

König, Johannes

**Zur Bildung von Kompetenzniveaus im Pädagogischen Wissen von
Lehramtsstudierenden: Terminologie und Komplexität kognitiver
Bearbeitungsprozesse als Anforderungsmerkmale von Testaufgaben?**

Lehrerbildung auf dem Prüfstand 2 (2009) 2, S. 244-262



Empfohlene Zitierung/ Suggested Citation:

König, Johannes: Zur Bildung von Kompetenzniveaus im Pädagogischen Wissen von Lehramtsstudierenden: Terminologie und Komplexität kognitiver Bearbeitungsprozesse als Anforderungsmerkmale von Testaufgaben? - In: Lehrerbildung auf dem Prüfstand 2 (2009) 2, S. 244-262 - URN: urn:nbn:de:0111-pedocs-147039

<http://nbn-resolving.de/urn:nbn:de:0111-pedocs-147039>

in Kooperation mit / in cooperation with:

VEP

www.vep-landau.de

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Herausgeber

Rainer Bodensohn, Reinhold S. Jäger und Andreas Frey
Bürgerstraße 23, 76829 Landau/Pfalz
Telefon: +49 6341 906 165, Telefax: +49 6341 906 166

Verlag

Empirische Pädagogik e. V.
Bürgerstraße 23, 76829 Landau/Pfalz
Telefon: +49 6341 906 180, Telefax: +49 6341 906 166
E-Mail: info@vep-landau.de
Homepage: <http://www.vep-landau.de>

Umschlaggestaltung

© Harald Baron

Druck

DIFO Bamberg

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung, werden vorbehalten. Kein Teil des Werks darf in irgendeiner Form (durch Fotografie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlags reproduziert oder unter Verwendung elektronischer Systeme verbreitet werden.

ISSN 1867-2779

© Verlag Empirische Pädagogik, Landau 2009

Inhalt

Editorial	201
Originalarbeiten	
Beck, K.: Implementation von Anreizstrukturen zur Qualitätsverbesserung in der Lehrerbildung	206
Hoffmann, L., Azizi Ghanbari, S. & Schott, F.: Untersuchung zur Entwicklung von Bildungsstandards: Rekonstruk- tion eines vorläufigen Kompetenzmodells und der Kompetenzdiagnos- tik am Beispiel einer Unterrichtseinheit für das Schulfach Ethik	225
König, J.: Zur Bildung von Kompetenzniveaus im Pädagogischen Wissen von Lehramtsstudierenden: Terminologie und Komplexität kognitiver Bearbeitungsprozesse als Anforderungsmerkmale von Testaufgaben?	244
Košinár, J.: Das Konzept „Ganzheitliche Stressprävention“ – Überprüfung der Effektivität eines Trainingsprogramms in der Lehrerausbildung	263
Rothland, M.: Kooperation und kollegiale Unterstützung im Lehrerberuf. Empirische Befunde zum Berufsbild und zu berufsrelevanten Einstellungen angehender Lehrkräfte	282
Schubarth, W., Speck, K., Seidel, A. & Wendland, M.: Unterrichtskompetenzen bei Referendaren und Studierenden. Empirische Befunde der Potsdamer Studien zur ersten und zweiten Phase der Lehrerausbildung	304
Weresch-Deperrois, I., Bodensohn, R. & Jäger, R. S.: Curriculare Standards in der Praxis: Einschätzung ihres Stellenwerts, ihrer Anwendungshäufigkeit, Schwierigkeit und Bedeutung in der Lehrerausbildung und universitären Vorbereitung im Bachelor-Studium der Lehrerbildung – eine Erkundungsstudie	324
Frey, A. & Jäger, R. S.: Der entwicklungspädagogische Ansatz in der Lehrerbildung	346
Richtlinien zur Manuskriptgestaltung	361
Impressum	368

Contents

Editorial	201
Articles	
Beck, K.: Implementation of a structure of incentives to improve the quality of teacher education	206
Hoffmann, L., Azizi Ghanbari, S. & Schott, F.: An empirical study towards the development of educational standards: Reconstruction of a preliminary competency model and of the compe- tence diagnostics for a teaching unit in the school subject ethics	225
König, J.: On the modelling of proficiency levels of future teachers' pedagogical knowledge: Can test item terminology and complexity of cognitive pro- cesses be used to describe item difficulty?	244
Košinár, J.: The concept of "Holistic Stress Prevention" – surveying the effective- ness of a teacher training program	263
Rothland, M.: Cooperation and Support Among Teachers. Empirical findings about the occupational image and about attitudes of prospective teachers concerning their job	282
Schubarth, W., Speck, K., Seidel, A. & Wendland, M.: Teaching competencies for trainees and students. Empirical findings of the Potsdam studies on first and second phase of teacher education	304
Weresch-Deperrois, I., Bodensohn, R. & Jäger, R. S.: Curricular standards in practice: Appraisal of their importance, fre- quency of application, difficulty, relevance in teacher education and pre- paration by the university in the field of BA-teacher education – an explorative study	324
Frey, A. & Jäger, R. S.: The developmental pedagogical approach in teacher education	346

Originalbeitrag

Johannes König

Zur Bildung von Kompetenzniveaus im Pädagogischen Wissen von Lehramtsstudierenden: Terminologie und Komplexität kognitiver Bearbeitungsprozesse als Anforderungsmerkmale von Testaufgaben?

Zusammenfassung: Kompetenzmodelle zur Messung von Schülerleistungen sehen mittlerweile häufig die Bildung von Kompetenzniveaus vor, um erreichte Leistungen anschaulich und kriteriumsbezogen beschreiben zu können. Ein solches Vorgehen findet bislang kaum Realisierung bei der Erfassung und Modellierung von Kompetenzen angehender Lehrkräfte. Die aktuelle Diskussion um Standards in der Lehrerausbildung impliziert jedoch die Forderung nach einer solchen Umsetzung. Der Beitrag exploriert eine Möglichkeit der Bildung von Kompetenzniveaus im fachübergreifenden, pädagogischen Wissen angehender Lehrkräfte der ersten Ausbildungsphase. Verwendet wird der TEDS-M Test zur Erfassung des pädagogischen Wissens und eine Stichprobe von 197 Studierenden der Universität zu Köln. Als Anforderungsmerkmale werden die Terminologie der Testaufgaben und die Komplexität kognitiver Bearbeitungsprozesse in ein entsprechendes Kompetenzmodell einbezogen. Die Verteilungen der geschätzten Itemschwierigkeitsparameter nach Anforderungsmerkmalen sind weitgehend erwartungskonform und ermöglichen die Bildung von Kompetenzniveaus. Häufigkeitsverteilungen der Studierenden nach Ausbildungsstadium und Ausbildungsgang verweisen auf die inhaltliche Validität der gebildeten Kompetenzniveaus.

Schlagwörter: Lehrerausbildung – Leistungstest – Kompetenzniveaus – Pädagogisches Wissen

On the modelling of proficiency levels of future teachers' pedagogical knowledge: Can test item terminology and complexity of cognitive processes be used to describe item difficulty?

