pedocs
Open Access Erziehungswissenschaften

DIPF

Klemenz, Stefan; König, Johannes

Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften. Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung

Zeitschrift für Pädagogik 65 (2019) 3, S. 355-377

Quellenangabe/ Reference:
Klemenz, Stefan; König, Johannes: Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften. Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung - In: Zeitschrift für Pädagogik 65 (2019) 3, S. 355-377 - URN: urn:nbn:de:0111-pedocs-239470 - DOI: 10.25656/01:23947

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-239470
https://doi.org/10.25656/01:23947

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

http://www.juventa.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
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der Leibniz-Gemeinschaft

D 7484

ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 3 Mai/Juni 2019

■ *Aktuelle Forschungsdiskurse*

… zur inklusiven Schule

… zur LehrerInnenbildung und zum Lehrberuf

… zur Digitalisierung von Bildung

BELTZ JUVENTA

ZEITSCHRIFT FÜR PÄDAGOGIK
Jahrgang 65 – Heft 3
Mai/Juni 2019

Inhaltsverzeichnis

Aktuelle Forschungsdiskurse

… zur inklusiven Schule

Adina Küchler/Alina Ivanova
Anerkennung als Voraussetzung für inklusives und migrationspädagogisches Handeln in der Schule 315

Jan Jochmaring
Übergänge von Schüler/innen mit sonderpädagogischem Förderbedarf in die Berufsausbildung. Eine Auswertung von Sekundärstatistiken 335

… zur LehrerInnenbildung und zum Lehrberuf

Stefan Klemenz/Johannes König
Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften. Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung 355

Tobias Kärner/Caroline Bonnes/Christian Schölzel
Bewertungstransparenz im Referendariat 378

Colin Cramer/Martin Harant/Samuel Merk/Martin Drahmann/ Marcus Emmerich
Meta-Reflexivität und Professionalität im Lehrerinnen- und Lehrerberuf 401

… zur Digitalisierung von Bildung

Sigrid Hartong
Bildung 4.0? Kritische Überlegungen zur Digitalisierung von Bildung als erziehungswissenschaftliches Forschungsfeld 424

Diskussion

Fritz Bohnsack
Noch einmal „Anders als gedacht". Ein Kommentar zu Michael Knolls Dewey-Aufsatz in der Zeitschrift für Pädagogik (2018, Heft 5) 445

Besprechungen

Johannes Drerup
Anke Redecker: Das kritische Selbst. Bildungstheoretische Reflexionen im Anschluss an Hugo Gaudig, Marian Heitger, Käte Meyer-Drawe und Immanuel Kant 451

Heinz-Elmar Tenorth
Hartmut Rosa/Wolfgang Endres: Resonanzpädagogik. Wenn es im Klassenzimmer knistert (2. Aufl., mit einem Nachwort von Reinhart Kahl).
Hartmut Rosa/Wolfgang Endres/Jens Beljan: Resonanz im Klassenzimmer. 48 Impulskarten zur Resonanzpädagogik mit 16-seitigem Booklet.
Hartmut Rosa/Claus G. Buhren/Wolfgang Endres: Resonanzpädagogik und Schulleitung. Neue Impulse für die Schulentwicklung.
Hartmut Rosa: Resonanz. Eine Soziologie der Weltbeziehung 453

Berno Hoffmann
Eckart Conze/Susanne Rappe-Weber (Hrsg.): Die deutsche Jugendbewegung. Historisierung und Selbsthistorisierung nach 1945 (Jugendbewegungen und Jugendkulturen, Bd. 14) 458

Ulrich Heimlich
Tanja Sturm/Monika Wagner-Willi (Hrsg.): Handbuch schulische Inklusion 461

Jane Weiß
Albert Scherr/Lena Sachs: Bildungsbiographien von Sinti und Roma. Erfolgreiche Bildungsverläufe unter schwierigen Bedingungen 463

Dokumentation

Pädagogische Neuerscheinungen 467

Impressum U3

Table of Contents

Current Research Discourses

… on Inclusive Schools

Adina Küchler/Alina Ivanova
Recognition as a Prerequisite for Inclusive and Migration Pedagogical Action in School 315

Jan Jochmaring
Transitions of Students with Special Educational Needs into Vocational Training: an evaluation of secondary statistics 335

… on Teacher Training and the Teaching Profession

Stefan Klemenz/Johannes König
Modeling Proficiency Levels of Pre-Service Teachers' Pedagogical Knowledge: On the criterial description of cross-subject teacher education outcomes 355

Tobias Kärner/Caroline Bonnes/Christian Schölzel
Assessment Transparency in Teacher Training 378

Colin Cramer/Martin Harant/Samuel Merk/Martin Drahmann/ Marcus Emmerich
Meta-Reflexivity and Professionalism in the Teaching Profession 401

… on Digitalisation of Education

Sigrid Hartong
Education 4.0? Critical reflections on the digitalisation of education as an educational research field 424

Discussion

Fritz Bohnsack
Once again "Different than thought". A comment on Michael Knoll's article on Dewey in this periodical (2018, 5) 445

Book Reviews 451

New Books 467

Impressum U3

Stefan Klemenz/Johannes König

Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften

Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung

Zusammenfassung: In der aktuellen Lehrer*innenbildungsforschung spielt eine kriteriumsbezogene Testwertinterpretation zur Beschreibung von Kompetenzen nur eine untergeordnete Rolle, obwohl sie im Vergleich zu normbezogenen Aussagen als angemessener bewertet wird. Eine an Standards ausgerichtete Lehrer*innenausbildung erfordert jedoch die inhaltliche Beschreibung von Anforderungsniveaus, weil erst dadurch konkrete Aussagen darüber getroffen werden können, welche Personen über welche Kompetenzen verfügen. Der Beitrag begegnet dem Desiderat, indem ein Kompetenzniveaumodell im pädagogischen Wissen von angehenden Lehrkräften konstruiert und geprüft wird. In die Analyse wird eine Stichprobe von 1451 angehenden Lehrkräften aus Deutschland und Österreich einbezogen, deren pädagogisches Wissen über zwei Messzeitpunkte längsschnittlich mithilfe des im Rahmen von TEDS-M entwickelten Tests erfasst wurde. In das Kompetenzniveaumodell werden auf Grundlage theoretischer Ansätze die Aufgabenmerkmale Komplexität kognitiver Bearbeitungsprozesse und terminologische Anforderungen integriert. Sie können einen bedeutsamen Anteil der Varianz in den Aufgabenschwierigkeiten aufklären und werden zur Definition von Kompetenzniveaus herangezogen. Die Verteilung der angehenden Lehrkräfte auf die gebildeten Kompetenzniveaus fällt erwartungskonform aus und verweist auf eine hohe Änderungssensitivität des Modells. Ferner unterstreichen die Befunde der vorliegenden Analyse die Konstruktvalidität der Testwertinterpretation.

Schlagworte: Professionelles Wissen, Lehrer*innenbildung, Kompetenzniveaus, Längsschnittanalyse, Pädagogisches Wissen

1. Einleitung

Im Zuge der Ergänzung einer Input- um eine Output-Orientierung in Bildungssystemen findet eine zunehmende Ausrichtung an Standards und Kompetenzen auch in der Ausbildung von Lehrer*innen statt (vgl. König, 2009). Die von der KMK formulierten Standards in den bildungswissenschaftlichen Studien der Lehrer*innenausbildung (vgl. KMK, 2004) sind trotz Kritik „als strategischer und argumentativer Bezugspunkt der Arbeit an den Strukturen und Curricula dieses Elements der Lehrerbildung unbestritten wichtig und nachhaltig geworden“ (Terhart, 2014, S. 315) und bilden eine zentrale normative Bezugsgröße der Kompetenzen, die Absolvent*innen bei Berufseintritt erlangt haben sollen.

