2007). Der Umgang mit Heterogenität steht zudem an erster Stelle der KMK-Definition von Ausbildungszielen für alle Lehrämter (KMK 1997a, 1997b, 1997c). Die drei Anforderungen, Unterricht zu planen, das Erreichen von Lernzielen zu kontrollieren und mit sozialer Ungleichheit umzugehen, weisen gleichzeitig Berührungspunkte zu den Lehrschwerpunkten der am erziehungswissenschaftlichen Studium beteiligten Disziplinen Erziehungswissenschaft (in diesem Fall vor allem der Allgemeinen Didaktik), Psychologie (Pädagogische Psychologie) und Soziologie (Bildungssoziologie) auf, sodass ihre Auswahl als solide begründet gelten kann.

Blömeke, Felbrich und Müller (2008) untersuchen für die deutsche Stichprobe die Struktur des pädagogischen Wissens. In einer konfirmatorischen Faktorenanalyse wird deutlich, dass ein Modell, das zwischen den drei Wissensbereichen Unterrichtsplanung, Lernzielkontrolle und Umgang mit sozialer Ungleichheit unterscheidet, die Antworten der Studierenden und Referendare deutlich besser widerspiegelt als ein Modell, das davon ausgeht, pädagogisches Wissen sei homogen. Die Zusammenhänge zwischen dem erfassten Wissen in den drei Bereichen sind äußerst niedrig ( $04 \leq \phi \leq .34$ ). Umfangreiches Wissen in einem der drei Bereiche geht also nicht notwendigerweise mit umfangreichem Wissen in einem der anderen beiden Bereiche einher.

Dieses Ergebnis stützt langjährige wissenschaftstheoretische Diskussionen zur Struktur des pädagogischen Wissens. Oelkers (1989) und Tenorth (1989) haben wiederholt darauf hingewiesen, dass zwischen verschiedenen Bereichen und Funktionen des pädagogischen Wissens unterschieden werden muss, wenn man es differenziert beschreiben will. Sie führen dies auf den handlungsbezogenen Charakter der Disziplin zurück. Benner (1987) hebt als Eigenheit der Pädagogik ebenfalls hervor, dass sie sich funktional von der Logik der Praxis aus bestimme. Insofern wird auch aus allgemeinpädagogischer Perspektive das Herangehen gestützt, Kernanforderungen an Lehrkräfte zu identifizieren und entsprechendes Wissen zu erfassen – schon um zu klären, „ob und wie sie (die pädagogische Praxis und das sie regierende Wissen; d. Verf.) einer solchen prinzipientheoretischen Struktur entsprechen, die ihnen von außen zugeschrieben wird“ (Oelkers & Tenorth 1991, S. 27).

Aus institutioneller Perspektive bietet eine mehrdimensionale Modellierung des pädagogischen Wissens den Vorteil, differenzierte Erkenntnisse über den Prozess des Wissenserwerbs angehender Lehrkräfte während der Ausbildung erhalten zu können. Es ist nicht

unplausibel anzunehmen, dass die Angebotsstruktur im erziehungswissenschaftlichen Studium zu unterschiedlichen Zeitpunkten unterschiedliche Wirkungen in Teilgebieten entfaltet. Das erziehungswissenschaftliche Studium setzt sich aus mehreren Disziplinen und Komponenten zusammen, deren Gewichtung, Vielfalt und Verbindlichkeit erheblich variieren können (Nolle 2004). Der Vorwurf inhaltlicher Beliebigkeit und mangelnder Systematik gehört zum wiederkehrenden Repertoire darauf ausgerichteter Reviews (Terhart 2000; Keuffer & Oelkers 2001; Schaefers 2002). Sofern mit der Erfassung des fachübergreifenden, pädagogischen Wissens angehender Lehrer der Anspruch verbunden ist, ausbildungsabhängiges Wissen abzubilden – und dieser Anspruch wird in der vorliegenden Studie verfolgt –, dürfte es wenig valide sein, pädagogisches Wissen in seiner Struktur ausschließlich als Einheit zu betrachten, da differenzielle Ausbildungswirkungen überdeckt werden würden. Im Folgenden wird daher das pädagogische Wissen angehender Lehrer mehrdimensional konzeptualisiert. *Inhaltlich* liegt der Mehrwert eines solchen Vorgehens gegenüber MT21 darin, das Wissen angehender Lehrkräfte extensiver beschreiben und damit einen differenzierteren Einblick in deren Wissensstruktur erhalten zu können. *Methodisch* liegt der Mehrwert gegenüber MT21 darin, Annahmen zur Mehrdimensionalität mithilfe einer Rasch-Skalierung (Rost 1996) prüfen zu können und damit anschlussfähig zu sein an einen forschungsmethodischen Standard, der mittlerweile in der Schulleistungsforschung selbstverständlich geworden ist.

## 2 Theoretischer Rahmen der Strukturierung des pädagogischen Wissens

Für eine detaillierte Ausarbeitung der beruflichen Anforderungen an Lehrkräfte greifen wir auf empirisch nachgewiesene Basisdimensionen von Unterrichtsqualität zurück und verknüpfen diese mit Modellen der Allgemeinen Didaktik. Während sich die Allgemeine Didaktik vornehmlich mit den Komponenten und Prozessen von Unterricht auseinandersetzt, versucht die empirische Unterrichtsforschung die Qualität von Unterricht anhand von Merkmalen zu beschreiben, die mit Schülerleistungen in Zusammenhang stehen (Brophy 1999; Baumert et al. 2004). Die zahlreichen Untersuchungen zur Unterrichtsqualität haben in den vergangenen Jahrzehnten eine Fülle von Einzelergebnissen hervorgebracht. Ditton (2000) hat einen Modellrahmen zu wirksamen Faktoren des Unterrichts vorgeschlagen, dessen Struktur an den Ansatz *Quality-Appropriateness-Incentives-Time* (QAIT) von Slavin (1994) angelehnt ist. Dieser verknüpft vier Unterrichtsfaktoren, die sich in empirischen Studien wiederholt als bedeutsam erwiesen haben:

- Qualität der Instruktion (*Quality*),
- Angemessenheit des Anspruchsniveaus (*Appropriateness*),
- Motivierung (*Incentives*) und
- Unterrichtszeit (*Time*).

Da weitere Systematisierungen mit diesem Modell gut harmonieren (siehe z.B. Helmke 2003; Baumert et al. 2004; Good & Brophy 2007), kann die Eingrenzung von Unterrichtsqualität auf diese Basisdimensionen als relativ abgesichert gelten.

Unsere Konzeptualisierung des fachübergreifenden, pädagogischen Wissens angehender Lehrkräfte folgt diesem Diskurs. Allerdings kann eine Systematisierung zur Quali-

tät von Unterricht, basierend auf Untersuchungsergebnissen zum Zusammenhang von Unterrichtsmerkmalen und Schülervariablen, nicht unverändert einer Modellierung von Lehrerkompetenzen zugrunde gelegt werden: Erstens legen überzeugende Modelle zur Beschreibung von Unterricht eine Trennung von Unterrichtsmerkmalen und Lehrerkompetenzen nahe (vgl. z.B. Helmke 2003). Zweitens reicht es nicht aus, allein Schülerleistungen als Kriterium für die Gestaltung von Unterricht zu verwenden, da auf der Seite der Lehrkräfte zahlreiche Überlegungen eine Rolle spielen – etwa zu Zielen, Inhalten und Methoden von Unterricht –, die in der Unterrichtsgestaltung berücksichtigt werden müssen (Blömeke, Herzig & Tulodziecki 2007). Solche Überlegungen werden in Modellen der Allgemeinen Didaktik beschrieben (vgl. z.B. Heimann, Otto & Schulz 1965; Klafki 1985; Tulodziecki, Herzig & Blömeke 2004). Im Folgenden beschreiben und begründen wir daher sowohl aus der Perspektive der empirischen Unterrichtsforschung als auch unter didaktischen Gesichtspunkten die Operationalisierung des fachübergreifenden, pädagogischen Wissens von Lehrkräften als „Fachleute für das Lehren und Lernen“ (KMK 2004, S. 3) anhand von fünf Dimensionen beruflicher Anforderungen:

- Strukturierung von Unterricht,
- Motivierung,
- Umgang mit Heterogenität,
- Klassenführung und
- Leistungsbeurteilung.

## 2.1 Strukturierung von Unterricht

Anhand welcher Kriterien Unterricht strukturiert und geplant werden kann, ist zentraler Gegenstand der Allgemeinen Didaktik. Entsprechend ist die Planung von Unterricht wesentlicher Gegenstand der Lehrerbildung. Untersuchungen zur Unterrichtsqualität verweisen ebenfalls auf die Bedeutung der Lehrerinstruktionen (z.B. Strukturiertheit der Darbietung des Stoffs, der Aufgabenstellungen und -folgen), welche in der Komponente „Qualität der Instruktion“ (*Quality*) im QAIT-Ansatz (Slavin 1994) ihren Niederschlag findet. Insofern kommt dem Wissen zur Strukturierung von Unterricht bei der Erfassung des fachübergreifenden Wissens von zukünftigen Lehrkräften ein besonderer Stellenwert zu.

Vor allem in den ersten Jahren des Berufslebens dürfte die Instruktionsqualität einer Lehrperson in hohem Maße durch ihre Analyse und Planung von Unterricht bedingt sein, da hierbei die Unterrichtsstruktur gedanklich antizipiert wird. Aufgrund dieses Mechanismus stellt die Unterrichtsplanung ein besonders geeignetes Themenfeld dar, das Strukturierungs- und Instruktionswissen von angehenden Lehrern zu testen. In Anlehnung an zentrale didaktische Modelle (insbesondere Heimann, Otto & Schulz 1965; Klafki 1985; Tulodziecki, Herzig & Blömeke 2004) und Erkenntnisse aus der empirischen Unterrichtsforschung fassen wir unter dieser Dimension die folgenden beruflichen Anforderungen an Lehrkräfte zusammen:

- komponentenbezogene Planung und Analyse von Unterricht (z.B. Bedingungs- vs. Entscheidungsfelder),

- prozessbezogene Planung und Analyse von Unterricht (z.B. Phasenmodelle von Unterricht, konkrete Strukturierungshilfen für Schüler im Unterricht) sowie
- curriculare Strukturierung von Unterricht (z.B. Klassifikation von fachübergreifenden Lernzielen).

## 2.2 Motivierung

Schulische Lernprozesse können unter didaktischen Gesichtspunkten als „inszenierte“ Situationen betrachtet werden, in denen Schüler zur Auseinandersetzung mit Aufgabenstellungen angeregt werden sollen (vgl. Tulodziecki, Herzig & Blömeke 2004). Insofern handelt es sich um Lernarrangements, die vorbestimmt sind und von den Lernenden nicht vollständig eigenverantwortlich gestaltet werden können. Einer besonderen Motivierung kommt für schulische Lernprozesse insofern große Bedeutung zu. Nicht für alle Schüler kann Interesse im Sinne einer langfristig überdauernden und stabilen positiven Beziehung zwischen einer Person und einem Inhalt (vgl. Krapp 2001) vorausgesetzt werden. Die *Incentive*-Komponente des QAIT-Modells steht daher für die Ausprägung, inwieweit Lehrkräfte mit ihrem Unterricht eine Motivation von Schülern im Unterricht empirisch gewährleisten (vgl. Slavin 1994; Ditton 2000).