Summary: To provide a concrete description of attained achievement, competency measurement contains the modelling of proficiency levels. This option is regularly applied in research studies on student achievement, whereas it has hardly been taken into account by studies measuring future teachers' knowledge yet. However, the current discussion on teacher education standards requires such an application. This report explores a strategy how proficiency levels of first phase future teachers' pedagogical knowledge can be modelled. The TEDS-M test instrument measuring pedagogical knowledge is applied and a sample of 197 students at the University of Cologne is used. The terminology of test items as well as the complexity of cognitive processes are used to describe test item difficulty and are incorporated into a corresponding measurement model. Item parameter estimation distributions differentiated by the criteria used to describe test difficulty are nearly as expected and enables the definition of proficiency levels. Frequency distributions of future teachers differentiated by teacher education program and period of training point out the content validity of the modelled proficiency levels.

Key words: achievement test – pedagogical knowledge – proficiency levels – teacher education

1. Einleitung

Mit den „Standards für die Lehrerbildung“ der Kultusministerkonferenz (KMK, 2004a) haben die Bundesländer ein neues Steuerungsinstrument erhalten. Dieses

weitet die Ergänzung einer *input*- um eine *output*-Orientierung in Bildungssystemen auch auf den Bereich der Lehrerbildung aus und erweckt die Zuversicht, eine Orientierungshilfe für die Entwicklung angemessener Ausbildungsprogramme sowie deren Evaluation abzugeben. Folglich kann im deutschsprachigen Raum eine zunehmende Orientierung an Standards und Kompetenzen im Bereich der Lehrerbildung beobachtet werden (vgl. z. B. Oser & Oelkers, 2001; KMK, 2004a, b; Hilligus & Rinkens, 2006; Lüders & Wissinger, 2007). Gleichzeitig stellt die Standarddiskussion die empirische Lehrerbildungsforschung vor die große Herausforderung, geeignete Verfahren zu entwickeln, um die gesetzten Ziele empirisch überprüfen zu können und damit differenziert Einsicht in Lern- und Ausbildungsprozesse sowie Wirkungsweisen von Ausbildungsprogrammen und ihren Charakteristika zu erhalten. Nicht zuletzt soll damit der Reformdiskurs und die Wirksamkeitsdebatte zur Lehrerbildung auf eine empirische Basis gestellt werden.

Die Entwicklung geeigneter Verfahren zur Erfassung von Ergebnissen der Lehrerbildung gestaltet sich allerdings als komplexes Unterfangen, unter anderem weil dieses nicht losgelöst von Fragen zur detaillierten Ausrichtung von Ausbildungsstandards betrachtet werden kann: So verlangt etwa die Gliederung einer Standarderreichung in „Minimum“, „Regelfall“ und „Gut“ (KMK, 2004b) konsequenterweise nicht nur konkrete Operationalisierungen von Kompetenzen in Testaufgaben, mit denen anschließend Daten gewonnen und verarbeitet werden sowie ein Kontinuum von Fähigkeitswerten skaliert wird. Die Festlegung von Standarddifferenzierungen dieser Art enthält vielmehr den Anspruch einer *kriteriumsorientierten* Bewertung erreichter Leistungen: Während die kontinuumbezogene Modellierung von Fähigkeitswerten bereits Aussagen über „bessere“ und „schlechtere“ Ergebnisse, insbesondere im sozialen Vergleich, z. B. zwischen Studienanfänger¹ und Studienabsolventen, zulässt, wird die eigentliche Idee der Standardorientierung erst dann vollständig eingelöst, wenn das eingesetzte Messverfahren auch eine Beurteilung erreichter Kompetenzen anhand von Kriterien erlaubt, die eine konkrete Beschreibung der mit dem Testergebnis erfassten Leistung ermöglichen (vgl. bereits Angoff, 1984). In Bezug auf die standardisierte Erfassung von Kompetenzen findet dieser Anspruch seine methodische Realisierung in der Modellierung von „Kompetenzstufen“ bzw. „Kompetenzniveaus“ (Hartig, 2007; Hartig & Klieme, 2006): Mit der Skalierung eines Leistungstests erhält man eine quantitative Messung auf einer kontinuierlichen Skala. Die daraus resultierenden numerischen Werte eignen sich für weitergehende Analysen, beispiels-

¹ Um den Lesefluss zu erleichtern, wird im Folgenden für Lehrerinnen und Lehrer, Schülerinnen und Schüler usw. nur die männliche Geschlechtsform angegeben. Selbstverständlich sind jedoch immer beide Geschlechter gemeint.

weise für Zusammenhangs- oder Unterschiedsanalysen. In Hinblick auf eine Beschreibung der mit dem Test erfassten, spezifischen Kompetenzen, die eine Person X mit dem Wert Y auf der Skala erreicht hat, bleiben die Zahlenwerte jedoch relativ unkonkret. Eine solche Beschreibung für jeden einzelnen Wert auf der Testskala vorzunehmen würde praktisch gesehen einen unangemessenen Aufwand bedeuten (Beaton & Allen, 1992; Hartig, 2007). Es empfiehlt sich daher, die Leistungstestskala in überschaubare Abschnitte zu unterteilen, für die eine kriteriumsorientierte Beschreibung jener Kompetenzen erfolgt, die für die Erreichung eines Skalenwertes in diesem Abschnitt notwendig sind. Die Skalenabschnitte werden dann als „Kompetenzniveaus“ bezeichnet (Hartig, 2007). Vorteile einer solchen Möglichkeit, Personenfähigkeitswerte kriteriumsorientiert zu interpretieren, werden zum einen in theoretischer Hinsicht gesehen – z. B. um vertieften Einblick in die Validität eines untersuchten Konstrukts zu erhalten –, zum anderen in der praktischen Bedeutung für die Leistungsdiagnostik und Evaluation im Bildungsbereich (vgl. Watermann & Klieme, 2006).

Während im Bereich der Kompetenzmessung von Schülern die Modellierung von Kompetenzstufen bzw. Kompetenzniveaus durch die großen Schulleistungstudien wie TIMSS oder PISA sowie die Maßnahmen zur Schulentwicklung (z. B. Lernstandserhebungen) mittlerweile geläufig ist, stellt sie für den Bereich der Kompetenzmessung von angehenden wie berufstätigen Lehrern ein bislang kaum bearbeitetes Forschungsfeld dar. Eine wichtige Ausnahme ist die Studie *Mathematics Teaching in the 21st Century* (MT21; Blömeke, Kaiser & Lehmann, 2008), in der Kompetenzniveaus im fachbezogenen Wissen angehender Mathematiklehrkräfte modelliert werden. Bei der Konstruktion mathematikbezogener Kompetenzniveaus profitiert MT21 sicherlich von bereits geleisteten Vorarbeiten auf Schülerebene. Für den Bereich des fachübergreifenden, pädagogischen Wissens angehender Lehrkräfte – ein Gegenstandsbereich, der grundlegend auf andere Weise strukturiert ist – liegen zwar mittlerweile verschiedene Ansätze zur standardisierten Erfassung vor (Baer et al., 2007; Blömeke, Kaiser & Lehmann, 2008; König & Blömeke, 2009a, b, c; König, Peek & Blömeke, 2008; Schulte, Bögeholz & Watermann, 2008; Seifert, Hilligus & Schaper, 2009); eine Modellierung von Kompetenzniveaus wurde für diesen Bereich hingegen bislang nicht angegangen. Der vorliegende Beitrag möchte dieser defizitären Forschungslage begegnen.