Reformprozesse der Lehrer*innenbildung haben den Bedarf kenntlich gemacht, die intendierten Ziele empirisch zu überprüfen und detaillierte Einblicke in Lern- und Ausbildungsprozesse zu erlangen (vgl. König, 2010). Damit geht die Notwendigkeit einher, geeignete Modellierungs- und Messverfahren zu entwickeln. In den vergangenen Jahren ist eine deutliche Zunahme an Forschungsansätzen zur standardisierten Erfassung von Lehrer*innenkompetenzen festzustellen, die das pädagogische Wissen im Anschluss an Shulman (1987) übereinstimmend als wesentliche Facette professioneller Kompetenz ansehen (vgl. Voss, Kunina-Habenicht, Hoehne & Kunter, 2015). Im Laufe der Ausbildung sollen angehende Lehrer*innen dieses generische, pädagogische Wissen erlangen, um auf ihren Beruf vorbereitet zu werden (vgl. Terhart, 2012).

Um die so implizierte Annahme zum Erwerb von pädagogischem Wissen, das als fächerübergreifendes, unterrichtsbezogenes Wissen verstanden wird, zu prüfen, wird im Rahmen der Studie *Entwicklung von berufsspezifischer Motivation und pädagogischem Wissen in der Lehrer*innenausbildung* (EMW) in Deutschland, Österreich und der Schweiz neben motivationalen Aspekten das pädagogische Wissen angehender Lehrpersonen, die 2011 ihre Ausbildung starteten, längsschnittlich im Turnus von zwei Jahren erfasst.[1] Eingesetzt wird dazu ein Testinstrument, das bereits in der TEDS-M Studie entwickelt und angewendet wurde (vgl. König, Blömeke, Paine, Schmidt & Hsieh, 2011). Die mittels derartiger Instrumente gewonnenen Personenfähigkeitswerte werden in der Regel normorientiert interpretiert und eignen sich dazu, mit anderen Variablen in Beziehung gesetzt oder gruppenspezifisch verglichen zu werden (vgl. Hartig, 2007). Anhand einer normorientierten Messung lässt sich die Erreichung von Standards allerdings nur unzureichend prüfen, da keine Aussage darüber getroffen werden kann, ob eine bestimmte Teilpopulation ein bestimmtes Kriterium erreicht hat (vgl. Rost, 2004a). Kompetenzniveaumodelle können diesem Defizit begegnen, da sie einen Bezug zwischen numerischen Testwerten und konkreten, fachbezogenen Kompetenzen herstellen. Diese sogenannte kriteriumsorientierte Testwertinterpretation wird häufig als angemessener angesehen, jedoch ist sie bis heute in der Lehrer*innenbildungsforschung kaum realisiert worden (vgl. Schaper, 2009).

Während Kompetenzniveaumodellierungen in Schulleistungsstudien als etabliertes Verfahren bekannt sind und bestimmte Diskussionen erst möglich gemacht haben – z. B. die Ergebnisinterpretation sog. „Risikoschüler" in PISA und daran anknüpfende Handlungsempfehlungen (vgl. OECD, 2016) –, werden sie bei der Zielgruppe von Lehrkräften bislang nur selten vorgenommen. Der Beitrag begegnet diesem Desiderat, indem ein Kompetenzniveaumodell im pädagogischen Wissen konstruiert wird. Hierzu werden Aufgabenmerkmale, die sich in einer Studie zur Niveaumodellierung (König, 2009) bewährt haben, aufgegriffen, theoretisch verankert und an einer Längsschnittstichprobe angehender Lehrkräfte analysiert. Das Modell wird zudem anhand Kriterien der Konstruktvalidität geprüft.

---

1 Die EMW-Studie wird von der Rhein-Energie-Stiftung Köln (Projektnummer: W-13-2-003) gefördert und von Prof. Dr. Johannes König (Universität zu Köln) und Prof. Dr. Martin Rothland (Universität Siegen) geleitet.

2. Theoretischer Rahmen

2.1 Konzeptualisierung des pädagogischen Wissens

Pädagogisches Wissen (general pedagogical knowledge, GPK) von Lehrer*innen kann zunächst als „knowledge, with special reference to those broad principles and strategies of classroom management and organization that appear to transcend subject matter" (Shulman, 1987, S. 8) verstanden werden. Es lässt sich vom fachdidaktischen (pedagogical content knowledge, PCK) und fachlichen Wissen (content knowledge, CK) abgrenzen (vgl. Baumert & Kunter, 2006; Shulman, 1987). Während sich CK und PCK auf das zu unterrichtende Fach beziehen, stehen beim GPK fächerübergreifende unterrichtliche Herausforderungen wie die Klassenführung im Mittelpunkt. Die pädagogischen und fachlichen Wissensfacetten lassen sich konzeptuell relativ klar voneinander abgrenzen. Das PCK hingegen, das als „special amalgam of content and pedagogy" (Shulman, 1987, S. 8) aufgefasst werden kann und sowohl Bereiche des pädagogischen als auch des fachwissenschaftlichen Wissens umfasst, erweist sich konzeptuell und empirisch bislang als weniger trennscharf (König et al., 2018).

Kernaufgabe von Lehrer*innen ist die „gezielte und nach wissenschaftlichen Erkenntnissen gestaltete Planung, Organisation und Reflexion von Lehr- und Lernprozessen sowie ihre individuelle Bewertung und systemische Evaluation" (KMK, 2004, S. 3). Um diese beruflichen Anforderungen erfolgreich zu bewältigen, ist GPK notwendig, das „Kenntnisse über das Lernen und Lehren, die sich auf die Gestaltung von Unterrichtssituationen beziehen und die fachunabhängig, das heißt auf verschiedene Fächer und Bildungsbereiche anzuwenden sind" umfasst (Voss et al., 2015, S. 194). GPK bezieht sich dementsprechend deutlich auf Anforderungen des Unterrichtens, insbesondere mit Blick auf Klassenführung, fächerübergreifende Aspekte von Lehr-Lern-Methoden, Umgang mit Heterogenität und Schüler*innenlernen sowie Leistungsbeurteilung.

Die Expertiseforschung stellt für die Ausübung des Lehrer*innenberufs eine umfangreiche Wissensbasis in den Fokus (vgl. Bromme, 1992). Im Anschluss daran und aktuellen kompetenztheoretischen Ansätzen folgend wird GPK übereinstimmend als zentrale Facette des professionellen Wissens (vgl. Shulman, 1987; Bromme, 1992; Terhart, 1991) sowie als eine zentrale kognitive Komponente der professionellen Kompetenz von Lehrer*innen angesehen (vgl. Baumert & Kunter, 2006; Kunter et al. 2011; König & Klemenz, 2015). Kompetenzen werden im Folgenden als „*kontextspezifische kognitive Leistungsdispositionen,* die sich funktional auf Situationen und Anforderungen in bestimmten *Domänen* beziehen", verstanden (Klieme & Leutner, 2006, S. 879). Gemäß dieser funktionalen Eingrenzung von Kompetenz auf kognitive Fähigkeiten wird GPK als wesentliche Voraussetzung betrachtet, um spezifische Anforderungen im Lehrer*innenberuf zu bewältigen, und kann somit der hier angestrebten Kompetenzniveaumodellierung zugrunde gelegt werden.