Zu berücksichtigen ist in diesem Zusammenhang, dass es nicht nur darum geht, einzelne Schüler für ihre Aufgaben im Unterricht zu motivieren, sondern dass parallel dazu die ganze Lerngruppe zur Mitarbeit bzw. selbstständigen Arbeit im Unterricht anzuregen ist (Good & Brophy 2007). Unter den möglichen Themen fokussieren wir in unserer Studie auf

- die Leistungsmotivation (z.B. intrinsische vs. extrinsische Motivation) sowie
- Motivierungsstrategien im Unterricht (z.B. Verknüpfung mit lebensweltlichen Erfahrungen der Schüler).

## 2.3 Umgang mit Heterogenität

Der Umgang mit einer heterogenen Schülerklientel gehört zu den größten Herausforderungen des Unterrichts (vgl. Helmke 2003; Horstkemper 2004). Dies gilt für die Grundschule ebenso wie für das gegliederte Sekundarstufenschulwesen. Eine auf einzelne Schüler oder auf Gruppen von Schülern bezogene Individualisierung der Lehr-Lernprozesse im Unterricht stellt insofern eine wichtige Aufgabe von Lehrkräften dar. Individualisierung wird durch didaktische Maßnahmen innerer Differenzierung ermöglicht, die sich auf unterschiedliche Methoden, Lerninhalte, Lernmaterialien, unterschiedliche Lernziel-niveaus und Techniken der Motivierung von Schülern beziehen können (Weinert 1997; Bönsch 2004). Im Vergleich zu dieser Mikroebene sind auf einer Makroebene organisationssoziologische und schultheoretische Bezüge hervorzuheben, etwa die Organisation des Bildungswesens, die Institutionalisierung von Lernprozessen oder bildungspolitische Entwicklungen, welche in den Unterricht hineinwirken (Fend 1980, 2006; Bönsch 2004). Im QAIT-Ansatz werden entsprechende Maßnahmen der Differenzierung und Individualisierung schulischen Lernens unter der Komponente „Angemessenheit des Anspruchsniveaus“ subsumiert (vgl. Slavin 1994; Ditton 2000).



Für die Operationalisierung der Anforderung „Umgang mit Heterogenität“ fokussieren wir vor diesem Hintergrund auf zwei Themenbereiche:

- Kenntnisse zu Differenzierungsmaßnahmen und deren Umsetzung im Unterricht (z.B. äußere vs. innere Differenzierung, Differenzierung von Aufgabenstellungen nach Lern-typen) sowie
- Kenntnisse zur Methodenvielfalt und zu deren Einsatz im Unterricht (z.B. Vor- und Nachteile offener Unterrichtskonzepte, Methodenmuster konventionellen Unterrichts).

## 2.4 Klassenführung

Die Unterscheidung von quantitativen und qualitativen Merkmalen ist kennzeichnend für die Beschreibung von Unterricht in Modellen der empirischen Unterrichtsforschung (vgl. zusammenfassend z.B. Gruehn 2000). Der für aktives Lernen zur Verfügung stehenden Zeit kommt in Bezug auf Unterrichtseffektivität eine Schlüsselrolle zu (Doyle 1986; Weinert 1996; Helmke 2003). Im Rahmen der Testung professioneller Kompetenz von Lehrkräften steht nicht die bereitgestellte Lernzeit für die Schüler im Mittelpunkt, als vielmehr die Sicherung einer effektiven Nutzung dieser Zeit durch die Lehrkraft. Vom Gesichtspunkt der Klassenführung her ist dabei vornehmlich an Maßnahmen zu denken, mit denen eine Lehrkraft störungsarmen Unterricht ermöglicht (Kounin 1976; Doyle 1986). Auch Ditton (2000) führt die beiden Bereiche Klassenführung und Unterrichtszeit in seiner an den QAIT-Ansatz von Slavin (1994) angelehnten Darstellung zusammen. Unter dieser Dimension subsumieren wir daher die folgenden beiden Themen:

- störungspräventive Unterrichtsführung (z.B. Planungsaspekte, konkretes Lehrerverhalten) und
- effektive Nutzung der Unterrichtszeit (z.B. Unterrichtsrouinen).

## 2.5 Leistungsbeurteilung

Die Leistungsbeurteilung von Schülern bleibt im QAIT-Modell weitgehend unberücksichtigt. In der jüngsten Diskussion um die Diagnosekompetenz von Lehrkräften und in den Standards für die Lehrerbildung ist sie jedoch sehr prominent (vgl. KMK 2004; Ziegenspeck 2004; Good & Brophy 2007). Aus didaktischer Sicht stellen die Überprüfung des Lernerfolgs und seine Bewertung ebenfalls wichtige Anforderungen an Lehrkräfte dar (Tulodziecki, Herzig & Blömeke 2004). Schon während des Unterrichtsverlaufs entsteht die Frage nach einer formativen Feststellung des Lernerfolgs. In Klassenarbeiten tritt die Leistungsbeurteilung ganz in den Mittelpunkt, wobei Gütekriterien eine angemessene Qualität der Messung sichern und Bezugsnormen beachtet werden müssen (Bohl 2004). Außerdem wird von Lehrern zunehmend eine Beteiligung an Parallel- oder Vergleichsarbeiten bzw. Lernstandserhebungen gefordert.

Für die Schüler hat die Leistungsbeurteilung eine große Bedeutung, entscheidet sie doch darüber, ob weitere Bildungschancen eröffnet oder verwehrt werden. Lehrkräfte handeln dabei vor dem Hintergrund einer spannungsreichen Diskussion (z.B. um die Widersprüchlichkeit von Fördern und Auslesen oder von Lernen und Leisten) und müs-



sen unterschiedlichen Funktionen von Leistungsbeurteilungen gerecht werden (z.B. Diagnose, Orientierung, Prognose oder Legitimation; Ziegenspeck 1999). Wir fokussieren vor diesem Hintergrund auf

- Funktionen und Formen der Leistungsbeurteilung (z.B. Ziffernzeugnisse vs. alternative Formen),
- zentrale Kriterien (z.B. Gütekriterien) und
- Urteilsfehler (z.B. Voreingenommenheiten von Lehrern).

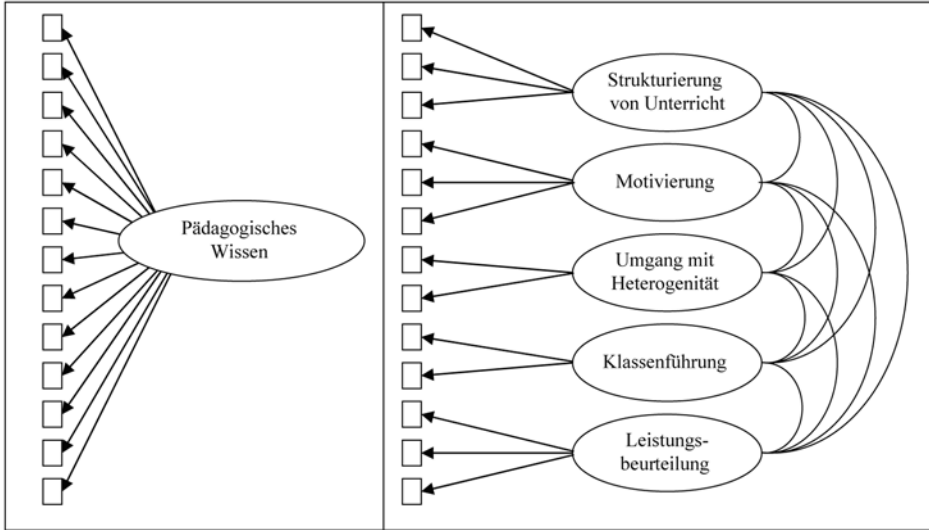
## 2.6 Formen des pädagogischen Wissens von Lehrkräften

Bei der Konzeptualisierung des pädagogischen Wissens aus der Perspektive beruflicher Anforderungen ist zu beachten, dass dieses in zwei Formen vorliegt: in Form deklarativen und in Form prozeduralen Wissens. Prozedurales Wissen stellt dabei jene Form dar, die besonders handlungsrelevant ist, während das ausschließliche Vorliegen deklarativen Wissens Schwierigkeiten bei der Umsetzung von Wissen in die Praxis mit sich führen kann (Gruber & Renkl 2000; Anderson et al. 2001). Prozedurales Wissen ist dadurch gekennzeichnet, dass es situations- und ablauforientiert organisiert ist („Handlungsschemata“). Das heißt in Bezug auf Lehrpersonen, dass sie Unterricht schrittweise anhand typischer Abfolgen wahrnehmen, planen und durchführen (Aebli 1983; Putnam 1987). Wir streben mit unserem Leistungstest an, durch eine entsprechende Gestaltung der Testaufgaben beiden Wissensformen gerecht zu werden.

## 3 Fragestellung

Anliegen des vorliegenden Beitrags ist die Untersuchung der Struktur des fachübergreifenden, pädagogischen Wissens angehender Lehrpersonen. Wir gehen in einem ersten Schritt der Frage nach, ob dieses Wissen anhand der fünf theoretisch herausgearbeiteten beruflichen Anforderungen organisiert ist oder ob es eine homogene Struktur aufweist. In einem zweiten Schritt stellen wir die Frage, welche Zusammenhänge diese verschiedenen inhaltlichen Wissensbereiche – so sie sich denn als unterschiedlich erweisen – aufweisen.<sup>2</sup> Entsprechende Erkenntnisse tragen sowohl zur Klärung theoretischer Grundfragen der Erziehungswissenschaft und Bildungsforschung als auch zur Klärung der Wirksamkeit von Lehrerausbildung bei.

Ausgehend von unserem kompetenzorientierten Ansatz sowie in Übereinstimmung mit wissenschaftstheoretischen Überlegungen, dass sich pädagogisches Wissen funktional von der Logik der Praxis aus bestimme, und institutionenbezogenen Reviews des erziehungswissenschaftlichen Studiums nehmen wir an, dass sich im fachübergreifenden, pädagogischen Wissen von angehenden Lehrkräften die Dimensionen Strukturierung von Unterricht, Motivierung, Umgang mit Heterogenität, Klassenführung und Leistungsbeurteilung unterscheiden lassen. Demgegenüber steht die Annahme, dass nicht die praktischen beruflichen Anforderungen, mit denen sich Lehrkräfte konfrontiert sehen, ihr pädagogisches Wissen strukturieren, sondern dass dieses eine homogene Struktur auf-



**Abb. 1:** Schematische Darstellung des eindimensionalen Modells (links) und des fünfdimensionalen Modells (rechts) pädagogischen Wissens

weist. Technisch würde dies bedeuten, dass sämtliche Items auf einem Generalfaktor laden. Abbildung 1 verdeutlicht die beiden, sich gegenüberstehenden Annahmen.