2. Theoretischer Rahmen und Fragestellungen

Zur Bildung von Kompetenzniveaus im pädagogischen Wissen von Lehramtsstudierenden wird das im Rahmen von TEDS-M entwickelte Testinstrument eingesetzt. Dieses berücksichtigt fünf Inhaltsdimensionen, die unter Bezug auf die Bildungsstandards der KMK (2004a) sowie aus einer Verschränkung der Per-

spektiven von Allgemeiner Didaktik und Lehr-Lern-Forschung definiert wurden. Zu diesem inhaltsbezogenen Teil der Testkonzeption – in Abbildung 1 unterhalb des horizontalen Pfeils dargestellt – liegen bereits erste Ergebnisse aus Strukturanalysen vor (vgl. König & Blömeke, 2009a, b), auf die daher nachfolgend nicht weiter eingegangen werden soll. In diesem Beitrag soll hingegen die anforderungsbezogene Struktur des Testinstruments fokussiert werden. Zum einen besteht diese in kognitiven Prozessen, die bei der Bearbeitung der Testaufgaben jeweils im Vordergrund stehen, zum anderen in zentralen Merkmalen der Testaufgaben, die explizit über deren Schwierigkeit bestimmen dürften. Zur Unterscheidung verschiedener Qualitäten der kognitiven Anforderungen, die bei der Bearbeitung der Testaufgaben an die angehenden Lehrkräfte gestellt werden, wurde die von Anderson und Krathwohl (2001) revidierte und erweiterte Bloom'sche Taxonomie kognitiver Prozesse aufgegriffen und auf die Testaufgaben bezogen. Im Fokus stehen dabei drei Dimensionen kognitiver Prozesse: erinnern, verstehen / analysieren, kreieren.

Erinnern. Zur Bearbeitung von Testaufgaben, die diesem kognitiven Prozess zugeordnet werden, wird Wissen benötigt, das in Erinnerung gerufen werden muss (z. B. Lehrbuchwissen und bzw. oder Erfahrungswissen); die Bearbeitung benötigt in dieser Hinsicht kognitive Prozesse wie das Nennen bzw. Erkennen. Testaufgaben dieses Typs fordern bei der Befragung angehende Lehrer auf,

- eine definitorische Nennung zu geben,
- eine Aufzählung von Elementen eines Phänomens, Begriffs oder Konzepts vorzunehmen oder
- einen Begriff oder ein Konzept zu erkennen oder zu identifizieren.

Verstehen / analysieren. Zur Bearbeitung dieser Testaufgaben wird Wissen benötigt, das in Erinnerung gerufen werden muss (z. B. Lehrbuchwissen und bzw. oder Erfahrungswissen), das darüber hinaus aber auch mit einer Problemstellung in Verbindung gebracht werden muss; die Bearbeitung benötigt in dieser Hinsicht kognitive Prozesse wie das Erklären bzw. Vergleichen. Bei der Bearbeitung der Testaufgaben dieser Dimension wurde den angehenden Lehrpersonen abverlangt,

- einen Sachverhalt, ein Konzept, ein Aspekt zu erklären oder zu charakterisieren,
- einen Begriff oder ein Konzept auszuwählen und damit eine Erklärung abzugeben oder
- einen Sachverhalt, eine Situation, eine Falldarstellung, einen oder mehrere Begriffe zu vergleichen, zu kategorisieren, zu ordnen oder zu interpretieren.

Kreieren. Zur Bearbeitung der Aufgaben, die dieser Dimension zugeordnet werden, wird wiederum Wissen benötigt, das in Erinnerung gerufen werden muss (z. B. Lehrbuchwissen und bzw. oder Erfahrungswissen), das darüber hinaus aber auch mit einer Problemstellung und einer typischen Situation in Verbindung ge-

bracht werden muss; im Gegensatz zur vorherigen Dimension des Verstehens und Analysierens wird den Befragten für die Bearbeitung jedoch abverlangt,

- Handlungsoptionen zu entwickeln und zu formulieren, die ihnen für die Bewältigung einer typischen Situation zur Verfügung stehen,
- praktisches Wissen und Können, das als propositionale mentale Repräsentationen beschrieben werden kann, zu explizieren.

Von diesen drei Dimensionen wird jedoch kein expliziter Einfluss auf die Schwierigkeit einer Testaufgabe erwartet, da für jede der drei kognitiven Prozesse Testaufgaben mit einem breiten Spektrum unterschiedlicher Schwierigkeitsgrade denkbar und im Instrument integriert sind. Verknüpft man jedoch die kognitiven Bearbeitungsprozesse mit weiteren Anforderungsmerkmalen, welche die Schwierigkeit einer Testaufgabe explizit beeinflussen dürften, so sind Segmente auf einem Kompetenzkontinuum denkbar wie sie – der Einfachheit halber stark schematisiert – in Abbildung 1 oberhalb des horizontalen Pfeils dargestellt sind. Der horizontale Pfeil steht dabei für die Annahme eines zunehmenden Schwierigkeitsgrades der Testaufgaben. Berücksichtigt wird zum Einen die *Komplexität der kognitiven Bearbeitung*: Muss lediglich ein einzelner Aspekt eines Konzepts oder muss ein komplexer Sachverhalt erinnert und genannt werden? Soll die befragte Person auf eine Problemstellung hin lediglich eine einzelne Handlungsoption kreieren oder werden auch Alternativlösungen in Form vielfältiger Handlungsoptionen verlangt? Zum anderen wird dabei die in den Testaufgaben enthaltene *Terminologie* berücksichtigt: Geschieht die kognitive Bearbeitung von Testaufgaben zur Erfassung von pädagogischem Wissen über Unterricht auf einem sprachlichen Niveau, das angehenden Lehrern bereits aufgrund ihrer schulischen Sozialisation vertraut ist, oder wird die Verwendung einer Professionssprache, wie sie typischerweise erst in der universitären Lehrerausbildung erworben wird, vorausgesetzt?