Bei der Messung von Kompetenzen kann zwischen Kompetenzstruktur- und Kompetenzniveaumodellen differenziert werden (vgl. Hartig & Klieme, 2006). Strukturmodelle unterscheiden, welche und wie viele Teilkompetenzen eine Kompetenz aufweist.

Niveaumodelle hingegen beschreiben, über welche spezifischen kognitiven Fähigkeiten Personen bei unterschiedlichen Kompetenzausprägungen verfügen (vgl. Hartig & Klieme, 2006; Schaper, 2009).

2.2 Kompetenzstrukturmodell

Das in dieser Studie vorliegende Kompetenzstrukturmodell pädagogischen Wissens rückt den Unterricht und die oben dargestellten Anforderungen als „Kerngeschäft" von Lehrer*innen in den Mittelpunkt. Unter Einbezug der KMK-Standards (vgl. KMK, 2004) sowie Ansätzen der Allgemeinen Didaktik und der Lehr-Lern-Forschung (vgl. detailliert König & Blömeke, 2009) wurden fünf inhaltliche Dimensionen des GPK definiert: Umgang mit Heterogenität, Strukturierung von Unterricht, Klassenführung, Motivierung und Leistungsbeurteilung. Die genannten Dimensionen fungierten als Basis für die Erstellung des TEDS-M Testinstruments, dessen Konzeption und Operationalisierung sich in mehreren empirischen Studien zur Lehrer*innenforschung bewährt haben und auch bei Kritiker*innen Anerkennung findet (vgl. Lüders, 2012; Voss et al., 2015). Die Testinhalte korrespondieren weitgehend mit den in den KMK-Standards (2004) als Kernaufgaben formulierten Anforderungen (vgl. König & Blömeke, 2010a). Das Testinstrument wurde mehrfach Validitätsprüfungen unterzogen, wodurch u. a. Konstruktvalidität und curriculare Validität belegt werden konnten (vgl. König & Klemenz, 2015).

3. Kompetenzniveaumodell

Bei der Modellierung von Kompetenzniveaus stehen weniger inhaltliche Facetten im Fokus der Betrachtung, sondern vielmehr die auf einem bestimmten Niveau verfügbaren kognitiven Fähigkeiten, die zur Bewältigung berufsspezifischer Anforderungen in einem bestimmten Inhaltsbereich von Bedeutung sind. Daran anschließend verwenden wir im Folgenden die Konzepte *kognitive Komplexität* und *Anwendung unterrichtsbezogener Terminologie* und stellen ihre Relevanz für Kernaufgaben von Lehrer*innen heraus. Sie dienen als Rationale für die Auswahl der in das Modell integrierten Aufgabenmerkmale.

Kognitive Komplexität
Mit dem Konzept der kognitiven Komplexität (vgl. Bieri et al., 1966; Scott, 1962; Peterson & Scott, 1983) werden unterschiedliche Qualitäten von informationsverarbeitenden Prozessen eines Individuums beschrieben. Dabei wird nicht ein spezieller Inhalt in den Mittelpunkt gerückt, sondern die für die Informationsverarbeitung vorhandenen Wissensstrukturen (vgl. Mandl & Huber, 1978). In den verschiedenen Ansätzen des Konzepts lässt sich im Wesentlichen eine konstituierende Komponente herausstellen: Kognitive Komplexität wird übereinstimmend „durch die *Zahl der Dimensionen* beschrieben, die eine Person bei der Strukturierung und Beurteilung eines Objektes oder Sachver-

halts verwendet“ (von Eye, 1999, S. 82). Je mehr Wissenselemente in eine Entscheidungsfindung einfließen, desto komplexer ist das Niveau der Informationsverarbeitung. Die verfügbaren Informationen über Anforderungssituationen, Handlungsmöglichkeiten und -präferenzen sowie Bewertungen und Werturteile werden in entsprechend viele differenzierte Dimensionen zerlegt, wodurch mehrere Perspektiven eingenommen werden können und eine flexible Wissensanwendung begünstigt wird (vgl. Huber, 1982). Bromme (1992) stellt fest, dass sich eine Anhäufung einzelner Wissenselemente nicht nur in ihrer Anzahl widerspiegelt, sondern damit auch eine größere Anzahl an Bedeutungseinheiten von Wissen („Chunks“) einhergeht. Diese können einzelne Begriffe umfassen, aber auch zu größeren Einheiten wie Begriffskategorien und Konzepten zusammengefasst werden. Eine breite Basis verdichteter Chunks wird als essenziell für die Prozeduralisierung von Wissen betrachtet (vgl. Anderson, 1982).

Bezugnehmend auf das Konstrukt des pädagogischen Wissens, wie es nachfolgend untersucht wird, kann angenommen werden, dass eine hohe kognitive Komplexität förderlich wirkt, bei unterrichtsbezogenen Kernanforderungen, mehrperspektivische Sichtweisen einnehmen und vielfältige Handlungsoptionen kreieren zu können (vgl. auch Testaufgaben-Beispiel in Kapitel 3.2). Diese Fähigkeit stellt insbesondere bei pädagogischen Reflexionsprozessen, mit denen Wissen und Handeln weiterentwickelt werden kann (vgl. Combe & Kolbe, 2008) und denen auch in der Expertiseforschung bei der Entwicklung des Novizen zum Fortgeschrittenen Anfänger und Experten ein hoher Stellenwert beigemessen wird (vgl. Berliner, 2004), einen zentralen Aspekt dar. Bei Reflexionsprozessen wird nicht das Ziel verfolgt, eine eindeutige Lösung zu finden, sondern vielschichtige Problemsituationen aus mehreren Sichtweisen zu betrachten (vgl. Schön, 1983). Somit kann verhindert werden, dass lediglich ausgewählte Kriterien bzw. subjektive Theorien einbezogen, unter Umständen wichtige und zentrale Aspekte vernachlässigt, und somit negative Handlungsentscheidungen gefällt werden (vgl. Wyss, 2013).

Ferner deutet ein hohes Niveau kognitiver Komplexität auf eine besser strukturierte und verdichtete Wissensbasis hin. Wenn GPK sehr gut vernetzt (vgl. Baumert & Kunter, 2006) und in größeren, bedeutungsreicheren Einheiten repräsentiert ist, fördert dies „die Wahrnehmung, Informationsaufnahme und -verarbeitung sowie das Problemlöseverhalten“ (Gruber & Stöger, 2011, S. 252). Unter Nutzung von Chunks können neue Gegebenheiten in unterrichtsbezogenen Situationen schneller erkannt, gespeichert, mit Gedächtnisinhalten verknüpft und rascher in Handlungsvorschlägen umgesetzt werden (vgl. Wipperfürth, 2015). Eine kategoriale Wahrnehmung, durch die lösungsdienliche Strukturen sowie funktionale Zusammenhänge in Situationen erkannt werden können, stellt eine Voraussetzung für situatives Lehrer*innenhandeln dar (vgl. Bromme, 1992) und wird durch komplexe kognitive Strukturen begünstigt. Das Merkmal gewinnt demnach auch für die Organisation von Lehr-Lernprozessen an Bedeutung.