Hinsichtlich der Verknüpfung der fünf Wissensbereiche erwarten wir einerseits einen vergleichsweise engen Zusammenhang zwischen den Dimensionen *Strukturierung von Unterricht* und *Umgang mit Heterogenität*. Diese beiden Anforderungen stehen sich inhaltlich relativ nahe. Beispielsweise dürften Kenntnisse zur Methodenvielfalt – ein Themenbereich der Dimension *Umgang mit Heterogenität* – eng an Kenntnisse zur *Strukturierung von Unterricht* gekoppelt sein. Zum anderen erwarten wir einen engen Zusammenhang zwischen den Dimensionen *Motivierung* und *Leistungsbeurteilung*, da beispielsweise Formen der Leistungsbeurteilung in großer Nähe zu Kenntnissen zur Leistungsmotivation von Schülern stehen.

Unsere Annahmen zur Zusammenhangsstruktur der Wissensbereiche sehen wir gestützt durch die institutionelle Struktur der Lerngelegenheiten im erziehungswissenschaftlichen Studium. Diese sind institutionell verschiedenen akademischen Disziplinen zugeordnet (insbesondere der Erziehungswissenschaft und der Psychologie). Während die Strukturierung von Unterricht und der Umgang mit Heterogenität Kernfragen der Allgemeinen Didaktik darstellen, also Bestandteil von Lehrveranstaltungen in der Erziehungswissenschaft sind, gehören Motivierung und Leistungsbeurteilung – sowie partiell auch die Klassenführung – zu den klassischen Themen der Pädagogischen Psychologie. Angesichts der Schwierigkeiten mit der Abstimmung der Lehre über Fakultätsgrenzen hinweg (Terhart 2000; Keuffer & Oelkers 2001; Horstkemper 2004) sowie des multidisziplinären Charakters des erziehungswissenschaftlichen Studiums (Baumert & Roeder 1990) nehmen wir jeweils größere Konsistenz für die Lehrangebote innerhalb der

beteiligten akademischen Disziplin an, welche mit einem konsistenteren Wissenserwerb aufseiten der angehenden Lehrer einhergehen sollte.

## 4 Untersuchungsmethode

### 4.1 Stichprobe

Die dargelegte Fragestellung wird im Rahmen einer Pilotstudie zur Erfassung des fachübergreifenden, pädagogischen Wissens von angehenden Lehrkräften für die internationale Vergleichsstudie *Teacher Education and Development Study – Learning to Teach Mathematics* (TEDS-M) untersucht. TEDS-M ist die erste Studie der *International Association for the Evaluation of Educational Achievement* (IEA) im tertiären Bereich sowie die erste international-vergleichende *large-scale* Studie zur Lehrerausbildung, die mit repräsentativen Stichproben arbeitet (Tatto et al. 2008; Blömeke et al. 2009). In TEDS-M werden das fachliche und das fachdidaktische Wissen angehender Lehrkräfte getestet; die Zielpopulation stellen angehende Mathematiklehrerinnen und -lehrer dar, die sich im letzten Jahr ihres Referendariats bzw. Vorbereitungsdienstes befinden. Die vorliegende Pilotstudie diente der Entwicklung eines Tests für das fachübergreifende, pädagogische Wissen angehender Lehrkräfte.

Um den Zugang für TEDS-M zum Feld der zweiten Ausbildungsphase unberührt zu lassen, wurde entschieden, die Pilotstudie in der ersten Ausbildungsphase durchzuführen. Unter forschungsökonomischen Gesichtspunkten war diese Entscheidung vorteilhaft. Auf Anfrage erklärten sich Hochschullehrkräfte von zehn Hochschulen in Deutschland und Österreich bereit, die Testung in ihren Lehrveranstaltungen zu Beginn des Wintersemesters 2007/08 durchzuführen.<sup>3</sup> Damit konnte einerseits gewährleistet werden, dass der Test unter Aufsicht durchgeführt wurde; andererseits ermöglichte der institutionelle Feldzugang zu Lehramtsstudierenden im Vergleich zu einem individuellen Zugang, Mechanismen der Selbstselektion entgegenzuwirken. Allerdings handelt es sich aus diesem Grund auch nicht um eine Gruppe von Personen, die durch zufällige Ziehung für die Teilnahme an der Studie ausgewählt wurden. Vielmehr stellen sie eine Gelegenheitsstichprobe von Lehramtsstudierenden dar, die einfach und kurzfristig zu erreichen waren. Für die vorliegende Studie stellt dies insofern kein Problem dar, als nicht die Gewinnung deskriptiver Erkenntnisse, sondern die *Struktur* des pädagogischen Wissens im Vordergrund steht.

An der Pilotstudie nahmen 802 Lehramtsstudierende teil, davon 310 (38,7%) von sechs Hochschulen in fünf deutschen Bundesländern und 492 (61,3%) von vier Hochschulen in Österreich. 679 (84,7%) der Befragten waren weiblich, 123 (15,3%) männlich. Der höchste Bildungsabschluss von 587 (73,2%) der Befragten war die Allgemeine Hochschulreife, die Fachhochschulreife oder die fachgebundene Hochschulreife, 183 (22,8%) verfügten über eine Zwischenprüfung, ein Vordiplom oder einen Bachelor-Abschluss und 30 (3,7%) besaßen bereits einen (Fach-)Hochschulabschluss.<sup>4</sup> In der ersten Gruppe befinden sich neben deutschen Studierenden im Grund- bzw. Bachelorstudium österreichische Lehramtsstudierende aller Semester, da in Österreich keine Zwischenprüfungen durchgeführt werden. Insgesamt können 264 (32,9%) Personen als „fortgeschrittene Lehramtsstudierende“ identifiziert werden, da sie sich zum Zeitpunkt der Befragung entweder im

Haupt- bzw. Masterstudium oder in einem deutlich höheren Ausbildungssemester befanden. Als Ausbildungsgang verfolgten 397 (49,5%) der Befragten das Grundschullehramt (Klasse 1 bis 4), 94 (11,7%) ein stufenübergreifendes Lehramt (Grund-, Haupt-, Realschule), 123 (15,3%) das Lehramt für die Sekundarstufe I, 119 (14,8%) das Lehramt für die Sekundarstufen I und II und weitere 64 (8,0%) verfolgten entweder das Lehramt für die Sonderschule oder die Berufsschule.<sup>5</sup>

## 4.2 Instrumente

Zur Erfassung des fachübergreifenden Wissens wurde unter der Leitung des deutschen TEDS-M-Teams von Erziehungswissenschaftlern und Psychologen aus Deutschland, Taiwan und den USA – teilweise auch unter Aufgreifen von Anregungen durch parallel laufende Initiativen im Bereich der pädagogisch-psychologischen Testentwicklung (vgl. z.B. Schulte, Bögeholz & Watermann 2008)<sup>6</sup> – ein umfangreicher Pool an Testaufgaben theoriegeleitet und unter besonderer Berücksichtigung von Erkenntnissen der Allgemeinen Didaktik sowie der empirischen Unterrichtsforschung entwickelt. Für jede der fünf beruflichen Anforderungen wurden sowohl geschlossene als auch offene Antwortformate konzipiert (vgl. exemplarisch die Aufgaben in den Abb. 2 bis 6). Nur vereinzelt konnte dabei auf vorhandene Testaufgaben zurückgegriffen werden (vgl. die Testaufgabe in Abb. 3 oder die Endnote 6).<sup>7</sup>

Dieser erste deduktive Entwicklungsschritt wurde um einen induktiven Schritt ergänzt. In einer intensiven Erprobung wurden sämtliche Testaufgaben von angehenden und berufstätigen Lehrern sowie Lehrerausbildnern, Erziehungswissenschaftlern und Psychologen probeweise ausgefüllt und diskutiert. Auf dieser Basis erfolgte eine erste Überarbeitung der theoriegeleitet entwickelten Testaufgaben, wobei vor allem Fragen der Verständlichkeit im Vordergrund standen. Anschließend wurden die Testaufgaben durch externe nationale und internationale Experten für Fragen der Lehrerausbildung unterschiedlicher Fachrichtungen (u.a. Matthias Baer, Jere Brophy, Gabriele Kaiser) hinsichtlich verschiedener Kriterien begutachtet. Bei diesem Schritt standen vor allem Fragen der inhaltlichen Validität, der Zugehörigkeit der Items zu einer der fünf Dimensionen und des Schwierigkeitsgrads im Vordergrund. Testaufgaben, die einem oder mehreren dieser Kriterien nicht genügten, wurden ausgeschlossen.

Auf der Basis des verbleibenden Testaufgaben-Pools wurde die Entscheidung über die endgültige Teststruktur erneut theoriegeleitet vorgenommen. Alle fünf beruflichen Anforderungen sollten mit einer Zahl an Items vertreten sein, die Rasch-Skalierungen ermöglichen würden. Dabei war unser Ziel, die Anforderungen angesichts ihrer gleichermaßen hohen Bedeutung für die Gestaltung von Lehrprozessen annähernd gleichgewichtig im Test durch Items zu repräsentieren, die allen genannten Kriterien genügen. Für den Bereich *Klassenführung* ist dies in dem engen Zeitrahmen, der von der internationalen Projektleitung für TEDS-M vorgegeben war, allerdings nur begrenzt gelungen, sodass hier weniger Items zu finden sind.

Angesichts der speziell bei der Erfassung von pädagogischem Wissen bestehenden Schwierigkeit, bei Testaufgaben mit geschlossenem Antwortformat *a priori* richtige und falsche Antworten so wenig wie möglich simplifizierend und schematisierend festzulegen, kam eine relativ große Anzahl von Testaufgaben mit offenem Antwortformat zum Einsatz. Im Vergleich zu geschlossenen Aufgaben gewährleisteten diese zudem in höhe-

Stellen Sie sich vor, Sie helfen einer angehenden Lehrperson bei der Auswertung ihres Unterrichts, weil sie dies noch nie gemacht hat.

Welche Fragen würden Sie ihr stellen, damit sie ihren Unterricht angemessen analysiert?

Nennen Sie zehn zentrale Fragen und formulieren Sie diese bitte aus.