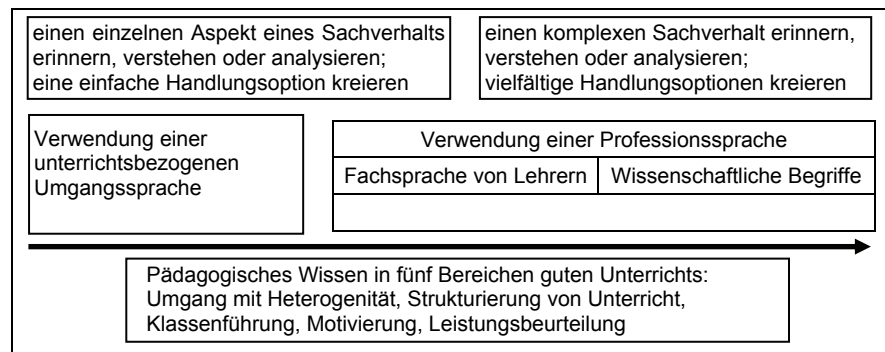


Abbildung 1: Schematische Verknüpfung von Inhaltsdimensionen, kognitiven Prozessen und expliziten Schwierigkeitsmerkmalen zu einem Kompetenzmodell

2.1 Terminologische Anforderungen

Der Gegenstandsbereich des pädagogischen Wissens angehender Lehrkräfte dürfte in großen Teilen nur mithilfe der domänenspezifischen Terminologie zugänglich sein. Dadurch ergibt sich die Notwendigkeit, die terminologischen Anforderungen, die die Testaufgaben implizieren, als Merkmal zur Bestimmung der Aufgabenschwierigkeit aufzugreifen. Aus der Diskussion zum Lehrwissen und zur Lehrersprache (Terhart, 1993) lässt sich eine Teilung der Lehrersprache in drei Niveaus ableiten: *Praxisnahe, unterrichtsbezogene Umgangssprache*, die Lehrkräfte verwenden müssen, um sich mit Schülern z. B. über die Unterrichtsorganisation oder mit Eltern z. B. über die Lernleistung ihrer Kinder zu verständigen. Wörter wie „Gruppenarbeit“, „Stillarbeit“, „Unterrichtstempo“ oder „Lern-diagnose“ sind dafür beispielhaft. Auch wenn solche Wörter zum Teil sogar von der Lehr-Lernforschung als Kategorien zur Beschreibung der Unterrichtswirklichkeit aufgegriffen werden (z. B. Gruehn, 2000; Helmke, 2003), tragen sie für Lehramtsstudierende insofern keinen die Aufgabenschwierigkeit erhöhenden Fachsprachencharakter, da sie den Studierenden aufgrund ihrer langjährigen schulischen Erfahrung geläufig sind. Für angehende Lehrkräfte besitzen diese Wörter daher eher den Charakter einer Umgangssprache. Davon hebt sich das zweite sprachliche Niveau ab, welches abstraktere Termini enthält, die Schülern, Eltern und damit auch Lehramtsstudierenden am Anfang ihrer Ausbildung nicht notwendigerweise aus dem schulpädagogischen Kontext geläufig sind, etwa „Phasenmodelle von Unterricht“, „innere Differenzierung“ oder „intrinsische Motivation“. Dieses zweite Niveau ist aber immer noch so konkret, dass es sich auf eine *Fachsprache der Lehrer* bezieht, die in der Praxis üblicherweise verwendet wird und gegebenenfalls sogar dort zum Teil entstanden ist. Es weist somit einen starken Professionsbezug auf. Dadurch unterscheidet es sich vom dritten sprachlichen Niveau, das eine *wissenschaftlich-spezifische Terminologie* enthält, etwa „Akkomodation“, „operante Konditionierung“ oder „internal-variable Beurteilung von Misserfolgen“. Zwar ist für solche Begriffe kennzeichnend, dass sie im Wissenschaftsdiskurs erzeugt wurden. Sie ermöglichen jedoch Zugang zu spezifischen Theorien, die potenziellen Nutzen bei der professionellen Reflexion berufsbezogener Problemstellungen besitzen. Wissenschaftssprache in diesem Sinne kann als indirekter Indikator für ein breites und vertieftes pädagogisches Wissen über Unterricht aufgefasst werden, welches bei der Bewältigung beruflicher Anforderungen zur Anwendung kommen kann. Aus diesen Überlegungen lässt sich ein terminologisches Anforderungsmerkmal konstruieren, um das sprachliche Anspruchsniveau der eingesetzten Testaufgaben zu bestimmen (Tabelle 1).

Tabelle 1: Stufung der Aufgaben anhand ihrer terminologischen Anforderungen in drei sprachliche Niveaus

Stufe	Sprachliches Niveau	Charakteristika der Aufgaben
3	Wissenschaftssprache	hohes Abstraktionsniveau
2	Fachsprache von Lehrkräften	mittleres Abstraktionsniveau
1	Umgangssprache	niedriges Abstraktionsniveau

2.2 Komplexität kognitiver Bearbeitungsprozesse

Die kognitive Bearbeitung von Testaufgaben in Form der zuvor beschriebenen Prozesse des Erinnerns, Verstehens / Analysierens und Kreierens kann auf mehr oder weniger komplexe Weise geschehen. Unterschiede in der Komplexität kognitiver Bearbeitungsprozesse wurden daher bei der Entwicklung der Testaufgaben des hier verwendeten Testinstruments berücksichtigt. In konzeptioneller Hinsicht geschah dies in Anlehnung an den Ansatz der kognitiven Komplexität (Mandl & Huber, 1978; von Eye, 1999). Eine zentrale Facette der kognitiven Komplexität ist die Anzahl der Dimensionen, „die eine Person bei der Strukturierung oder Beurteilung eines Objekts oder Sachverhalts verwendet.“ (von Eye, 1999, S. 1). Diese Facette wird als „Breite“ bezeichnet. Angenommen wird, dass Personen, die sich in der Komplexität ihrer kognitiven Strukturen unterscheiden, unterschiedlich viele Dimensionen verwenden, um ein Objekt oder einen Sachverhalt zu strukturieren oder zu beurteilen.

Die Integration der Breitenfacette in Testaufgaben zur Bestimmung der Komplexität kognitiver Bearbeitungsprozesse wird insbesondere bei solchen Testaufgaben deutlich, die eine typische Handlungssituation oder -anforderung schildern, zu denen die Befragten anschließend potenzielle Handlungsoptionen kreieren sollen: Wird bei der Beantwortung einer solchen Aufgabe eine vergleichsweise hohe Anzahl an unterschiedlichen Dimensionen (bzw. Handlungsoptionen) herangezogen, so darf ein breiter Wissensumfang vermutet werden, über den diese Person verfügt. Wird dagegen eine vergleichsweise kleine Anzahl an solchen Dimensionen (bzw. Handlungsoptionen) verwendet, so dürfte die Person nur über einen schmalen Wissensumfang verfügen.

Dies sei an einer beispielhaften Testaufgabe illustriert (Abbildung 2). In dieser Aufgabe wird nach drei Unterrichtsmethoden bzw. -konzepten gefragt, welche die Teilhabe der Schüler an Prozessen der inneren Differenzierung im Unterricht ermöglichen sollen. Werden nun drei solche Unterrichtsmethoden bzw. -konzepte genannt bzw. kreiert (z. B. (1) Wochenplan, (2) Stationslernen, (3) Lernbuffet), so stellt die befragte Person ein breiteres Wissen unter Beweis als eine Person, die

lediglich ein oder zwei solcher Methoden bzw. Konzepte nennen bzw. kreieren kann.²

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)

Abbildung 2: Testaufgabe, für deren Lösung ein breiter Wissensumfang benötigt wird (Quelle: König & Blömeke, 2009b)

Da der Schwierigkeitsgrad einer Aufgabe ein höherer ist, wenn ein breiter Wissensumfang unter Beweis gestellt werden muss, als wenn ein schmaler gefordert wird, lässt sich daraus ein Merkmal zur Bestimmung der Aufgabenschwierigkeit ableiten (Tabelle 2): Die mit Testaufgaben bzw. Testitems definierten Anforderungen, deren erfolgreiche Bewältigung lediglich einen – im Sinne der Breitenfacette der kognitiven Komplexität – schmalen Wissensumfang erfordern, stellen ein niedriges Komplexitätsniveau kognitiver Bearbeitung dar. Sie werden daher einer ersten Stufe zugeordnet. Anforderungen hingegen, die darüber hinaus einen breiteren Wissensumfang erfordern, stellen ein hohes Komplexitätsniveau dar und werden somit einer zweiten Stufe zugeordnet. Auf diese Weise kann die Komplexität kognitiver Bearbeitungsprozesse operationalisiert werden.