Anwendung unterrichtsbezogener Terminologie

Es kann angenommen werden, dass der Zugang zum Gegenstandsbereich pädagogischen Wissens insbesondere über eine spezifische Terminologie möglich wird (vgl. König, 2009). Denn „wie alles Wissen, so ist auch pädagogisches Wissen an eine be-

stimmte Sprache gebunden“ (Terhart, 1991, S. 136). Alltagssprache über Erziehung unterscheidet sich von berufsspezifischer Sprache und diese wiederum von wissenschaftsinterner Sprache von Erziehungswissenschaftler*innen (vgl. auch Bastian & Helsper, 2000). In konzeptioneller Hinsicht wurden in Anlehnung an die Diskussion von Terhart (1991) zum GPK, seiner Funktion und sprachlichen Form im Bereich des Lehrer*innenberufs drei Sprachniveaus formuliert (vgl. König, 2009): (1) praxisnahe, unterrichtsbezogene Umgangssprache, (2) Fachsprache von Lehrer*innen und (3) eine wissenschaftsspezifische Sprache. Die Sprachniveaus lassen sich durch spezifische Termini charakterisieren (vgl. Kapitel 3.2).

Ein fundierter Umgang mit unterrichtsbezogener Terminologie auf einem hohen sprachlichen Niveau deutet auf ein hohes Niveau spezialisierten, pädagogischen Wissens hin. Angehende Lehrpersonen besitzen in diesem Fall tiefreichende Kenntnisse über zentrale theoretische Begriffe und Konzepte in der Domäne der Pädagogik und können diese in bestimmten Situationen abrufen, was eine Anwendung wahrscheinlicher macht (vgl. Bromme, 1992). Die Fähigkeit, Erfahrungen, Wahrgenommenes und konkrete Unterrichtserlebnisse mit pädagogischem Wissen in Beziehung zu setzen, ist ein zentraler Bestandteil der Aufgaben von Lehrer*innen und ein wichtiger Impuls für die professionelle Unterrichtsentwicklung (vgl. Wipperfürth, 2015).

Darüber hinaus ist es für die Anwendung spezifischer Terminologie erforderlich, GPK zu explizieren. Helsper (2002) betont, dass es für professionelle Lehrer*innen entscheidend ist, nicht nur über implizites Wissen zu verfügen, sondern dieses auch abrufen, in explizites Wissen übersetzen und zum impliziten schulisch-unterrichtlichen Wissen ins Verhältnis setzen zu können. Damit kann es in Form (selbst)reflexiven Wissens den unterrichtlichen Routinen im Sinne einer kritisch-reflexiven Evaluation gegenübergestellt und für reflexive Transformationen geöffnet werden. Schön (1983) unterscheidet Reflexionsprozesse in der Handlung (reflection-in-action) von solchen über die Handlung (reflection-on-action). Insbesondere die Reflexion über die Handlung erfordert expliziertes Wissen für eine Handlungsstrategie, um es analysierbar und re-organisierbar sowie mittelbar zu machen (vgl. Altrichter & Posch, 2007). Demnach spielt eine spezifische Terminologie bei der eigenen Reflexion eine wesentliche Rolle. Sie wird jedoch ebenso im Austausch mit Professionskolleg*innen bedeutsam. Das GPK, das einer Handlung zu Grunde liegt, wird sprachlich geordnet ausgedrückt, wodurch es erst möglich wird, Entscheidungen und Handlungen mit Professionskolleg*innen diskutieren und kritisch prüfen sowie ihnen gegenüber begründen zu können (vgl. Altrichter & Posch, 2007).

Terhart (2002, S. 18) betont zudem die „Kommunikationsfähigkeit über Inhalte, Strukturen und Probleme des unterrichtsfachlichen, pädagogisch-didaktischen und schulbezogenen Bereichs“ als bedeutsame Fähigkeit angehender Lehrer*innen. Es kann angenommen werden, dass Personen, die über unterrichtsbezogene Terminologie auf einem hohen Niveau verfügen, diese auch in ihrer Kommunikation anwenden.

Zusammenfassend lässt sich annehmen, dass Lehrer*innen, die im Rahmen ihres GPK über komplexe kognitive Strukturen verfügen und ihr umfassendes Wissen mittels differenzierter, unterrichtsbezogener Terminologie explizieren können, eher dazu in der

Lage sind, zentrale berufsspezifische Anforderungen erfolgreich zu bewältigen. Hierbei wurden Bezüge zu Anforderungen der KMK-Standards – Reflexion, Evaluation und Organisation von Lehr-Lernprozessen (vgl. KMK, 2004) – und weiteren zentralen Aufgaben im Lehrer*innenberuf hergestellt. Die Merkmale kognitive Komplexität und Anwendung unterrichtsbezogener Terminologie sind somit von substanzieller Bedeutung für Lehrer*innen und können als Grundlagen für ein Kompetenzniveaumodell im pädagogischen Wissen herangezogen werden.

3.1 Methodisches Vorgehen

Um psychometrische Kompetenzniveaumodelle zu entwickeln, kommen verschiedene Verfahren in Betracht. Insbesondere der modellgeleitete Ansatz (vgl. Kapitel 5.4) zur Vorhersage von Aufgabenschwierigkeiten durch Aufgabenmerkmale (vgl. bspw. Gorin & Embretson, 2006; Hartig, 2007), welcher auch in der DESI-Studie (vgl. Hartig, 2007) eingesetzt wurde, kann zu den häufig angewendeten und etablierten Verfahren gezählt werden.

Die vorliegende Studie verfolgt diesen Ansatz, da er mehrere Vorteile bietet (vgl. Hartig & Frey, 2012): Anhand einer kriteriumsorientierten Skaleninterpretation kann unabhängig von der Verteilung der gemessenen Kompetenz eine qualitative Beschreibung von Testwerten vorgenommen werden. Bei dem ebenfalls häufig angewendeten *proficiency scaling* Ansatz nach Beaton und Allen (1992), der bspw. in den PISA- und TIMSS-Studien eingesetzt wurde, wird hingegen das kontinuierliche Kompetenzmerkmal oftmals anhand willkürlicher Kriterien (bspw. gleiche Abstände auf der Testwertskala) mittels Schwellenwerten unterteilt und die Schwellen post-hoc auf Basis geeigneter Testaufgaben beschrieben (vgl. Harsch & Hartig, 2011). Aufgrund dessen und einem geringen inhaltlichen Bezug auf das zu messende Konstrukt steht der Ansatz hinsichtlich seiner Validität in Kritik (vgl. Harsch & Hartig, 2011). Die systematische Vorhersage von Schwierigkeitsunterschieden zwischen Testitems stellt zudem einen wesentlichen Schritt zur Konstruktvalidierung im Sinne der Konstruktrepräsentation dar (vgl. Embretson, 1983) und zählt zu einer empirischen Strategie, die sich diesbezüglich etabliert hat (vgl. Jenßen, Dunekacke & Blömeke, 2015). Die differenzierte Beschreibung schwierigkeitsbestimmender Aufgabenmerkmale und ein daraus abgeleitetes Messmodell liefern belastbare Hinweise über die Validität von Testwertinterpretationen (vgl. Hartig, 2007). Ferner kann die Dekomposition von Itemschwierigkeiten zur Optimierung von Testitems herangezogen werden (vgl. Hartig & Frey, 2010).