- 1)
- 2)
- 3)
- 4)
- 5)
- 6)
- 7)
- 8)
- 9)
- 10)

Beispiel	Originalantworten	Punkte
1.	1) Haben meine Schüler das Lernziel erreicht? 2) War das Unterrichtsmaterial brauchbar? 3) Wie war das Zeitmanagement? 4) 5) 6) 7) 8) 9) 10)	3
2.	1) Haben alle Kinder die wesentlichen Inhalte verstanden? 2) War meine zeitliche Planung richtig? 3) Was würde ich das nächste Mal ändern? 4) Konnte ich die Lernentwicklung aller Schüler überprüfen? 5) Waren Kinder über-/unterfordert? 6) War mein Material ansprechbar? 7) Gab es ein angemessenes Lernklima? 8) Hat es den Kindern Spaß bereitet und ihr Interesse geweckt? 9) Welche sozialen Kompetenzen konnte ich fördern? 10)	9

**Abb. 2:** Beispielaufgabe mit offenem Antwortformat zur Erfassung des deklarativen und anteilig prozeduralen Wissens zur beruflichen Anforderung *Strukturierung von Unterricht* (Antwortkasten in verkleinerter Darstellung) sowie Originalantworten mit Punktevergabe

rem Maße, nicht nur deklaratives, sondern anteilig auch prozedurales Wissen zu erfragen. Die folgenden Analysen basieren vor diesem Hintergrund auf 50 Testaufgaben, in denen insgesamt 136 Punkte erreicht werden konnten:

Bei welchen der folgenden Fälle handelt es sich um eine intrinsische Motivation, bei welchen um eine extrinsische Motivation?

*Kreuzen Sie ein Kästchen pro Zeile an.*

	intrinsische Motivation	extrinsische Motivation
<b>Ein Schüler lernt vor einer Mathematikarbeit, weil er...</b>		
A. für eine gute Note eine Belohnung erwartet.	<input type="checkbox"/> <sub>1</sub>	<input checked="" type="checkbox"/> <sub>2</sub>
B. einen Tadel für eine schlechte Note vermeiden möchte.	<input type="checkbox"/> <sub>1</sub>	<input checked="" type="checkbox"/> <sub>2</sub>
C. an mathematischen Problemen interessiert ist.	<input checked="" type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
D. seine Eltern nicht enttäuschen möchte.	<input type="checkbox"/> <sub>1</sub>	<input checked="" type="checkbox"/> <sub>2</sub>
E. seine gute Leistungsposition in der Klasse auch in Zukunft behalten möchte.	<input type="checkbox"/> <sub>1</sub>	<input checked="" type="checkbox"/> <sub>2</sub>

**Abb. 3:** Beispielaufgabe mit geschlossenem Antwortformat zur Erfassung des deklarativen Wissens zur beruflichen Anforderung *Motivierung* (in Anlehnung an Edelmann 2000, S. 269) sowie korrekte Lösungen (jeweils 1 Punkt pro richtiger Antwort)

- 5 Aufgaben (35 Punkte), mit denen das pädagogische Wissen zur *Strukturierung von Unterricht* erfasst wird (vgl. das Aufgabenbeispiel in Abb. 2),
- 13 Aufgaben (28 Punkte) zur Motivierung (vgl. Abb. 3),
- 15 Aufgaben (39 Punkte) zum Umgang mit Heterogenität (vgl. Abb. 4),
- 7 Aufgaben (12 Punkte) zur Klassenführung (vgl. Abb. 5) und
- 10 Aufgaben (22 Punkte) zur Leistungsbeurteilung (vgl. Abb. 6).

34 Aufgaben weisen ein offenes (vgl. die Aufgabenbeispiele in den Abb. 2, 4 und 5), 16 Aufgaben ein geschlossenes Antwortformat auf (vgl. Abb. 3 und 6). 28 Testfragen erfragen eher deklaratives Wissen (vgl. Abb. 3 und 6). 22 Testfragen schildern dagegen eine typische Unterrichtssituation, mit denen spezifische Anforderungen, die an eine Lehrkraft gestellt werden, verbunden sind, und erfragen unterschiedliche Handlungsoptionen (vgl. Abb. 2, 4 und 5). Mit diesem Fragetyp wird nicht nur deklaratives, sondern auch anteilig prozedurales Wissen erfasst, da die Erfragung unterschiedlicher Handlungsoptionen die potenzielle Situationsflexibilität einer angehenden Lehrperson berücksichtigt. Im Folgenden findet jedoch keine analytische Trennung von deklarativen und prozeduralen Wissensformen Anwendung, da das vorrangige Erkenntnisinteresse dieses Beitrags auf der inhaltlichen Wissensstruktur liegt.

Die Kodierung offener Antworten erfolgte mithilfe von Kategoriensystemen, die im mehrfachen Wechsel zwischen deduktiver und induktiver Herangehensweise entwickelt worden waren. Dies gewährleistete einerseits theoretisch begründete Kategorien und andererseits die erschöpfende Nutzung der in den Antworten vorzufindenden Informationen. Bei der Entwicklung der Kategoriensysteme wurden die Antworten aus rund 20 Prozent der Fragebögen einbezogen. Anschließend erfolgte eine Begutachtung der Kategoriensysteme durch Experten für Allgemeine Didaktik und empirische Unterrichts-

Nennen Sie drei Unterrichtskonzepte bzw. Unterrichtsmethoden, die es den Schüler(inne)n grundsätzlich ermöglichen, über die Differenzierung im Unterricht mit zu entscheiden.

1)  
2)  
3)

Beispiel	Originalantworten	Bewertung
1.	1) Lernzirkel 2) Freiarbeit 3) Lehrervortrag	2 Punkte für Lernzirkel und Freiarbeit
2.	1) Wochenplan 2) Stationenlernen 3) Lernbuffet	3 Punkte

**Abb. 4:** Beispielaufgabe mit offenem Antwortformat zur Erfassung des deklarativen und anteilig prozeduralen Wissens zur beruflichen Anforderung *Umgang mit Heterogenität* (Antwortkasten in verkleinerter Darstellung) sowie Originalantworten mit Punktvergabe

forschung aus Deutschland, Taiwan und den USA. Die Kodierungen der Antworten der restlichen Fragebögen erfolgten letztlich jeweils durch zwei geschulte Rater,<sup>8</sup> die unabhängig voneinander die offenen Antworten kodierten. Als Übereinstimmungsmaß wurde Cohen’s Kappa berechnet (Wirtz & Caspar 2002). Werte größer .75 gelten als sehr gute Übereinstimmung. Die für unser Anliegen berechneten Kappa-Werte variieren zwischen .69 und .98 mit einem Mittelwert von  $M = .85$  ( $SD = .08$ ). Unterhalb der Grenze von .75 liegen lediglich drei Testfragen, sodass die für die Kodierung entwickelten Kategoriensysteme als bewährt gelten können. Bei fehlender Übereinstimmung wurden in gemeinsamer Diskussion – auch unter Hinzuziehung von Experten – Einigungen erzielt. Schwierig zu kodierende Antworten wurden als „Grenzfälle“ dokumentiert, um im Anschluss ein konsistentes Vorgehen mit ähnlichen Antworten zu gewährleisten.

### 4.3 Datenanalyse

Jedem Befragten lagen rund zwei Drittel der Testaufgaben vor. Ein vollständig ausbalanciertes Testdesign, in dem die Testaufgaben ausgewogen nach den fünf beruflichen Anforderungen, eingeschätzter Item-Schwierigkeit, Antwortformat (offen/geschlossen) und erwarteter Bearbeitungszeit auf sechs Testhefte verteilt wurden, ermöglichte große Überschneidungen der eingesetzten Testaufgaben und eine hinreichende Verknüpfung der Daten.

Bei einem solchen Design hat man es im Zuge der Datenanalyse aufgrund des Rotationsdesigns und der Nichtbeantwortung von einzelnen Fragen mit verschiedenen Formen an fehlenden Werten zu tun. Im Zuge einer Skalierung nach den Methoden der *Item-*



Was kann eine Lehrkraft schon während der Vorbereitung ihres Unterrichts tun, um Störungen im Unterricht entgegenzuwirken?  
Nennen Sie drei vorbeugende Maßnahmen, die nach Ihrer Meinung während der Unterrichtsvorbereitung am wichtigsten sind.

1)  
2)  
3)

Beispiel	Originalantworten	Bewertung
1.	1) im Klassenzimmer herumlaufen 2) Störungen unterbinden 3)	0 Punkte, da Antworten keine vorbeugenden Maßnahmen während der Unterrichtsvorbereitung darstellen
2.	1) Regeln festlegen 2) Einhaltung der Regeln 3) Grenzen setzen	1 Punkt für das vorbereitende Festlegen von Regeln
3.	1) genaue Zeitplanung/-einteilung 2) mehrere Methodenwechsel vorsehen 3) Alternativen in der „Hinterhand“ haben, um flexibel den Plan im Fall von Störungen zu ändern	3 Punkte

**Abb. 5:** Beispielaufgabe mit offenem Antwortformat zur Erfassung des deklarativen und anteilig prozeduralen Wissens zur beruflichen Anforderung *Klassenführung* (Antwortkasten in verkleinerter Darstellung) sowie Originalantworten mit Punktevergabe

Wenn diagnostische Urteile fair und genau sein sollen, dann müssen sie drei Gütekriterien erfüllen. Welche sind das?

*Bitte nur ein Kästchen ankreuzen.*

A.	Neutralität, Reliabilität, Veridikalität	<input type="checkbox"/> 1
B.	Objektivität, Reliabilität, Validität	<input checked="" type="checkbox"/> 2
C.	Objektivität, Reliabilität, Veridikalität	<input type="checkbox"/> 3
D.	Neutralität, Reliabilität, Validität	<input type="checkbox"/> 4

**Abb. 6:** Beispielaufgabe mit geschlossenem Antwortformat zur Erfassung des deklarativen Wissens zur beruflichen Anforderung *Leistungsbeurteilung* mit korrekter Lösung (1 Punkt)

*Response*-Theorie sind beide Formen unproblematisch, da auf Itemebene angemessene Wege des Einbezugs von Fällen mit fehlenden Werten gefunden werden können. Die Skalierung der Daten erfolgt im vorliegenden Fall auf der Grundlage des dichotomen Raschmodells (Rost 1996). Testhefte, die über gemeinsame Items verbunden sind und ansonsten aus unterschiedlichen Items bestehen, können so auf eine gemeinsame Skala gebracht werden.

Die für die Skalierung genutzte Software *Conquest* (Wu, Adams & Wilson 1997; Wu 1997) weist jedem Item aufgrund seiner Lösungsquote einen Schwierigkeitsparameter und jeder befragten Person entsprechend der gezeigten Leistung einen Fähigkeitsparameter zu. Mit dem mehrdimensionalen *Random Coefficient Multinomial Logit*-Modell, das ebenfalls in *Conquest* implementiert ist, können mehrere Fähigkeiten gleichzeitig berücksichtigt werden, deren Zusammenhänge messfehlerbereinigt ausgegeben werden (Rost 1996; Wu & Adams 2006).