Tabelle 2: Stufung der Aufgaben anhand des benötigten Wissensumfangs in zwei Komplexitätsniveaus

Stufe	Benötigter Wissensumfang	Charakteristika der Aufgaben
2	breit	hohes Komplexitätsniveau kognitiver Bearbeitung
1	schmal	niedriges Komplexitätsniveau kognitiver Bearbeitung

² Gegebene Antworten wurden mithilfe eines differenzierten Kategoriensystems niedrig-inferent kodiert. Das niedrig-inferente Kategoriensystem wurde im Rahmen eines Expertenreviews zu größeren Wertungskategorien gebündelt.

2.3 Fragestellung und Hypothesen

Gegenstand dieses Beitrages ist die Exploration einer möglichen Strategie, Kompetenzniveaus im pädagogischen Wissen von Lehramtsstudierenden zu bilden. Dabei werden folgende Fragen fokussiert:

Korrespondieren die theoretisch angenommenen Schwierigkeitsmerkmale der Aufgaben bzw. Items mit den empirischen Aufgabenschwierigkeiten bzw. Itemschwierigkeiten?

Die Annahme vorausgesetzt, dass Lehrerausbildung wirksam und das Testinstrument in Bezug auf die Lehrerausbildung inhaltlich valide ist, sollten höhere Itemschwierigkeitswerte durch ein höheres sprachliches Niveau, welches in der Lehrerausbildung erworben wird, bestimmt werden; ferner sollten Aufgaben, die komplexere kognitive Bearbeitungsprozesse verlangen, einen höheren Schwierigkeitsgrad aufweisen als Aufgaben, deren Bearbeitung eine weniger komplexe Bearbeitung erfordern.

Können die theoretisch angenommenen Kompetenzniveaus Leistungsunterschiede zwischen Studierenden abbilden, die sich in unterschiedlichen Stadien ihrer Ausbildung befinden?

Wiederum ausgehend von der Annahme einer Wirksamkeit der Ausbildung und der inhaltlichen Validität des Testinstruments sollten Studierende, die sich in einem fortgeschrittenen Stadium ihrer Ausbildung befinden, häufiger ein höheres Kompetenzniveau erreichen als ihre Kommilitonen aus einem früheren Ausbildungsstadium.

3. Methode

3.1 Stichprobe

Verwendet wird eine Stichprobe von 197 Lehramtsstudierenden der Universität zu Köln, die im Mai und Juni 2008 im Rahmen von Lehrveranstaltungen unter Beobachtung und einheitlicher Instruktion befragt bzw. getestet wurden (zu Details der Stichprobenzusammensetzung und Datenerhebung vgl. König, Peek & Blömeke, 2008). Die Stichprobe wurde entsprechend ihres Ausbildungsstadiums anhand ihrer Semesteranzahl in drei Kohorten unterteilt: Kohorte 1 mit vier oder weniger Semestern (Grundstudium), Kohorte 2 mit fünf bis sieben Semestern (Hauptstudium) und Kohorte 3 mit acht oder mehr Semestern (Examenskandidaten). Tabelle 3 enthält die Fallzahlen der Untersuchungstichprobe für die drei Kohorten getrennt nach Ausbildungsgang.

Tabelle 3: Fallzahl für Befragte in verschiedenen Ausbildungsstadien, gegliedert nach Ausbildungsgang (König et al., 2008)

Kohorte	Semesteranzahl	Bezeichnung	GHRG	SP	GYG	Gesamt
(1)	4 oder weniger	Grundstudium	21	9	8	38
(2)	5 bis 7	Hauptstudium	44	31	21	96
(3)	8 oder mehr	Examenskandidaten	11	21	31	63
Gesamt			76	61	60	197

GHRG – Grund-, Haupt-, Real- und Gesamtschule, SP – Sonderpädagogik, GYG – Gymnasium und Gesamtschule.

3.2 Testinstrument und Skalierung

Eingesetzt wird das Testinstrument zur Erfassung des pädagogischen Wissens von angehenden Lehrkräften aller Schulstufen und Unterrichtsfächer, das im Rahmen der Studie „*Teacher Education and Development: Learning to Teach Mathematics*“ (TEDS-M; Tatto et al., 2008; Blömeke et al., 2009) entwickelt worden ist (König & Blömeke, 2009b). Details zur Testentwicklung sowie Testaufgabenbeispiele mit Lösungsantworten sind ausführlich in König und Blömeke (2009a, b, c) dokumentiert, genaue Angaben zur Skalierung des Leistungstests an der hier einbezogenen Stichprobe finden sich in König, Peek und Blömeke (2008).³ Hervorzuheben für den vorliegenden Beitrag ist dabei, dass der eingesetzte Leistungstest mithilfe der Software „Conquest“ (Wu, Adams & Wilson, 1997) einer eindimensionalen Rasch-Skalierung unterzogen wurde, mit der Aufgaben- bzw. Item-Schwierigkeitsparameter einerseits und Personenfähigkeitswerte andererseits auf einem gemeinsamen Kontinuum abgetragen werden (vgl. Rost, 1996). Der dadurch mögliche Vergleich von individuellen Messwerten mit den Schwierigkeiten der eingesetzten Testaufgaben stellt eine wichtige Voraussetzung für die Bildung von Kompetenzniveaus dar (vgl. Hartig, 2007; Hartig & Klieme, 2006).

3.3 Zuordnung der Aufgabenmerkmale zu den Testitems

Zunächst wurden die einbezogenen Testaufgaben auf Termini hin analysiert, die sich den drei sprachlichen Niveaus zuordnen lassen bzw. charakteristisch für diese sind. Bei offenen Aufgaben wurde der Aufgabenstamm, bei geschlossenen Aufgaben wurden neben dem Aufgabenstamm auch die Antwortoptionen in die Analyse einbezogen. Auf diese Weise wurden 86 relevante Termini zu einem

³ Die dort beschriebene Skalierung wird im Wesentlichen auch für die Analysen dieses Beitrages übernommen. Lediglich die *partial-credit* Items wurden in dichotome Items umgewandelt (d. h. vollständige Lösung mit 1, Teillösung mit 0 rekodiert) und das Skalierungsmodell neu berechnet. Die aus dem neuen Skalierungsmodell resultierenden WLE-Werte korrelieren mit den WLE-Werten, die in den Analysen von König et al. (2008) berichtet werden, fast perfekt ($r = .98$).