3.2 Dekomposition von schwierigkeitsbestimmenden Aufgabenmerkmalen

Auf Basis der in Kapitel 3 dargestellten theoretischen Ansätze wurden Aufgabenmerkmale abgeleitet, deren Abstufungen die interindividuellen Unterschiede in den kognitiven Strukturen und Prozessen der Personen abbilden (vgl. Hartig & Jude, 2007).

Aufgabenmerkmal Komplexität kognitiver Bearbeitungsprozesse

Kognitive Komplexität wurde operationalisiert, indem die für das Lösen notwendige Antwortstruktur mit Charakteristika der erforderlichen Wissensstrukturen in Beziehung gesetzt wurde (vgl. Prenzel, M., Häußler, P., Rost, J., & Senkbeil, M., 2002). Aufgaben mit einfachen kognitiven Bearbeitungsprozessen werden so definiert, dass für die Lösung lediglich ein einzelnes Wissenselement erforderlich ist. Bei einer Aufgabe mit komplexen kognitiven Bearbeitungsprozessen ist hingegen eine Antwortstruktur erforderlich, bei der mindestens zwei und bis zu fünf in Beziehung zueinander stehende Wissenselemente abgerufen und angewendet werden müssen. Die Einstufung einer Aufgabe hinsichtlich der Komplexität kognitiver Bearbeitungsprozesse erfolgt demnach anhand der für die Lösung erforderlichen Anzahl an Elementen. Ob ein oder mehrere Wissenselemente den Erwartungshorizont der jeweiligen Testaufgaben widerspiegeln, ist in dem Kategoriensystem der umfassenden Kodieranleitung festgelegt (vgl. König & Blömeke, 2010b).

Die folgende Testaufgabe verdeutlicht die Operationalisierung des Aufgabenmerkmals Komplexität kognitiver Bearbeitungsprozesse (siehe Abb. 1). Sie wurde als komplex eingestuft, da zur Lösung mehrere verknüpfte Wissenselemente erforderlich sind. Somit findet die Vernetzung der zu bearbeitenden Inhalte Berücksichtigung (vgl. Kauertz, Fischer, Mayer, Sumfleth & Walpuski, 2010). Formuliert eine Person mehrere Fragen, die hinsichtlich verschiedener Kategorien als korrekt gewertet werden, so wurden mehr Wissenselemente zur Strukturierung verwendet, als wenn bspw. sämtliche Fragen aufgrund von Redundanzen nur einer Kategorie zugeordnet werden können.

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)
…
10)

Abb. 1: Testaufgabe zur Erfassung von Wissen zur Strukturierung von Unterricht (König & Blömeke, 2009).

Aufgabenmerkmal terminologische Anforderungen

Das Konzept der Anwendung unterrichtsbezogener Terminologie wurde operationalisiert, indem die im Aufgabenstamm verwendete Terminologie nach oben dargestellter Kategorisierung eingestuft wurde (vgl. Kapitel 3). Zu praxisnaher, unterrichtsbezogener Umgangssprache werden Begriffe wie „Gruppenarbeit", „Stillarbeit" oder „Unterrichtstempo" gezählt, denen trotz ihres Fachsprachencharakters keine schwierigkeitserhöhenden Effekte zugeschrieben werden, da sie Lehramtsstudierenden aufgrund ihrer schulischen Sozialisation als Schüler*innen geläufig sind. Die Kategorie Fachsprache umfasst gängige Begriffe, von denen angenommen wird, dass sie Lehrer*innen im Diskurs über Unterricht verwenden. Dieses Niveau basiert auf einer abstrakteren Terminologie, die deutlichen Professionsbezug aufweist und angehenden Lehrpersonen nicht zwingend geläufig ist. Es ist durch Begriffe wie „innere Differenzierung" oder „intrinsische Motivation" gekennzeichnet. Von den beiden genannten Niveaus hebt sich das dritte insofern ab, als dass Termini hierzu gezählt werden, die überwiegend im Wissenschaftsdiskurs verwendet werden und ggf. auch dort entstanden sind. Im Mittelpunkt stehen abstrakte Aspekte von Unterricht und Termini, die nur selten zum Vokabular von Lehrpersonen gehören dürften. Die Betrachtung von Unterricht erfolgt hierbei aus einer hochspezialisierten Perspektive. Das Niveau wird trotz des hohen Abstraktionsgrades und Wissenschaftsbezugs für Lehrer*innen als bedeutsam erachtet, da es spezifische Theorien zugänglich macht, die insbesondere bei der professionellen Reflexion berufsbezogener Probleme erforderlich sind. Das wissenschaftliche Sprachniveau wird durch Begrifflichkeiten wie „Akkommodation" oder „operante Konditionierung" repräsentiert.

Mit einem Anstieg im Niveau des jeweiligen Aufgabenmerkmals – Komplexität kognitiver Bearbeitungsprozesse bzw. terminologische Anforderungen – sollte eine höhere Aufgabenschwierigkeit einhergehen. Bei der erfolgreichen Bewältigung von Aufgaben mit hoher Aufgabenschwierigkeit wird ein hohes Niveau bezüglich der jeweiligen kognitiven Fähigkeit unterstellt.

4. Fragestellungen

Hinsichtlich der Fragestellungen wird auf einen Ansatz zur Niveaumodellierung von König (2009) Bezug genommen, der auf Basis des TEDS-M Testinstruments ebenfalls die Merkmale Komplexität kognitiver Bearbeitungsprozesse und terminologische Anforderungen einbezogen hat. Sie stellten sich bei einer Analyse auf Basis einer Querschnittstichprobe angehender Lehrer*innen als vorhersagestark heraus. Es wird untersucht, ob sich diese Befunde replizieren lassen. Die methodische Weiterführung, differenzierte schwierigkeitsbestimmende Kriterien anstelle globaler Aufgabeneinstufungen zur Beschreibung von Kompetenzniveaus anzuwenden, ermöglicht detaillierte Aussagen darüber, wie groß die Effekte der einzelnen Merkmale auf die Schwierigkeit von Aufgaben ausfallen. Darüber hinaus bringt das methodische Vorgehen die Vorteile mit sich, *cut-off* Scores präzise bestimmen und begründen sowie statistische Aussagen

über die Modellgüte treffen zu können. Ferner wird nun ein Fokus auf die Validierung des Modells gelegt. Die Analyse gibt Hinweise auf die Konstruktrepräsentation- bzw. validität der Testwerte (vgl. Winther, 2010). Die umfassende Längsschnittstichprobe ermöglicht insbesondere mit Blick auf die Änderungssensitivität eine größere Aussagekraft. Zudem wird ein Kriterium der diskriminanten Validität in die Analyse miteinbezogen. Vor diesem Hintergrund wird folgenden Fragen nachgegangen:

1) Können anhand der Aufgabenmerkmale Komplexität kognitiver Bearbeitungsprozesse und terminologische Anforderungen die empirischen Aufgabenschwierigkeiten erwartungskonform vorhergesagt werden?

Es wird angenommen, dass komplexere kognitive Bearbeitungsprozesse und ein höheres sprachliches Niveau höhere empirische Itemschwierigkeiten bedingen. Die erfolgreiche Vorhersage von Aufgabenschwierigkeiten durch Aufgabenmerkmale kann als Indiz für Konstruktvalidität interpretiert werden (vgl. Kapitel 3.1).

2) Erreichen angehende Lehrer*innen im fortgeschrittenen Ausbildungsstadium häufiger höhere Kompetenzniveaus als zu Beginn der Lehrer*innenausbildung?