In die empirische Überprüfung der Struktur des pädagogischen Wissens wurden 136 dichotome Items einbezogen. Ausgeschlossen wurden jene Items, die bei einer der angestrebten Skalierungen keinen zufrieden stellenden Fit aufwiesen (*Weighted Mean Square* < 0.80 bzw. > 1.20; vgl. Adams 2002). Einzelne Items, die Werte zwischen 0.75 und 0.79 bzw. zwischen 1.21 und 1.25 annahmen, verblieben im Test, wenn ein Ausschluss aus theoretischen Gründen nicht angemessen erschien, um zu gewährleisten, dass ein möglichst breites inhaltliches Spektrum an Testfragen für jede der fünf beruflichen Anforderungen in den Analysen Berücksichtigung finden konnte (vgl. z.B. die Passung von Item 91 in Tab. 1). Die Streuung der Itemschwierigkeiten in Relation zu den Personenfähigkeiten kann als gut bezeichnet werden (vgl. Abb. 7). Trotz der Häufung von Items in der Mitte des Spektrums folgt die Verteilung der Itemschwierigkeiten gut der Streuung der Personenfähigkeiten. Insofern ist mit unserem Test über den ganzen Bereich hinweg eine präzise Schätzung des fachübergreifenden Wissens von angehenden Lehrkräften gegeben.<sup>9</sup>

## 5 Ergebnisse

Entsprechend der grafischen Veranschaulichung in Abb. 1 wurde für das pädagogische Wissen angehender Lehrkräfte die Passung eines Modells, das eine homogene Wissensstruktur, und die Passung eines Modells, das zwischen Wissen in den fünf Anforderungsbereichen unterscheidet, geschätzt. Der Vergleich der Anpassung beider Modelle an den Datensatz wurde über einen Index aus *Likelihood* und Parameteranzahl vorgenommen (vgl. Tab. 2). Die Differenz der beiden Indizes zeigt, dass das fünfdimensionale Modell die Antworten der Studierenden im Vergleich zum eindimensionalen Modell deutlich besser widerspiegelt (vgl. zum methodischen Vorgehen z.B. Rost 1996). Der Unterschied wird auch statistisch signifikant. Dieses Ergebnis deutet darauf hin, dass es wichtig ist, von einer anforderungsbezogenen Struktur des pädagogischen Wissens auszugehen und nicht Homogenität anzunehmen.

Für die Überprüfung der internen Konsistenz der fünf Testdimensionen ziehen wir die *Expected A Posteriori Estimation* (EAP) heran, die eine unverzerrte Beschreibung der Population liefert und die mehrdimensionale Modellstruktur berücksichtigt (vgl. Wu

**Tab. 1:** Itemschwierigkeitsparameter der eindimensionalen Skalierung für die fünf in den Abb. 2 bis 6 aufgeführten Beispielaufgaben sowie Angaben zum *Weighted Item Fit* (*Mean Square*, Konfidenzintervall, T-Wert)

Beispiel- Aufgabe in Abbildung	Teil- Antwort	Item in Abb. 7	Item-schwierig- keitsparameter	Mean Square	Konfidenz- intervall	T-Wert
2	1)	66	-1.3130	0.92	0.91, 1.09	-1.8
	2)	67	-1.0470	0.88	0.92, 1.08	-3.1
	3)	68	-0.5700	0.84	0.93, 1.07	-4.7
	4)	69	-0.1090	0.84	0.93, 1.07	-4.7
	5)	70	0.5080	0.85	0.92, 1.08	-3.9
	6)	71	1.1720	0.85	0.88, 1.12	-2.6
	7)	72	1.8590	0.90	0.83, 1.07	-1.1
	8)	73	2.6090	0.94	0.73, 1.27	-0.4
	9)–10)	74	3.4380	0.96	0.57, 1.43	-0.1
	A.	87	-1.3440	0.95	0.91, 1.09	-1.2
3	B.	88	-1.1410	1.05	0.92, 1.08	1.2
	C.	89	-2.0780	0.93	0.87, 1.13	-1.0
	D.	90	-1.1880	1.07	0.92, 1.08	1.6
	E.	91	0.5630	1.21	0.91, 1.09	4.5
	1)	13	-0.2420	0.86	0.93, 1.07	-4.3
4	2)	14	0.6720	0.90	0.91, 1.09	-2.3
	3)	15	2.2890	0.99	0.77, 1.23	-0.0
	1)	75	-1.7890	0.96	0.89, 1.11	-0.7
5	2)	76	-0.5470	1.04	0.93, 1.07	1.1
	3)	77	1.0860	1.04	0.89, 1.11	0.7
6	–	123	-0.6020	0.97	0.94, 1.06	-0.9

1997). Das eindimensionale Modell zeigt mit einem Wert von .91 eine sehr hohe Reliabilität. Die Dimensionen des differenzierten Modells besitzen aufgrund der geringeren Itemanzahl in jedem Anforderungsbereich unterschiedlich hohe Reliabilitäten (vgl. Tab. 3). Sie liegen jedoch für vier der fünf Dimensionen in einem guten und selbst für die Anforderung der Klassenführung, für die das pädagogische Wissen mit nur 12 Items erfasst werden konnte, in einem akzeptablen Bereich.

Neben Kennwerten zur Reliabilität geben die Zusammenhänge der fünf Anforderungsbereiche einen wichtigen Einblick in die Struktur des fachübergreifenden Wissens. Tab. 4 zeigt die Ergebnisse. Insgesamt fallen die latenten Korrelationen niedrig aus, d.h. höheres Wissen in Bezug auf eine Anforderung geht nicht unbedingt systematisch mit höherem Wissen in Bezug auf eine andere Anforderung einher. Über die bessere Modellanpassung hinaus spricht auch dies für eine Ausdifferenzierung des pädagogischen Wissens.

Die höchsten Zusammenhänge ( $\geq .65$ ) bestehen wie erwartet zwischen dem Wissen zu den beiden didaktischen Anforderungen *Strukturierung von Unterricht* und *Umgang mit Heterogenität* einerseits sowie zwischen dem Wissen zu den beiden pädagogisch-psychologisch ausgerichteten Anforderungen *Motivierung* und *Leistungsbeurteilung* anderer-

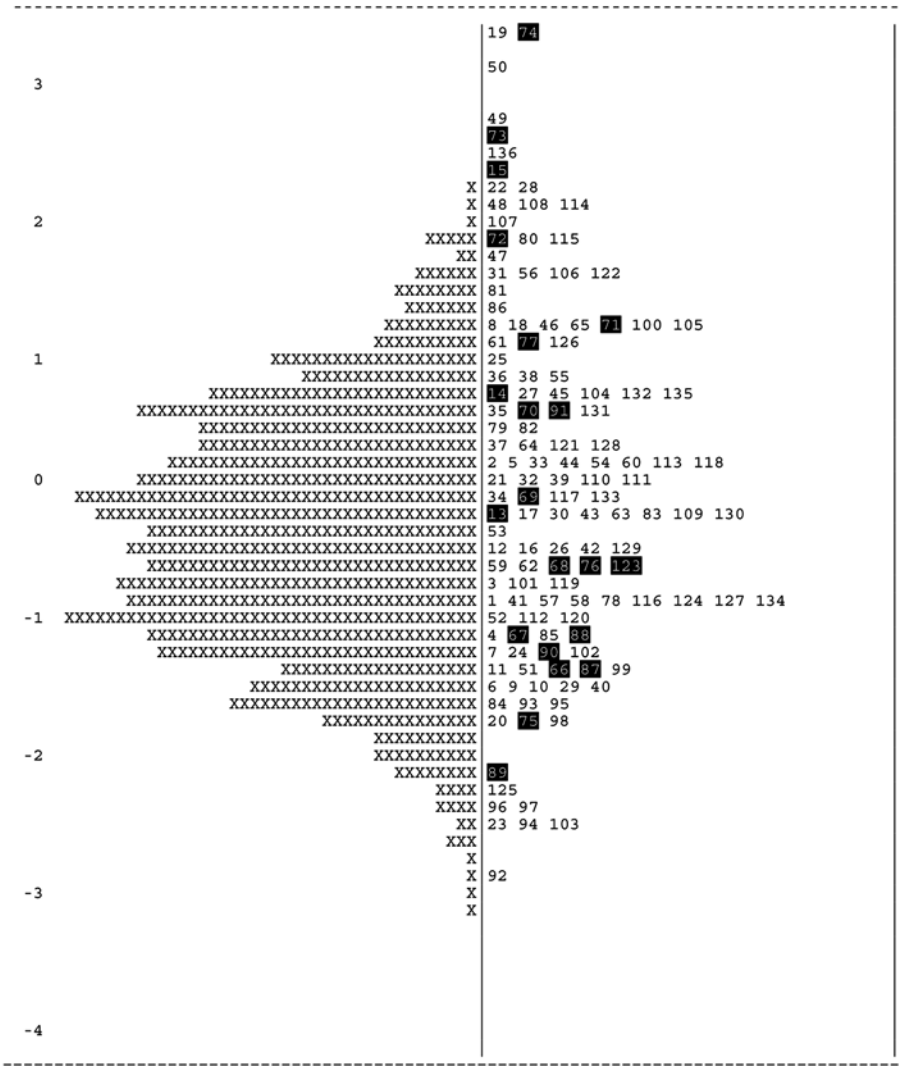


Abb. 7: Streuung der Itemschwierigkeiten in Relation zu den Personenfähigkeiten (Zahlen 1 bis 136: Items, ein X entspricht 1.1 Personen). Die Markierungen verdeutlichen die Items der fünf in den Abb. 2 bis 6 aufgeführten Beispielaufgaben (vgl. zur Zuordnung Tab. 1).

seits. Zwischen didaktischen und pädagogisch-psychologischen Wissensbereichen fallen die Zusammenhänge wie erwartet deutlich niedriger aus (.48 bzw. .49). Die niedrigsten latenten Korrelationen ( $\leq .42$ ) gehen mit der Dimension Klassenführung einher, die entgegen unserer Annahme – von der Dimension Motivierung abgesehen – nur lose mit den anderen Dimensionen verknüpft ist.

**Tab. 2:** Statistik zu den ein- und fünfdimensionalen Modellen

Modell	Abweichung	Zahl der geschätzten Parameter	Differenz		
			Abweichung	Parameter	<i>p</i>
1-dimensional	66246.96	137	3758.81	14	<.001
5-dimensional	62488.15	151			

Zusammenfassend lässt sich festhalten, dass sowohl die Modellanpassung als auch die Reliabilitäten und die Zusammenhänge der Wissensbereiche deutlich auf die angenommene fünfdimensionale Struktur des pädagogischen Wissens angehender Lehrkräfte hinweisen. Das Zusammenhangsmuster spiegelt die angenommene Gruppierung in didaktische und pädagogisch-psychologische Subdimensionen des Wissens, wobei Klassenführung mit beiden nur lose assoziiert ist. Die Interkorrelationen können somit als Beleg für die Konstruktvalidität angesehen werden.

Für eine genauere Untersuchung der Struktur des pädagogischen Wissens ist eine wichtige Frage, inwieweit dieses Ergebnis nicht nur für die gesamte Stichprobe angehender Lehrer in Deutschland und Österreich gilt, sondern auch für Subgruppen, denen aus Sicht der deutschen Lehrerausbildung ein bedeutsamer Stellenwert zugeschrieben werden kann. In unserem Fall stellt sich diese Frage für die Subgruppe der *deutschen* Lehramtsstudierenden ( $n=310$ ) sowie für die Subgruppe der *fortgeschrittenen* Lehramtsstudierenden ( $n=264$ ). Auch wenn die Stichprobengröße durch die Aufteilung jeweils deutlich reduziert wird, was mit Reliabilitätseinbußen einhergeht, die wiederum geringere Interkorrelationen nach sich ziehen, soll abschließend der Frage nachgegangen werden, inwieweit das pädagogische Wissen in diesen beiden Gruppen dieselbe Struktur aufweist.