Glossar zusammengestellt und von acht in der Lehrerausbildung tätigen Lehrkräften, Erziehungswissenschaftlern und Psychologen jeweils einem der drei Niveaus zugeordnet.⁴ Die acht unabhängig voneinander gegebenen Ratings zeigen mit einer durchschnittlichen Korrelation von $r = .73$ ($SD = .05$, $Min = .66$, $Max = .88$) und einem Cronbachs Alpha von $\alpha = .96$ eine gute bis sehr gute Übereinstimmung, sodass in einem ersten Schritt über den (gegebenenfalls gerundeten) arithmetischen Mittelwert der acht Ratings eine reliable Zuordnung der 86 Termini zu einem der drei sprachlichen Niveaus vorgenommen werden konnte. In einem zweiten Schritt wurden die Testaufgaben bzw. Testitems mithilfe der in ihnen enthaltenen Termini einem der drei sprachlichen Schwierigkeitsniveaus zugeordnet. Ferner wurden die Testitems hinsichtlich der Komplexität der jeweils erforderlichen Bearbeitungsprozesse als niedrig oder als hoch eingestuft. In Tabelle 4 ist die Verteilung der 75 dichotomen Items nach angenommenem sprachlichen Niveau und Komplexität kognitiver Bearbeitungsprozesse dargestellt. Da die Antworten dieser Untersuchungsstichprobe auf viele der Testaufgaben mit offenem Antwortformat insgesamt unterkomplex waren, konnten für diese Testaufgaben in der Skalierung keine komplexe, sondern lediglich eine einfache Wertung Berücksichtigung finden. Aus diesem Grund stellen Testitems, welche komplexe kognitive Bearbeitungsprozesse erfordern, eine relativ kleine Itemgruppe dar. Um die Vorhersage der empirischen Itemschwierigkeiten mit theoretisch angenommenen Merkmalen dennoch auf hinreichend viele Items beziehen zu können, werden daher im Folgenden lediglich vier Itemgruppen unterschieden: umgangssprachlich / niedrig (UN, 25 Items), fachsprachlich / niedrig (FN, 25 Items), wissenschaftlich/niedrig (WN, 16 Items) und umgangs- und fachsprachlich / hoch (UFH, 9 Items).

Tabelle 4: Verteilung der einbezogenen Items nach schwierigkeitsbestimmenden Merkmalen

		Komplexität kognitiver Bearbeitungsprozesse		Gesamt
		niedrig	hoch	
Terminologie	umgangssprachlich	25	3	28
	fachsprachlich	25	6	31
	wissenschaftlich	16	0	16
Gesamt		66	9	75

⁴ Den acht Expertinnen und Experten sei an dieser Stelle für ihre Unterstützung der Studie herzlich gedankt.

4. Ergebnisse

Die Verteilungen der empirischen Itemschwierigkeiten getrennt nach Itemgruppe unterscheiden sich (Abbildung 3 und Tabelle 5). Betrachtet man den jeweiligen Median der Itemgruppen, so lässt sich ein Anstieg in der Schwierigkeit nach sprachlichem Niveau interpretieren: Umgangssprachlich formulierte Items sind insgesamt einfacher zu beantworten als fachsprachlich formulierte, während Items, die eine wissenschaftliche Terminologie enthalten, schwieriger zu bearbeiten sind als fachsprachlich formulierte Items. Allerdings überlappen sich auch die Verteilungen der drei sprachlichen Itemgruppen, insbesondere die Gruppe der umgangssprachlichen und die Gruppe der fachsprachlichen Items. Die Gruppe von Items, welche komplexe kognitive Bearbeitungsprozesse erfordern, hebt sich hingegen deutlich von den ersten drei Itemgruppen ab und erweist sich somit als trennschärfer.

Da die Zuordnung der Items zu einer der vier Gruppen global erfolgte, empfiehlt es sich im Folgenden, die Abschnitte auf der Skala nach einem von Hartig (2007) beschriebenen und in DESI (vgl. Beck & Klieme, 2007) angewendeten Verfahren zu unterteilen: Als Schwellenwert wird (bei gerader Itemanzahl) die empirische Schwierigkeit des Medians oder (bei ungerader Anzahl) die empirische Schwierigkeit jenes Items verwendet, das sich unterhalb des Medians befindet. Auf diese Weise lassen sich die in Tabelle 5 aufgeführten Schwellenwerte festlegen. Zusätzlich sind in Tabelle 5 Kurzbeschreibungen der auf diese Weise gebildeten Kompetenzniveaus dokumentiert, die sich auf die inhaltlichen Definitionen kognitiver Bearbeitungsprozesse, welche unter Punkt 2 gegeben wurden, beziehen.⁵ Dabei gilt zu beachten, dass das jeweils höhere Niveau auch die Beherrschung des jeweils vorhergehenden Niveaus einschließt.

⁵ Auf sprachlich differenzierte Formulierungen der Kompetenzniveaus wird daher an dieser Stelle verzichtet.

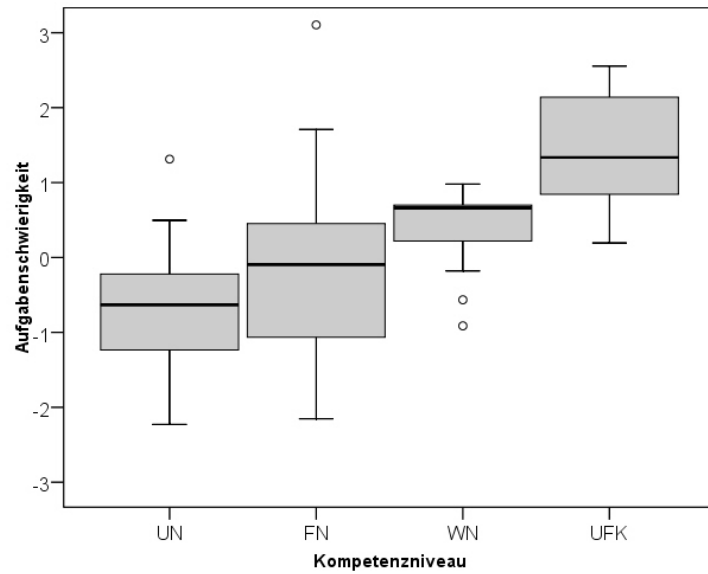


Abbildung 3: Verteilung der Itemschwierigkeiten getrennt nach Niveau

Tabelle 5: Kurzbeschreibung der Niveaus, Median und Streuung der Itemschwierigkeiten

Niveau	Itemschwierigkeiten	Median	Schwellenwert	Kurzbeschreibung der Kompetenzniveaus
Unterhalb I	-2.2 bis 1.3	-.63	–	Hinreichend beherrscht werden Testaufgaben zu pädagogischem Wissen über Unterricht mit den folgenden Anforderungen:
I	-2.2 bis 3.1	-.09	-.355	einfache kognitive Bearbeitungsprozesse auf umgangssprachlichem Niveau
II	-.91 bis .98	.66	.66	einfache kognitive Bearbeitungsprozesse auf fachsprachlichem Niveau
III	.19 bis 2.56	1.34	1.125	einfache kognitive Bearbeitungsprozesse auf wissenschaftlichem Niveau
				komplexe kognitive Bearbeitungsprozesse auf umgangssprachlichem und fachsprachlichem Niveau