Da wir der Lehrer*innenausbildung Wirksamkeit unterstellen, die sich in Lernzuwächsen bei den betrachteten Lehramtsstudierenden über den untersuchten Zeitraum zeigen sollte, nehmen wir an, dass fortgeschrittene Studierende höhere Kompetenzniveaus erreichen. Eine Bestätigung unserer Annahme deutet auf eine hohe Änderungssensitivität des Modells hin.

3) Zeigen sich geringe Effekte der Schulabschlussnoten angehender Lehrer*innen auf die Erreichung höherer Kompetenzniveaus?

Dem vorliegenden Beitrag liegt ein enges, auf kognitive Leistungsdispositionen bezogenes Verständnis von Kompetenz zugrunde (vgl. Kapitel 2.1), wodurch sich eine gewisse inhaltliche Überschneidung mit dem Intelligenzkonzept ergibt (vgl. Hartig & Klieme, 2006). Aufgrund dessen empfiehlt sich eine Analyse, ob der für das pädagogische Wissen konstruktirrelevante Faktor Intelligenz in der Messung eine Rolle spielt und somit das Konstrukt zumindest teilweise konfundiert sein könnte (vgl. Messick, 1995). Da die Konstruktion der Kompetenzniveaus nicht auf allgemeinen kognitiven Grundfertigkeiten basiert, sondern an kognitive Fähigkeiten im pädagogischen Wissen gekoppelt ist, die vorwiegend in Lerngelegenheiten im Lehramtsstudium erworben wurden, vermuten wir, dass die Schulabschlussnoten – als Indikator für allgemeine kognitive Grundfertigkeiten – der Studierenden nur einen geringen Einfluss auf die Erreichung höherer Kompetenzniveaus ausüben. Geringe Effekte können als Hinweis auf diskriminante Validität interpretiert werden und würden bisherige Befunde bestätigen (vgl. König, Tachtsoglou & Seifert, 2012).

5. Untersuchungsmethode

5.1 Stichprobe

Zur Bearbeitung der dargelegten Fragestellungen wird ein Teildatensatz der EMW-Studie verwendet, der Lehramtsstudierende aus Deutschland und Österreich umfasst, die zu den beiden Messzeitpunkten (MZP) im Herbst 2011 (1. Semester) und im Herbst 2013 (5. Semester) teilgenommen haben. Zum ersten MZP haben die Teilnehmer*innen ihr Bachelorstudium erst kürzlich begonnen, während sie sich zum zweiten MZP im letzten Jahr ihres Studiums befanden. Der Teildatensatz umfasst 1451 angehende Lehrer*innen aus 18 Hochschulstandorten (vgl. für eine detaillierte Stichprobenbeschreibung König & Klemenz, 2015). Sie verteilen sich auf vier nach den KMK-Standards definierte Lehramtstypen und wurden ausgewählt, um eine ausreichende Vergleichbarkeit zu gewährleisten. In dem Teildatensatz sind 33 Studiengänge an 18 Hochschulen vertreten. Im Durchschnitt ist ein Studiengang eines Standortes mit 44 Studierenden besetzt. Der Teildatensatz deckt somit ein breites Feld an verschiedenen Studiengängen ab.

Tab. 1 enthält die Fallzahlen der Untersuchungsstichprobe getrennt nach Land und Ausbildungsgang.

KMK-Typ	Bezeichnung	Deutschland		Österreich		Gesamt	
		Studierende	Hochschulen	Studierende	Hochschulen	Studierende	Hochschulen
1	Lehrämter der Grundschule bzw. Primarstufe (Lehramt für Volksschulen)	208	5	386	6	594	11
3	Lehrämter für alle oder einzelne Schularten des Sekundarbereichs I (Lehramt für Hauptschulen)	239	5	194	5	433	10
4	Lehrämter für den Sekundarbereich II (allgemeinbildende Fächer) oder für das Gymnasium (Lehramt für höhere Schulen)	136	5	46	1	182	6
6	Sonderpädagogische Lehrämter (Sonderpädagogisches Lehramt)	207	4	35	2	242	6
	Gesamt	790	10	661	8	1451	18

Tab. 1: Verteilung des hier verwendeten Teildatensatzes der EMW-Längsschnittstichproben von Studierenden und Hochschulen auf die Lehramtstypen der KMK (2009) sowie österreichische Bezeichnung in Klammern (König & Klemenz, 2015)

5.2 Testinstrument und Skalierung

Für die Testung des pädagogischen Wissens angehender Lehrkräfte wurde das TEDS-M Testinstrument angewendet (vgl. König & Blömeke, 2009). Es umfasst Aufgaben mit offenem und geschlossenem Antwortformat. Die Kodierung der offenen Antworten wurde von geschulten Rater*innen mithilfe umfassender Kodieranleitungen durchgeführt. Die Ergebnisse der Übereinstimmungsanalyse bei Doppelkodierungen weisen eine sehr gute Interrater-Reliabilität auf (durchschnittliches Cohens Kappa von 0,80).

Eine wichtige Voraussetzung für die kriteriumsbezogene Skalenbeschreibungen ist die Skalierung des Wissenstests auf Basis der Item-Response-Theorie (IRT). Die IRT-Skalierung ermöglicht es, Personenfähigkeitswerte und Itemschwierigkeiten auf einer gemeinsamen Skala abzubilden. Mithilfe der Software ConQuest (vgl. Wu, Adams & Wilson, 1997) wurden die Testdaten – einem Vorgehen der LEK Studie folgend (vgl. König, 2012) – zunächst getrennt für beide Messzeitpunkte skaliert. Die hohe Korrelation der Itemschwierigkeiten beider Messzeitpunkte ($r = 0{,}79$) verweist auf eine ausreichende Stichprobenunabhängigkeit, wodurch eine konkurrente Skalierung als zulässig angesehen werden kann (vgl. Bond & Fox, 2007). Anschließend wurden die Testdaten einer eindimensionalen Rasch-Skalierung im Ansatz virtueller Fälle (konkurrente Skalierung) unterzogen (vgl. Rost, 2004b). Bei diesem Vorgehen stellen Längsschnittprobanden, d. h. Personen, für die eine zweimalige Messung vorliegt, zwei Fälle in der Skalierungsdatei dar. Durch die simultane Skalierung aller Fälle wird dem Problem eingeschränkter Varianz, wie es bei einer Skalierung je Messzeitpunkt häufig vorliegt, begegnet. Mit der Parameterschätzung in ConQuest wird jedem Item auf Basis seiner empirischen Lösungshäufigkeit ein Schwierigkeitsparameter und jeder Testperson entsprechend dem Antwortverhalten ein Fähigkeitsparameter zugewiesen (Wu et al., 1997).

Vor dem Hintergrund vorheriger Skalierungsergebnisse (vgl. König, 2012; König & Klemenz, 2015) wurden 79 einzelne Items in die hier durchgeführte Skalierung einbezogen, die teilweise zusammengeführt wurden, um das Aufgabenmerkmal der kognitiven Komplexität abzubilden. Zur Verdeutlichung wird noch einmal auf die in Abbildung 1 dargestellte Testaufgabe Bezug genommen.