Die Itemstatistiken zeigen für die Testitems auch in diesen beiden Subgruppen eine sehr gute Modellanpassung. Maximal zwei Items überschreiten die Grenze eines *Weighted Mean Square*-Wertes von 1.25 in einer der Modellierungen. Tabelle 5 kann darüber hinaus entnommen werden, dass erneut das differenziertere Modell die Antworten der Studierenden signifikant besser widerspiegelt als das eindimensionale Modell. Dies gilt sowohl in der Subgruppe der deutschen als auch in jener der fortgeschrittenen Lehramtsstudierenden.

Trotz Verringerung der Stichprobengröße liegen die *EAP*-Reliabilitäten für das eindimensionale Modell bei hohen .88 für die deutschen und bei hohen .89 für die fortgeschrittenen Lehramtsstudierenden. In der fünfdimensionalen Modellierung wird mit unserem Test auch das Wissen in den beiden didaktischen Anforderungsbereichen für beide Subgruppen sehr präzise erfasst, da die jeweiligen Reliabilitäten über .8 liegen (vgl. Tab. 6). Die Reliabilitäten für die drei Dimensionen *Klassenführung*, *Motivierung* und *Leistungsbeurteilung* liegen für die Fortgeschrittenengruppe noch in einem akzeptablen Bereich (>.6), die Reliabilität für die Dimension *Klassenführung* entspricht sogar der für

**Tab. 3:** Reliabilitäten des fünfdimensionalen Modells

	Strukturierung von Unterricht	Umgang mit Heterogenität	Klassenführung	Motivierung	Leistungsbeurteilung
EAP	.85	.85	.65	.72	.71

**Tab. 4:** Messfehlerbereinigte Zusammenhänge zwischen den fünf Wissensbereichen

	(1)	(2)	(3)	(4)
(1) Umgang mit Heterogenität				
(2) Strukturierung von Unterricht	.65			
(3) Klassenführung	.37	.42		
(4) Motivierung	.49	.48	.59	
(5) Leistungsbeurteilung	.49	.49	.41	.70

die gesamte Stichprobe. Dagegen liegen diese Reliabilitäten für die deutsche Subgruppe, die sich aus Anfängern und Fortgeschrittenen zusammensetzt, deutlich niedriger (<.6). Für diese Subgruppe kann somit eine weniger homogene Struktur des pädagogischen Wissens zur *Motivierung* und zur *Leistungsbeurteilung* angenommen werden.

Die Zusammenhänge fallen aufgrund der Reliabilitätseinbußen zum Teil etwas niedriger aus. Sie weisen aber ein strukturell vergleichbares Muster wie in der Gesamtstichprobe auf, das im Hinblick auf die Einordnung des Wissens zur Klassenführung sogar noch erwartungsgemäßer ausfällt (vgl. Tab. 7). Die beiden didaktischen Dimensionen und die drei pädagogisch-psychologischen Dimensionen hängen sowohl bei deutschen Lehramtsstudierenden als auch bei Fortgeschrittenen untereinander jeweils deutlich stärker zusammen ( $\geq .49$ , in der deutschen Gruppe mit einer Ausnahme sogar  $\geq .55$ ) als über die Dimensionen hinweg ( $\leq .42$ , in der deutschen Gruppe sogar  $\leq .30$ ). Insgesamt ist festzustellen, dass die Zusammenhänge in der Fortgeschrittenengruppe deutlich homogener ausfallen als in der deutschen Stichprobe, die sich aus Anfängern und Fortgeschrittenen zusammensetzt. Dies kann als weiterer Indikator für die Angemessenheit unserer Teststruktur gewertet werden, die im Hinblick auf Ergebnisse der Lehrerausbildung konzipiert wurde.

## 6 Zusammenfassung und Diskussion

Die Forschungslage zum fachübergreifenden, pädagogischen Wissen von angehenden und praktizierenden Lehrkräften ist äußerst schmal. Lässt man die Studien mit Skalen zur Selbsteinschätzung außer Acht, liegt praktisch kein geeigneter theoretischer Ansatz zur Erfassung des pädagogischen Wissens vor. Eine empirische Erfassung wurde bislang ebenfalls kaum angegangen. Der vorliegende Beitrag begegnet diesem Forschungsdefizit,

**Tab. 5:** Statistik zu den ein- und fünfdimensionalen Modellen in den Gruppen der deutschen und der fortgeschrittenen Lehramtsstudierenden

Modell	Deutsche Lehramtsstudierende					Fortgeschrittene Lehramtsstudierende				
	Abweichung	Parameter	Differenz			Abweichung	Parameter	Differenz		
			Abw.	Par.	p			Abw.	Par.	p
1-dim.	25655.9	137	1488.5	14	<.001	21347.5	137	1217.6	14	<.001
5-dim.	24167.4	151				20129.9	151			

**Tab. 6:** Interne Konsistenzen des fünfdimensionalen Modells (deutsche Lehramtsstudierende/fortgeschrittene Lehramtsstudierende)

	<b>Strukturierung von Unterricht</b>	<b>Umgang mit Heterogenität</b>	<b>Klassenführung</b>	<b>Motivierung</b>	<b>Leistungsbeurteilung</b>
EAP	.81/.81	.81/.82	.59/.64	.52/.64	.55/.63

indem ein Testkonzept vorgestellt und die Struktur des damit erfassten pädagogischen Wissens an einer Stichprobe von 802 Lehramtsstudierenden untersucht wurde.

Lehrkräfte werden von uns unter Bezug auf die Standards für die Lehrerbildung der KMK (2004, S. 3) als „Fachleute für das Lehren und Lernen“ verstanden, deren „Kernaufgabe ... die gezielte und nach wissenschaftlichen Erkenntnissen gestaltete Planung, Organisation und Reflexion von Lehr- und Lernprozessen sowie ihre individuelle Bewertung und systemische Evaluation“ ist. Unter Bezug auf die empirische Forschung zu Basisdimensionen der Unterrichtsqualität und auf Modelle der Allgemeinen Didaktik wurden fünf zentrale berufliche Anforderungen an Lehrkräfte theoretisch begründet, die den Ausgangspunkt für die Testentwicklung bildeten.

Leitende Annahme unserer Studie war, dass das pädagogische Wissen von Lehrkräften in sich keine vollständig homogene Struktur aufweist, sondern dass es anhand dieser fünf Anforderungen der beruflichen Praxis strukturiert werden kann. Die Testfragen bezogen sich nicht allein auf deklaratives Wissen, sondern erfragten auch anteilig prozedurales Wissen, wobei offene und geschlossene Antwortformate zum Einsatz kamen.

Die angenommene Mehrdimensionalität des pädagogischen Wissens wurde mittels des Rasch-Modells überprüft, indem ein eindimensionales Modell, das die Homogenitätsannahme repräsentiert, einem fünfdimensionalen Modell, das die differenzierte Wissensstruktur widerspiegelt, gegenüber gestellt wurde. Die Ergebnisse machen deutlich, dass letztere Annahme angemessener ist, und zwar sowohl in Bezug auf die gesamte Stichprobe als auch in Bezug auf die beiden Subgruppen der deutschen Lehramtsstudierenden und der fortgeschrittenen Studierenden.

Unter wissenschaftstheoretischen Gesichtspunkten stützen unsere Daten damit die These von Benner (1987), dass sich das pädagogische Wissen funktional von der Logik der Praxis aus bestimmt. Sie rechtfertigt auch die Orientierung an der Kompetenztheorie, für die ein Ausgehen von beruflichen Anforderungen konstitutiv ist (Bromme 1992, 1997; Bromme & Haag 2004; Weinert 2001). Für zukünftige Studien zur Lehrerbildung bedeutet unser Ergebnis, dass die Erfassung von Ergebnissen des pädagogischen

**Tab. 7:** Messfehlerbereinigte Zusammenhänge der fünf Wissensbereiche (deutsche Lehramtsstudierende/fortgeschrittene Lehramtsstudierende)

	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>
(1) Umgang mit Heterogenität				
(2) Strukturierung von Unterricht	.55/.49			
(3) Klassenführung	.26/.40	.33/.34		
(4) Motivierung	.27/.34	.33/.40	.57/.63	
(5) Leistungsbeurteilung	.39/.39	.29/.42	.44/.51	.56/.64



Anteils der Ausbildung sehr differenziert erfolgen muss, möchte man nicht entscheidende Wirkungen übersehen. In Ergänzung zu einer eindimensionalen Erfassung, die durch eine mehrdimensionale Testanlage nicht ausgeschlossen wird, weist eine Unterscheidung von verschiedenen Wissensbereichen den Vorteil auf, dass spezifische Rückschlüsse auf Stärken und Schwächen der Lehrerbildung möglich werden. Erst auf diese Weise können beispielsweise Einflüsse von didaktischen Lehrveranstaltungen gegenüber Einflüssen von pädagogisch-psychologischen Lehrveranstaltungen abgegrenzt werden.

Die *Reliabilitäten* erweisen sich für die gesamte Stichprobe und für die Stichprobe der fortgeschrittenen Lehramtsstudierenden als akzeptabel bis gut. Die in allen (Teil-)Stichproben besonders hohe interne Konsistenz des Wissens der Studierenden zur Strukturierung von Unterricht und zum Umgang mit Heterogenität dürfte widerspiegeln, dass diese Themen in der Lehrerbildung in Form fachdidaktischer und allgemeindidaktischer Lehrveranstaltungen in höherem Maße abgedeckt werden als Themen der Pädagogischen Psychologie wie Motivierung und Leistungsbeurteilung, und zwar von Beginn des Studiums an. Für das Wissen zu den drei pädagogisch-psychologisch geprägten beruflichen Anforderungen an Lehrkräfte zeigt sich vor allem in der deutschen Stichprobe eine etwas geringere interne Konsistenz. Dies ist vermutlich auf die Zusammensetzung dieser Stichprobe aus Anfängern und Fortgeschrittenen zurückzuführen. Am Ende der universitären Ausbildung fallen die Reliabilitäten höher aus, was auf einen homogeneren Wissensbestand hindeutet. Traditionell sieht das Lehramtsstudium entsprechende Lehrveranstaltung eher für das Haupt- als für das Grundstudium vor, wobei der Zeitpunkt der Belegung oft frei gewählt werden kann.