Um vertieften Einblick in die so vorgenommene Unterteilung der Skala in Kompetenzniveaus zu erhalten, ist die Verteilung der befragten Studierenden auf die Niveaus von Interesse. Da sich in vorangegangenen Analysen deutliche Differenzen zwischen pädagogisch orientierten und fachwissenschaftlichen Studiengängen zeigten (vgl. dazu detailliert König, Peek & Blömeke, 2008), sollen im Folgenden die Häufigkeitsverteilungen entsprechend getrennt für beide Ausbildungsganggruppen dargestellt werden. Die Studierenden der pädagogisch orientierten Studiengänge GHRG und SP verteilen sich erwartungskonform auf die vier Niveaus (Abbildung 4): Etwa ein Viertel der Studierenden der ersten Kohorte erreichen Niveau I, drei Viertel verbleiben unter diesem ersten Niveau. Mehr als die Hälfte der Studierenden der zweiten Kohorte erreichen bereits Niveau I, während rund ein Drittel der dritten Kohorte Niveau II bzw. Niveau III erreicht. Die Gymnasialstudierenden dagegen verbleiben maximal auf Niveau I und kommen nicht über dieses hinaus (Abbildung 5).

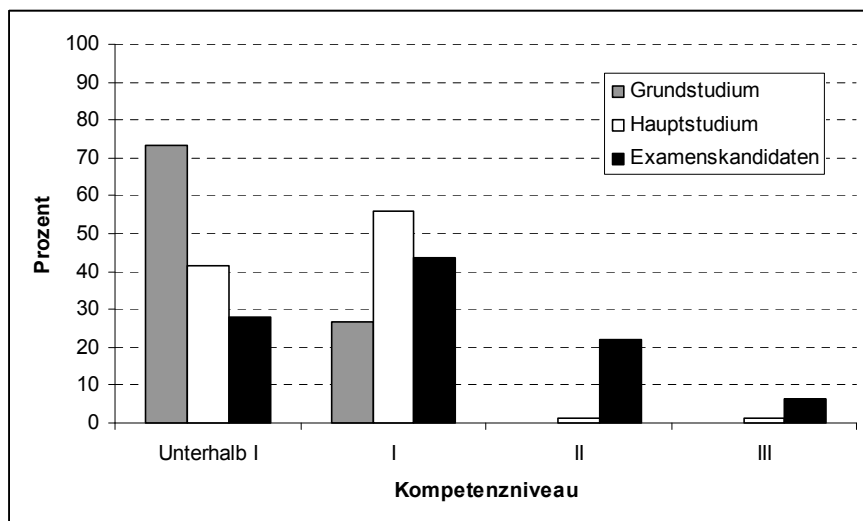


Abbildung 4: Verteilung der GHRG- und SP-Studierenden auf Kompetenzniveaus, gegliedert nach Kohorte

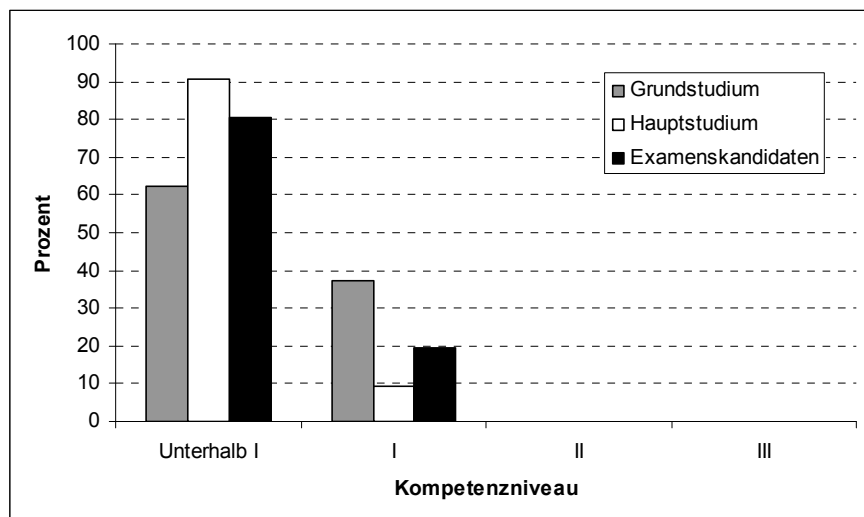


Abbildung 5: Verteilung der GYG-Studierenden auf Kompetenzniveaus, gegliedert nach Kohorte

5. Diskussion

Die Debatte um eine stärkere Orientierung an Standards und Kompetenzen in der Lehrerausbildung erfordert die Entwicklung geeigneter Verfahren, anhand derer das Erreichen zuvor festgelegter Ziele empirisch überprüft werden kann. Eine kriteriumsorientierte Evaluation erreichter Ausbildungsleistungen erfordert aus methodischer Sicht eine Modellierung von Kompetenzniveaus wie sie in Schulleistungsstudien bereits häufig Anwendung findet. Da es in der empirischen Lehrerausbildungsforschung bislang an entsprechenden Modellierungsansätzen mangelt, wurde im vorliegenden Beitrag eine mögliche Strategie zur Modellierung von Kompetenzniveaus im pädagogischen Wissen von Lehramtsstudierenden exploriert.

Die Testaufgaben des im Rahmen von TEDS-M entwickelten Tests zur Erfassung des pädagogischen Wissens von angehenden Lehrkräften (König & Blömeke, 2009b) wurden anhand zweier Merkmale – terminologische Anforderungen der Aufgaben und Komplexität kognitiver Bearbeitungsprozesse – klassifiziert, von denen ein expliziter Einfluss auf die Schwierigkeit der Testaufgaben angenommen wurde. Anschließend wurden die empirisch ermittelten Itemschwierigkeiten auf die theoretisch angenommenen zurückgeführt. Auf diese Weise konnten nach einem in DESI beschriebenen und verwendeten Verfahren (vgl. Beck &

Klieme, 2007) drei Kompetenzniveaus gebildet werden. Die Verteilung von Lehramtsstudenten unterschiedlicher Ausbildungsgänge und unterschiedlicher Ausbildungsstadien verweisen auf die inhaltliche Validität der gebildeten Kompetenzniveaus: Studierende, die pädagogisch orientierte Studiengänge verfolgen, erreichen in fortgeschrittenen Ausbildungsstadien auch zunehmend häufiger die Kompetenzniveaus II und III; Studierende, deren Studiengang stark fachwissenschaftlich ausgerichtet ist und einen geringeren Umfang pädagogischer Lerngelegenheiten vorsieht, verbleiben dagegen auf Kompetenzniveau I (vgl. zu einer differenzierten Erläuterung der Leistungsunterschiede nach Studiengang König, Peek & Blömeke, 2008).