Die Aufgabe wurde ursprünglich in vier inhaltliche Fokusse untergliedert, auf denen jeweils eine richtige Wertung möglich war (vgl. König, 2012). Für die Umsetzung des Merkmals der kognitiven Komplexität wurde die Aufgabe erst dann als korrekt gewertet, wenn Lösungen auf mehreren Fokussen erzielt wurden. Methodisch wurde dies durch eine Zusammenfassung betreffender Items zu einem Item umgesetzt, das erst bei mehreren richtigen Lösungen als korrekt gewertet wurde. Aufgrund dessen fließt eine reduzierte Anzahl von 48 Items in das vorliegende Modell ein, wodurch das Konstrukt GPK etwas weniger differenziert abgebildet werden kann. Die Kurzform des TEDS-M Testinstruments (vgl. König & Blömeke, 2010b) verweist jedoch auf sehr gute Ergebnisse auch mit einer geringeren Anzahl an Items. Die Korrelationsanalysen (s. u.) des neu strukturierten Tests mit dem Test in Langform unterstreichen dies. Die Skalierung weist eine gute Reliabilität auf (EAP/PV-Reliabilität: 0,80, WLE-Reliabilität: 0,77, beide vergleichbar mit Cronbach's Alpha). Die Itemschwierigkeitswerte streuen

über eine Range von rund fünf Logits (−2,78 bis 2,20). Die Diskriminationswerte der Items liegen durchschnittlich in einem Bereich von 0,35 (Min.: 0,10, Max.: 0,65;), der Weighted Mean Square der Items fällt in einen akzeptablen Bereich von 0,80 bis 1,20. Lediglich ein Item weist einen Wert von 1,21 auf. Aufgrund inhaltlicher Gesichtspunkte wurde dieses Item beibehalten.

In einem weiteren Schritt wurde die Vergleichbarkeit der in diesem Artikel durchgeführten Skalierung mit der Skalierung über die Gesamtanzahl der Items (vgl. König & Klemenz, 2015) geprüft. Hierzu wurden sowohl die Personenfähigkeitswerte der Teilnehmer*innen als auch die Itemschwierigkeiten der 34 unveränderten Items mit Ergebnissen aus der vorherigen Skalierung (vgl. König & Klemenz, 2015) korreliert. Die Korrelation nach Pearson beträgt unter Einbezug der Personenfähigkeiten r = 0,90, unter Einbezug der Itemschwierigkeiten r = 0,99. Die sehr hohen Korrelationen können als Indiz dafür angesehen werden, dass die Skalierungsergebnisse miteinander vergleichbar sind (vgl. Bond & Fox, 2007).

5.3 Zuordnung der Aufgabenmerkmale zu den Testitems

Im Kategoriensystem der Kodieranleitung ist der Erwartungshorizont dazu, wie viele Wissenselemente zur Lösung einer Aufgabe erforderlich sind, für jede Testaufgabe festgelegt (vgl. König & Blömeke, 2010b; vgl. Kapitel 3.2). Aufgaben, für deren Lösung leidglich ein Wissenselement notwendig ist, wurden dem niedrigen Niveau und Aufgaben, für deren Lösung mehrere Wissenselemente notwendig sind, dem hohen Niveau kognitiver Komplexität zugewiesen. Anhand dieses Vorgehens wurden 14 Aufgaben hinsichtlich der Komplexität kognitiver Bearbeitungsprozesse aufgrund ihres erforderlichen Wissensumfangs als hoch eingestuft. Die Merkmalskombination komplexe Bearbeitungsprozesse und wissenschaftliches Sprachniveau ist im Itempool, der in die Analyse einfließt, nicht realisiert. Das Modell kann demnach ein Niveau mit der anspruchsvollsten Merkmalskombination nicht abbilden.

Hinsichtlich der terminologischen Anforderungen wurden die Testaufgaben dahingehend analysiert, welche unterrichtsbezogenen Termini sie enthalten. So wurden 86 Termini identifiziert, die anschließend von acht in der Lehrer*innenausbildung tätigen Lehrkräften, Erziehungswissenschaftler*innen und Psycholog*innen den jeweiligen Niveaus zugeordnet wurden. Die acht unabhängigen Ratings weisen sehr gute Übereinstimmungen auf ((ICC$_{\text{unjustiert}}$(2, 8) = 0,95). Auf Basis der Mittelwerte aller acht Ratings wurden die Items einer Kategorie zugeordnet. 18 Items wurden einem umgangssprachlichen, 24 Items einem fachsprachlichen und 6 Items einem wissenschaftssprachlichen Niveau zugewiesen. Die Verteilung der Items hinsichtlich ihrer schwierigkeitsbestimmenden Merkmalskombinationen ist in Tabelle 2 dargestellt.

Komplexität	Sprachniveau	Anzahl Items
Einfach	Umgangssprachlich	9
Einfach	Fachsprachlich	19
Einfach	Wissenschaftlich	6
Komplex	Umgangssprachlich	9
Komplex	Fachsprachlich	5
		48

Tab. 2: Verteilung der einbezogenen Testitems nach schwierigkeitsbestimmenden Aufgabenmerkmalen

5.4 Bildung von Kompetenzniveaus auf Basis einer multiplen Regressionsanalyse

Die schwierigkeitsbestimmenden Merkmale können eingesetzt werden, um die empirischen Itemparameter vorherzusagen. Hierzu fließen die Aufgabenmerkmale als Prädiktoren in Form von Dummy-Variablen in eine multiple Regressionsanalyse ein. Auf Basis der Regression können die Schwellen zwischen den Kompetenzniveaus definiert werden. Für die Auswahl der Prädiktoren zur Schwellenwertdefinition werden einerseits die absolute Größe der Regressionsgewichte und andererseits positive Vorzeichen, die eine schwierigkeitserhöhende Wirkung implizieren, als wesentlich erachtet. Die Signifikanz wird nicht als das wichtigste Kriterium betrachtet, da zunächst keine Prüfung eines allgemeingültigen Modells im Vordergrund steht, das für eine Population von Items Gültigkeit besitzen soll (vgl. Hartig, 2007).

Nach der Auswahl geeigneter Prädiktoren werden sie für bestimmte Anforderungskonfigurationen der kriteriumsbezogenen Skalenbeschreibung verwendet. Hartig (2007) verweist dabei auf vier inhaltliche bzw. empirische Gütekriterien, die bei der Auswahl der Merkmalskombination von Bedeutung sind. Es ist darauf zu achten, dass ausreichend Aufgaben existieren, bei denen die Merkmalskombinationen auch tatsächlich realisiert und die Merkmalskombinationen inhaltlich gut geeignet sind, um Unterschiede zwischen Proband*innen niedrigerer und höherer Kompetenz zu unterscheiden. Ferner sollten die empirischen Schwierigkeiten nicht zu stark von den erwarteten Schwierigkeiten abweichen und die Schwellen auf der Kompetenzskala nicht zu dicht beieinander liegen.

Die Schwellenwerte der Kompetenzniveaus ergeben sich aus der Summe der Regressionsgewichte der jeweils kombinierten Merkmale. Die Regressionskonstante entspricht einer maximal einfachen Aufgabe (vgl. Hartig, 2007). Der Wert, bei dem ein Merkmal bzw. eine Merkmalskombination erstmalig auftreten, kann als untere Schwelle des Kompetenzniveaus charakterisiert werden (vgl. Winther, 2010). Durch Addition der Regressionskonstante und einem Merkmal bzw. Merkmalskombinationen ergeben sich die jeweiligen *cut-off* Scores (vgl. Hartig, 2007).