Dass die Dimension *Klassenführung* insgesamt eine geringere Reliabilität aufweist, kann einerseits darauf zurückgeführt werden, dass die Skala deutlich kürzer ist. Dieses Defizit konnten wir für TEDS-M bereits beheben, sodass diese Interpretation mit Vorliegen der Ergebnisse Anfang 2010 überprüft werden kann. Es ist aber auch denkbar, dass das Antwortverhalten mindestens zum Teil durch eine schwächere Berücksichtigung des Themengebietes in der Lehrerbildung bedingt ist (vgl. z.B. Helmke 2003). *Klassenführung* ist – erneut vor allem in Deutschland – traditionell kein Bestandteil der Didaktik-Ausbildung und steht auch in Veranstaltungen der Pädagogischen Psychologie gegenüber den beiden Bereichen *Motivierung* und *Leistungsbeurteilung* zurück.

Die *Zusammenhänge* der fünf Wissensbereiche sind insgesamt nicht sehr hoch. Auch dies spricht eher für das ausdifferenzierte Modell anstelle einer Homogenitätsannahme. Das Wissen zu den beiden didaktischen Anforderungen *Strukturierung von Unterricht* und *Umgang mit Heterogenität* auf der einen Seite sowie das Wissen zu den beiden pädagogisch-psychologisch geprägten Anforderungen *Motivierung* und *Leistungsbeurteilung* (sowie partiell auch *Klassenführung*) hängen untereinander erwartungsgemäß jeweils am stärksten zusammen. Dieses Muster spiegelt die klassischerweise an den beiden sehr unterschiedlichen Disziplinen Erziehungswissenschaft bzw. Psychologie orientierte Vermittlung des fach-übergreifenden, pädagogischen Wissens in der Lehrerbildung. Generell zeigen sich zum Ende der Lehrerbildung stärkere Zusammenhänge, die als eine umfassendere Vernetztheit der Wissensbereiche als Folge der Lehrerbildung interpretiert werden können.

Analysen zur Struktur des mit einem Test erfassten Wissens stellen – wenn sich die theoretischen Annahmen wie im vorliegenden Fall empirisch bestätigen lassen – eine

wichtige Form der Konstruktvalidierung dar (vgl. Borsboom, Mellenbergh & van Heerden 2004). Gleichzeitig kommt die vorliegende Untersuchung der Forderung nach, Unterrichtsqualität theoriegeleitet und multivariat zu verstehen (vgl. Einsiedler 1997). Bevor auf weitere Forschungsaufgaben eingegangen wird, sind allerdings zwei Aspekte kritisch zu diskutieren: die Eingrenzung des pädagogischen Wissens auf unterrichts- und beurteilungsbezogene Anforderungen an Lehrer sowie die Eingrenzung der kompetenzorientierten Messung auf Wissen.

Zu Ersterem ist festzuhalten, dass Lehrkräfte einem breiteren Anforderungsspektrum zu genügen haben, als in der vorliegenden Studie untersucht wird. Wenngleich die KMK (2004) in den Standards zur Lehrerbildung die hier untersuchten Dimensionen als den beruflichen Kern definiert hat, gehören Erziehen sowie Mitwirken an der Schulentwicklung zu wichtigen Aufgaben, deren Bedeutung nicht negiert werden soll. Zu beiden Aufgabengebieten ist es allerdings derzeit schwierig, einen Test zu entwickeln, da kaum empirische Erkenntnisse vorliegen und Erziehen zudem normativ besetzt ist. Zur Eingrenzung des Tests auf Wissen ist anzumerken, dass für die Erfassung von fachübergreifenden Überzeugungen als zweitem Bestandteil von Kompetenz (Weinert 2001) bereits Konzepte vorliegen, die sich auch empirisch bewährt haben (vgl. insbesondere Müller, Felbrich & Blömeke 2008). Sie wurden in MT21 und sie werden auch in TEDS-M erhoben, sodass ihre gemeinsame Berücksichtigung in Datenanalysen möglich ist. Mit dem vorliegenden Beitrag wurde hingegen das Ziel verfolgt, das verbleibende Forschungsdefizit im Bereich des pädagogischen Wissens angehender Lehrer zu verringern.

In zukünftigen Arbeiten wird es darum gehen, vertiefende Strukturanalysen durchzuführen, etwa die Bildung von Kompetenzniveaus, und das Testinstrument weiter zu validieren. In Bezug auf Letzteres ist festzuhalten, dass die Qualität eines Instruments zur Erfassung des Wissens angehender Lehrkräfte auch daran gemessen werden muss, ob es in der Lage ist, ausbildungsabhängiges Wissen angehender Lehrer zu erheben. Studien zum Zusammenhang zwischen Merkmalen der Lehrerbildung und dem erworbenen Wissen angehender Lehrer könnten hier wichtige Einblicke liefern. Erste Validierungsanalysen, in denen das Ausbildungsstadium angehender Lehrkräfte als Außenkriterium einbezogen und ihre Abiturnote kontrolliert wurde, belegen die Ausbildungsabhängigkeit des hier getesteten Wissens (König, Peek & Blömeke 2008; König & Blömeke 2009). Ferner ist dies Gegenstand der internationalen Vergleichsstudie TEDS-M unter Verwendung einer für ganz Deutschland repräsentativen Stichprobe angehender Lehrkräfte im letzten Jahr ihrer Ausbildung, deren Ergebnisse Anfang 2010 vorgestellt werden. Zudem wird TEDS-M einen internationalen Vergleich des fachübergreifenden pädagogischen Wissens mit den USA und Taiwan ermöglichen, die ebenfalls nach dem hier vorgestellten Testkonzept vorgehen. Damit dürfen wichtige Einblicke in das pädagogische Wissen angehender Lehrer – u.a. weiterführende Erkenntnisse zu seiner Struktur – über einen mitteleuropäischen Referenzrahmen hinaus sowie generell internationale Anschlussfähigkeit in der Erfassung von Kompetenzen angehender Lehrer erwartet werden.

## Anmerkungen

- 1 Um den Lesefluss zu erleichtern, wird im Folgenden für Lehrerinnen und Lehrer, Schülerinnen und Schüler usw. nur die männliche Geschlechtsform verwendet. Selbstverständlich sind immer beide Geschlechter gemeint.
- 2 Das Erkenntnisinteresse des vorliegenden Beitrags richtet sich also auf die inhaltliche Dimensionierung des pädagogischen Wissens, nicht auf eine Dimensionierung des pädagogischen Wissens in deklarative und prozedurale Formen.
- 3 Für die Mithilfe bei der Durchführung dieser Studie danken wir herzlich diesen Hochschullehrkräften sowie allen weiteren Personen, die dabei mitgewirkt haben. Außerdem gilt unser Dank allen Lehramtsstudierenden, die sich die Mühe gemacht haben, an dem Test teilzunehmen.
- 4 Für zwei befragte Personen (0,2%) liegen uns keine Angaben über den höchsten Bildungsabschluss vor.
- 5 Für fünf befragte Personen (0,6%) liegen uns keine Angaben über den verfolgten Ausbildungsgang vor.
- 6 Wir danken Frau Klaudia Schulte für die Möglichkeit, einzelne Testaufgaben aus dem von ihr entwickelten Pädagogisch-Psychologischen Test für unser Anliegen nutzen zu dürfen.
- 7 Bei nachweislich wissenschaftlichen Gründen sind – unter Gewährleistung der Testsicherheit in Form einer schriftlichen Vertraulichkeitserklärung – die Testaufgaben an einem Rechner der Autoren unter Aufsicht vollständig einsehbar.
- 8 Für die höchst sorgfältig durchgeführten Kodierarbeiten danken wir herzlich Jessica Ebrahimi, Sabine Freigang, Doreen Ressel, Anne-Kathrin Schille und Katharina Schoop.
- 9 Bei der Interpretation der Itemkennwerte in Tab. 1 ist zu beachten, dass diese aus der eindimensionalen Skalierung stammen. In der fünfdimensionalen Skalierung verbessert sich die Item-Passung noch einmal deutlich. Beispielsweise fällt die Item-Fit-Statistik für das Item 91 in der eindimensionalen Skalierung grenzwertig aus (vgl. Tab. 1: *Mean Square* 1.21, Konfidenzintervall (0.91, 1.09), T-Wert 4.5), während die Item-Fit-Statistik für das Item 91 in der fünfdimensionalen Skalierung akzeptabel ist (*Mean Square* 1.08, Konfidenzintervall (0.92, 1.08), T-Wert 1.7).

## Literatur

- Abs, H. J., Döbrich, P., Vögele, E., & Klieme, E. (2005). *Skalen zur Qualität der Lehrerbildung – Dokumentation der Erhebungsinstrumente: Pädagogische Entwicklungsbilanzen an Studienseminaren*. Frankfurt a.M.: Gesellschaft zur Förderung Pädagogischer Forschung.
- Adams, R. (2002). Scaling PISA cognitive data. In R. Adams & M. Wu (Eds.), *PISA 2000 technical report* (pp. 99–108). Zugriff am 07. Mai 2009, <http://www.oecd.org/dataoecd/53/19/33688233.pdf>.
- Aebli, H. (1983). *Zwölf Grundformen des Lehrens*. Stuttgart: Klett-Cotta.
- Allemann-Ghionda, C., & Terhart, E. (2006). Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf. Zur Einleitung in das Beiheft. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf* (Zeitschrift für Pädagogik, 51. Beiheft, S. 7–11). Weinheim: Beltz Verlag.