Die Ergebnisse stützen die theoretischen Annahmen aus der Diskussion zur Qualität der Lehrersprache, der zufolge die Terminologie einer Lehrerfachsprache anspruchsvolleren Charakter trägt als eine von Lehrern und Schülern zur gegenseitigen Verständigung über Unterricht verwendete Umgangssprache, während eine wissenschaftliche Terminologie wiederum eine höhere Schwierigkeit aufweist als die Terminologie der Fachsprache (Terhart, 1993). Dass die Kenntnis wissenschaftlicher Begriffe eine höhere Schwierigkeit impliziert als die Beherrschung einer professionsbezogenen Fachsprache könnte aber auch auf den hohen Spezialisierungsgrad zurückgeführt werden, der einer wissenschaftlichen Terminologie im Rahmen der universitären Lehrerausbildung zugeschrieben wird. Zukünftige Studien sollten der Frage nachgehen, inwieweit die Unterscheidung zwischen der Terminologie der Fachsprache der Lehrer und der wissenschaftlichen Terminologie auch in der zweiten Ausbildungsphase, die wesentlich praxisnäher als die erste Phase gestaltet ist und somit eine größere Nähe zur Fachsprache von Lehrkräften aufweisen dürfte, ebenfalls von Bedeutung ist. Ferner zeigen die Ergebnisse, dass die Verknüpfung kognitiver Bearbeitungsprozesse mit dem Merkmal der Komplexität eine Operationalisierung von Testitemschwierigkeit ermöglicht. Höhere Fähigkeitswerte im Test gehen einher mit Testaufgaben, die eine komplexe kognitive Bearbeitung verlangen.

Die hier vorgestellte Strategie verdeutlicht somit die Bedeutung der sprachlichen Qualität sowie der Komplexität kognitiver Bearbeitungsprozesse, welche als Merkmale von Testaufgaben zur Erfassung des pädagogischen Wissens von angehenden Lehrkräften berücksichtigt und für die kriteriumsorientierte Beschreibung erreichter Ausbildungsleistungen hinzugezogen werden können. Gleichzeitig werden aber auch Grenzen des hier dargestellten Vorgehens deutlich, die auf die Notwendigkeit zukünftiger Studien, die sich der Modellierung des pädagogischen Wissens widmen, verweisen: Zunächst gilt zu berücksichtigen, dass die hier explorierte Strategie der Bildung von Kompetenzniveaus in erster Linie der Beschreibung der im verwendeten Testinstrument integrierten Aufgaben dient.

Diese Bindung an die eingesetzten Testaufgaben erlaubt keine Generalisierung des formulierten Analysemodells auf andere Testaufgaben oder -instrumente (vgl. Hartig, 2007). Positiv gewendet wäre es also beispielsweise eine interessante Fragestellung, ob auch in anderen Testinstrumenten zur Erfassung pädagogischen Wissens die in den jeweiligen Testaufgaben verwendete Terminologie ein Merkmal darstellt, welches über die Schwierigkeit von Testaufgaben bzw. Testitems bestimmt. Ferner sollten die hier explorierten Merkmale zur Beschreibung der Aufgabenanforderungen um weitere Merkmale ergänzt werden, um zu einem differenzierteren Kompetenzmodell zu gelangen. Ein solches Kompetenzmodell wird derzeit in der vom Autor geleiteten und von der DFG geförderten Studie *Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden* (LEK) entwickelt.

Literatur

- Anderson, L. W. & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Longman.
- Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service.
- 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.
- Beaton, A. E. & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17, 191-204.
- Beck, B. & Klieme, E. (Hrsg.). (2007). Sprachliche Kompetenzen. Konzepte und Messung. Weinheim: Beltz.
- 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. & Hacke, S. (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.
- Eye, A. v. (1999). Kognitive Komplexität. Messung und Validität. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2, 81-96.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen*. Münster: Waxmann.

- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), Sprachliche Kompetenzen. Konzepte und Messung (S. 83-99). Weinheim: Beltz.
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), Leistung und Leistungsdiagnostik (S. 128-143). Heidelberg: Springer.
- Helmke, A. (2003). Unterrichtsqualität erfassen, bewerten, verbessern. Seelze: Kallmeyersche Verlagsbuchhandlung.
- Hilligus, A. H. & Rinkens, H.-D. (Hrsg.). (2006). Standards und Kompetenzen – neue Qualität in der Lehrerbildung? Berlin: LIT.
- KMK (2004a). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004. Verfügbar unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- KMK (2004b). Standards für die Lehrerbildung: Bericht der Arbeitsgruppe.
- König, J. & Blömeke, S. (2009a). Disziplin- oder Berufsorientierung? Zur Struktur des pädagogischen Wissens angehender Lehrkräfte. *Lehrerbildung auf dem Prüfstand*, 2, 126-147.
- König, J. & Blömeke, S. (2009b). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerbildung. *Zeitschrift für Erziehungswissenschaft*, 12, online first (Springer Verlag).
- König, J. & Blömeke, S. (2009c). Pädagogisches Wissen von österreichischen Lehramtsstudierenden. *Erziehung & Unterricht*, 159, 175-186.
- König, J., Peek, R. & Blömeke, S. (2008). Zum Erwerb von pädagogischem Wissen in der universitären Ausbildung: Unterscheiden sich Studierende verschiedener Lehrämter und Kohorten? *Lehrerbildung auf dem Prüfstand*, 1, 639-657.
- Lüders, M. & Wissinger, J. (Hrsg.). (2007). *Forschung zur Lehrerbildung: Kompetenzentwicklung und Programmevaluation*. Münster: Waxmann.
- Mandl, H. & Huber, G. L. (Hrsg.). (1978). *Kognitive Komplexität. Bedeutung – Weiterentwicklung – Anwendung*. Göttingen: Verlag für Psychologie.
- Oser, F. & Oelkers, J. (Hrsg.). (2001). *Die Wirksamkeit der Lehrerbildungssysteme*. Chur: Rüegger.
- Rost, J. (1996). *Testtheorie, Testkonstruktion*. Bern: Huber.
- Schulte, K., Bögeholz, S. & Watermann, R. (2008). Selbstwirksamkeitserwartungen und Pädagogisches Professionswissen im Verlauf des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft*, 11, 268-287.

- Seifert, A., Hilligus, A. H. & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstrumentes zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2, 82-103.
- 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.
- Terhart, E. (1993). Pädagogisches Wissen. Überlegungen zu seiner Vielfalt, Funktion und sprachlichen Form am Beispiel des Lehrerwissens. In J. Oelkers & H.-E. Tenorth (Hrsg.), *Pädagogisches Wissen* (S. 129-141). Weinheim: Beltz.
- Watermann, R. & Klieme, E. (2006). Modellierung von Kompetenzstufen mit Hilfe der latenten Klassenanalyse. *Empirische Pädagogik*, 20, 321-336.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software [computer program]*. Camberwell, Vic.: Australian Council for Educational Research.

Anschrift des Autors:

Dr. Johannes König, Humboldt-Universität zu Berlin, Philosophische Fakultät IV, Institut für Erziehungswissenschaften, Unter den Linden 6, 10099 Berlin,
E-Mail: Johannes.Koenig@staff.hu-berlin.de