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., & Pintrich, P. R. (Eds.) (2001). *A taxonomy for learning, teaching and assessing: a revision of bloom's taxonomy of educational objectives*. New York: Langenscheidt Elt.
- 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, 15–47.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520.
- Baumert, J., & Roeder, M. P. (1990). Expansion und Wandel der Pädagogik. Zur Institutionalisierung einer Referenzdisziplin. In L.-M. Alisch, J. Baumert & K. Beck (Hrsg.), *Professionswissen und Professionalisierung* (S. 79–128). Braunschweig: Universitätsverlag.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland. Ergebnisse des zweiten internationalen Vergleichs* (S. 314–354). Münster: Waxmann.
- Benner, D. (1987). *Allgemeine Pädagogik*. Weinheim: Juventa.
- Blömeke, S. (2004). Empirische Befunde zur Wirksamkeit der Lehrerbildung. In S. Blömeke, P. Reinhold, G. Tulodziecki & Wildt, J. (Hrsg.), *Handbuch Lehrerbildung* (S. 59–91). Bad Heilbrunn/Obb.: Klinkhardt.
- Blömeke, S. (2006). KMK-Standards für die LehrerInnenbildung in Deutschland. Ein Kommentar. *Journal für Lehrerinnen- und Lehrerbildung*, 1, 25–33.
- Blömeke, S., Felbrich, A., & Müller, C. (2008). Messung des erziehungswissenschaftlichen Wissens angehender Lehrkräfte. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 171–193). Münster: Waxmann.
- Blömeke, S., Herzig, B., & Tulodziecki, G. (2007). Zum Stellenwert empirischer Forschung für die Allgemeine Didaktik. *Unterrichtswissenschaft*, 35, 355–381.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Münster: Waxmann.
- Blömeke, S., Kaiser, G., Lehmann, R., König, J., Döhrmann, M., Buchholtz, C., et al. (2009). TEDS-M: Messung von Lehrerkompetenzen im internationalen Vergleich. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität – Bedingungen, Genese, Wirkungen und Messung* (S. 181–210). Weinheim: Beltz.
- Bönsch, M. (2004). *Intelligente Unterrichtsstrukturen. Eine Einführung in die Differenzierung*. Baltmannsweiler: Schneider Hohengehren.
- Bohl, T. (2004). *Prüfen und Bewerten im Offenen Unterricht* (2. Aufl.). Weinheim: Beltz.
- Borsboom, D., Mellenbergh, G. J., Heerden, J. van (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- 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.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule. Bd. 3* (S. 177–212). Göttingen: Hogrefe.
- Bromme, R., & Haag, L. (2004). Forschung zur Lehrerpersönlichkeit. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 777–793). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Brophy, J. (1999). *Teaching*. Brussels: International Academy of Education. Zugriff am 07. Mai 2009, <http://www.ibe.unesco.org/publications/EducationalPracticesSeriesPdf/prac01e.pdf>.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., & Dubberke, T. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9, 521–544.
- Ditton, H. (2000). Qualitätskontrolle und -sicherung in Schule und Unterricht. Ein Überblick über den Stand der empirischen Forschung. In A. Helmke, W. Hornstein & E. Terhart (Hrsg.), *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule* (Zeitschrift für Pädagogik: 41. Beiheft, S. 73–92). Weinheim: Beltz Verlag.
- Doyle, W. (1986). Classroom organization and management. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 392–431). New York: MacMillan.
- Edelmann, W. (2000). *Lernpsychologie*. Weinheim: BeltzPVU.
- Einsiedler, W. (1997). Unterrichtsqualität und Leistungsentwicklung. Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 225–240). Weinheim: BeltzPVU.
- Fend, H. (1980). *Theorie der Schule*. München: Auer.
- Fend, H. (2006). *Neue Theorie der Schule*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Good, T. L. & Brophy, J. E. (2007). *Looking in Classrooms*. Boston: Allyn & Bacon.
- Gruber, H., & Renkl, A. (2000). Die Kluft zwischen Wissen und Handeln: Das Problem des trägen Wissens. In G. H. Neuweg (Hrsg.), *Wissen – Können – Reflexion. Ausgewählte Verhältnisbestimmungen* (S. 155–174). Innsbruck: Studienverlag.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen*. Münster: Waxmann.
- Heimann, P., Otto, G., & Schulz, W. (1965). *Unterricht – Analyse und Planung*. Hannover: Schroedel.
- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Horstkemper, M. (2004). Erziehungswissenschaftliche Ausbildung. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.), *Handbuch Lehrerbildung* (S. 461–476). Bad Heilbrunn/Obb.: Klinkhardt.
- Keuffer, J., & Oelkers, J. (Hrsg.) (2001). *Reform der Lehrerbildung in Hamburg. Abschlussbericht der von der Senatorin für Schule, Jugend und Berufsbildung und der Senatorin für Wissenschaft und Forschung eingesetzten Hamburger Kommission Lehrerbildung*. Weinheim: Beltz.
- Klafki, W. (1985). *Neue Studien zur Bildungstheorie und Didaktik. Beiträge zur kritisch-konstruktiven Didaktik*. Weinheim: Beltz.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (2003). Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Frankfurt a.M.: DIPF. Zugriff am 07. Mai 2009, [http://www1.dipf.de//publikationen/volltexte/zur\\_entwicklung\\_nationaler\\_bildungsstandards.pdf](http://www1.dipf.de//publikationen/volltexte/zur_entwicklung_nationaler_bildungsstandards.pdf).
- KMK (1997a). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Rahmenvereinbarung über die Ausbildung und Prüfung für übergreifende Lehrämter der Primarstufe und für alle oder einzelne Schulformen der Sekundarstufe I (Lehramtstyp 2). Beschluss der Kultusministerkonferenz vom 28.2.1997 in der Fassung vom 05.02.2009. Zugriff am 07. Mai 2009, [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/1997/1997\\_02\\_28-RV-Lehramtstyp-2.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1997/1997_02_28-RV-Lehramtstyp-2.pdf).
- KMK (1997b). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Rahmenvereinbarung über die Ausbildung und Prüfung für Lehrämter für alle oder einzelne Schulformen der Sekundarstufe I (Lehramtstyp 3). Beschluss der Kultusministerkonferenz vom 28.2.1997 in der Fassung vom 05.02.2009. Zugriff am 07. Mai 2009, [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/1997/1997\\_02\\_28-RV-Lehramtstyp-3.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1997/1997_02_28-RV-Lehramtstyp-3.pdf).

- KMK (1997c). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Rahmenvereinbarung über die Ausbildung und Prüfung für ein Lehramt der Sekundarstufe II [allgemein bildende Fächer] oder für das Gymnasium (Lehramtstyp 4). Beschluss der Kultusministerkonferenz vom 28.2.1997 in der Fassung vom 05.02.2009. Zugriff am 07. Mai 2009, [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/1997/1997\\_02\\_28-RV-Lehramtstyp-4.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1997/1997_02_28-RV-Lehramtstyp-4.pdf).
- KMK (2004). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004. Zugriff am 07. Mai 2009, [http://sts-bs-darmstadt.bildung.hessen.de/gesetze\\_und\\_verordnungen/kmk\\_standards.pdf](http://sts-bs-darmstadt.bildung.hessen.de/gesetze_und_verordnungen/kmk_standards.pdf).
- König, J., & Blömeke, S. (2009). Pädagogisches Wissen von österreichischen Lehramtsstudierenden. *Erziehung & Unterricht*, 159(1/2), 175–186. Zugriff am 23. April 2009 [http://www.oebv.at/erziehung\\_unterricht/archiv/09\\_01/koenig\\_bloemeke.pdf](http://www.oebv.at/erziehung_unterricht/archiv/09_01/koenig_bloemeke.pdf).
- König, J., Peek, R., & Blömeke, S. (2008). Zum Erwerb von pädagogischem Wissen in der universitären Ausbildung: Unterscheiden sich Studierende verschiedener Lehramter und Kohorten? *Lehrerbildung auf dem Prüfstand*, 1(2), 639–657.
- Kounin, J. S. (1976). *Techniken der Klassenführung*. Stuttgart: Waxmann.
- Kozol, J. (2006). *The shame of the Nation. The restoration of apartheid in schooling in America*. New York: Crown.
- Krapp, A. (2001). Interesse. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 286–294). Weinheim: BeltzPVU.
- Müller, C., Felbrich, A., & Blömeke, S. (2008). Schul- und professionstheoretische Überzeugungen. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 277–302). Münster: Waxmann.
- Nolle, A. (2004). *Evaluation der universitären Lehrerinnen- und Lehrerausbildung*. München: Meidenbauer.
- Oelkers, J. (1989). *Die große Aspiration. Zur Herausbildung der Erziehungswissenschaft im 19. Jahrhundert*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Oelkers, J., & Tenorth, H.-E. (1991). Pädagogisches Wissen als Orientierung und als Problem. In J. Oelkers & H.-E. Tenorth (Hrsg.), *Pädagogisches Wissen* (Zeitschrift für Pädagogik: 27. Beiheft, S. 13–35). Weinheim: Beltz.
- Oser, F., & Oelkers, J. (Hrsg.) (2001). *Die Wirksamkeit der Lehrerbildungssysteme*. Chur: Rüegger.
- Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal*, 24(1), 13–48.
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Huber.
- Schaefers, C. (2002). Forschung zur Lehrerausbildung in Deutschland – Eine bilanzierende Übersicht der neueren empirischen Studien. *Schweizerische Zeitschrift für Bildungswissenschaften*, 24(1), 65–88.
- Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L., et al. (2007). The preparation gap: Teacher education for middle school mathematics in six countries – Mathematics teaching in the 21st century (MT21 Report). Zugriff am 07. Mai 2009, [http://usteds.msu.edu/related\\_research.asp](http://usteds.msu.edu/related_research.asp).
- Schubarth, W., & Pohlenz, P. (Hrsg.) (2006). *Qualitätsentwicklung und Evaluation in der Lehrerbildung. Die zweite Phase: das Referendariat*. Potsdam: Universitätsverlag.
- Schulte, K., Bögeholz, S., & Watermann, R. (2008). Selbstwirksamkeitserwartungen und Pädagogisches Professionswissen im Verlauf des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft*, 11, 268–287.

- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21, 141–157.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in Mathematics (TEDS-M): Conceptual framework*. East Lansing, MI: Michigan State University.
- Tenorth, H.-E. (1989). Professionstheorie für Pädagogen? *Zeitschrift für Pädagogik*, 35, 809–824.
- Terhart, E. (Hrsg.) (2000). *Perspektiven der Lehrerbildung in Deutschland. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Kommission*. Weinheim: Beltz.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz*. Münster: ZKL-Verlag. Zugriff am 07. Mai 2009, [http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-1151/Standards\\_fuer\\_die\\_Lehrerbildung\\_Eine\\_Expertise\\_fuer\\_die\\_Kultusministerkonferenz.pdf](http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-1151/Standards_fuer_die_Lehrerbildung_Eine_Expertise_fuer_die_Kultusministerkonferenz.pdf).
- Tulodziecki, G., Herzog, B., & Blömeke, S. (2004). *Gestaltung von Unterricht. Eine Einführung in die Didaktik*. Bad Heilbrunn/Obb.: Klinkhardt.
- Weinert, F. E. (1996). *Psychologie des Lernens und der Instruktion*. Enzyklopädie der Psychologie. Pädagogische Psychologie, Bd. 2. Göttingen: Hogrefe.
- Weinert, F. E. (1997). Notwendige Methodenvielfalt: Unterschiedliche Lernfähigkeit der Schüler erfordern variable Unterrichtsmethoden des Lehrers. In H. Olberding (Hrsg.), *Lernmethoden – Lehrmethoden – Wege zur Selbstständigkeit* (Friedrich Jahresheft 1997, S. 50–52). Velber: Friedrich.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–66). Göttingen: Hogrefe.
- Wiggan, G. (2007). Race, school achievement, and educational inequality: Toward a student-based inquiry perspective. *Review of Educational Research*, 77, 310–333.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wu, M. L. (1997). *The development and application of a fit test for use with generalised item response model*. Unpublished Master's Dissertation, University of Melbourne, Australia. Melbourne: University of Melbourne.
- Wu, M. L., & Adams, R. (2006). Modelling mathematics problem solving item responses using a multidimensional IRT model. *Mathematics Education Research Journal*, 18(2), 93–113.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-Aspect Test Software* [Computer program]. Camberwell, Vic.: Australian Council for Educational Research.
- Ziegenspeck, J. (1999). *Handbuch Zensur und Zeugnis in der Schule*. Bad Heilbrunn/Obb.: Klinkhardt.
- Ziegenspeck, J. (2004). Diagnose und Beurteilung. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.), *Handbuch Lehrerbildung* (S. 501–511). Bad Heilbrunn/Obb.: Klinkhardt.