6. Ergebnisse

Das Ergebnis der Regressionsanalyse (siehe Tab. 3) zeigt, dass sich alle drei Prädiktoren wie erwartet schwierigkeitssteigernd und als bedeutsam für die Vorhersage der Itemschwierigkeiten erweisen. Dies bestätigt die Annahmen im Rahmen der ersten Fragestellung. Das Merkmal der kognitiven Bearbeitungsprozesse ist äußerst vorhersagestark ($\beta = 0{,}74$). Hinsichtlich der sprachlichen Merkmale wird deutlich, dass die Itemschwierigkeiten von einem fachsprachlichen ($\beta = 0{,}24$) wie auch einem wissenschaftssprachlichen Niveau ($\beta = 0{,}21$) ähnlich stark beeinflusst werden. Das Kriterium der kognitiven Bearbeitungsprozesse wird signifikant ($p < 0{,}01$), die beiden terminologischen Kriterien liegen oberhalb des 5 %-Signifikanzniveaus (fachsprachliches Niveau $p = 0{,}06$; wissenschaftssprachlichen Niveau $p = 0{,}10$). Das Modell kann mit rund 43 % einen bedeutenden Teil der Varianz aufklären (adj. $R^2 = 0{,}43$, $F(3{,}44) = 12{,}67$, $p < 0{,}01$).

Abbildung 2 verdeutlicht die geringe Abweichung von erwarteter und empirischer Aufgabenschwierigkeit, was auf eine gute Passung zwischen theoretisch angenommener und empirischer Itemschwierigkeiten hindeutet. Die Mittelwerte der empirischen Itemschwierigkeiten je Kompetenzniveau unterscheiden sich signifikant ($F(2, 45) = 21{,}738$; $p < .001$), auch wenn Überlappungen erkennbar sind.

Unter Berücksichtigung der oben genannten Gütekriterien zur Auswahl der Merkmalskombinationen und Einbezug der Ergebnisse aus der multiplen Regressionsanalyse lassen sich vier Kompetenzniveaus sinnvoll konstruieren (siehe Tab. 4).

Adj. $R^2 = 0{,}43$	Regressionsgewicht	Standardfehler	Beta	p
Konstante	−0,942	0,267		0,001
Sprachniveau fachsprachlich	0,568	0,299	0,238	0,064
Sprachniveau wissenschaftlich	0,758	0,459	0,210	0,105
Komplexität	1,930	0,314	0,738	0,000

Tab. 3: Regressionsgewichte als Prädiktoren der Itemschwierigkeit

Kompetenzniveau	Schwellenwert (Logit)	Kurzbeschreibung
Unterhalb I	< −0,942	Einfache kognitive Bearbeitungsprozesse auf umgangssprachlichem Niveau können nicht mit hinreichender Wahrscheinlichkeit gelöst werden
I	−0,942	Einfache kognitive Bearbeitungsprozesse auf umgangssprachlichen Niveau
II	−0,374	Einfache kognitive Bearbeitungsprozesse auf fachsprachlichen bzw. wissenschaftssprachlichen Niveau
III	0,989	Komplexe kognitive Bearbeitungsprozesse auf umgangssprachlichen bzw. fachsprachlichen Niveau

Tab. 4: Schwellenwert und Kurzbeschreibung der Kompetenzniveaus

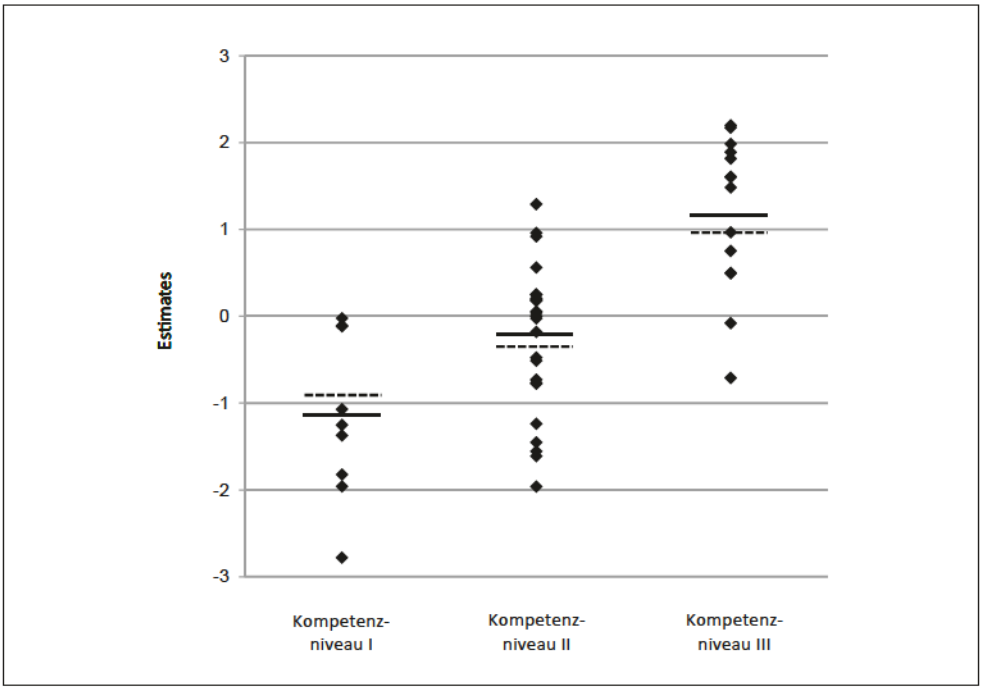

Abb. 2: Empirische Itemschwierigkeiten sowie Mittelwerte der erwarteten (gestrichelte Linie) und empirischen Itemschwierigkeiten (durchgezogene Linie) nach Kompetenzniveau

Beispielsweise errechnet sich der *cut-off* Score des Kompetenzniveaus II bei −0,374 Logits durch Addition der Regressionskonstante ($b_0 = -0{,}942$) mit dem unstandardisierten Regressionsgewicht des Merkmals Sprachniveau fachsprachlich ($b = 0{,}568$) (vgl. Hartig, 2007). Das Kompetenzniveau II umfasst in Kombination mit einfachen kognitiven Bearbeitungsprozessen die beiden Merkmale fachsprachlich und wissenschaftssprachlich, da eine Trennung in zwei Niveaus aufgrund der sehr nah bei einander liegenden *cut-off* Scores nicht sinnvoll erscheint (*cut-off* Score fachsprachlich: −0,374; *cut-off* Score wissenschaftssprachlich: −0,184).

Abbildung 3 visualisiert die Verteilung der angehenden Lehrer*innen auf die Kompetenzniveaus nach MZP. Zwischen den beiden MZP sind deutlich unterschiedliche Verteilungen der Studierenden auf die Niveaus erkennbar: Es zeigt sich wie bezüglich der zweiten Fragestellung angenommen, dass Studierende, die in ihrem Studium weiter fortgeschritten sind, signifikant häufiger höhere Kompetenzniveaus erreichen als Studierende zu Beginn ihres Studiums ($\chi^2 = 763{,}33$; df = 3; $p < 0{,}01$). Zum ersten MZP verbleiben nahezu drei Viertel der Proband*innen unterhalb des ersten Kompetenzniveaus. Sie können Aufgaben, die einfache kognitive Bearbeitungsprozesse auf umgangssprachlichem Niveau erfordern, nicht mit hinreichender Wahrscheinlichkeit lösen. Gut 16 % erreichen das erste und rund 10 % das zweite Kompetenzniveau. Zu Beginn

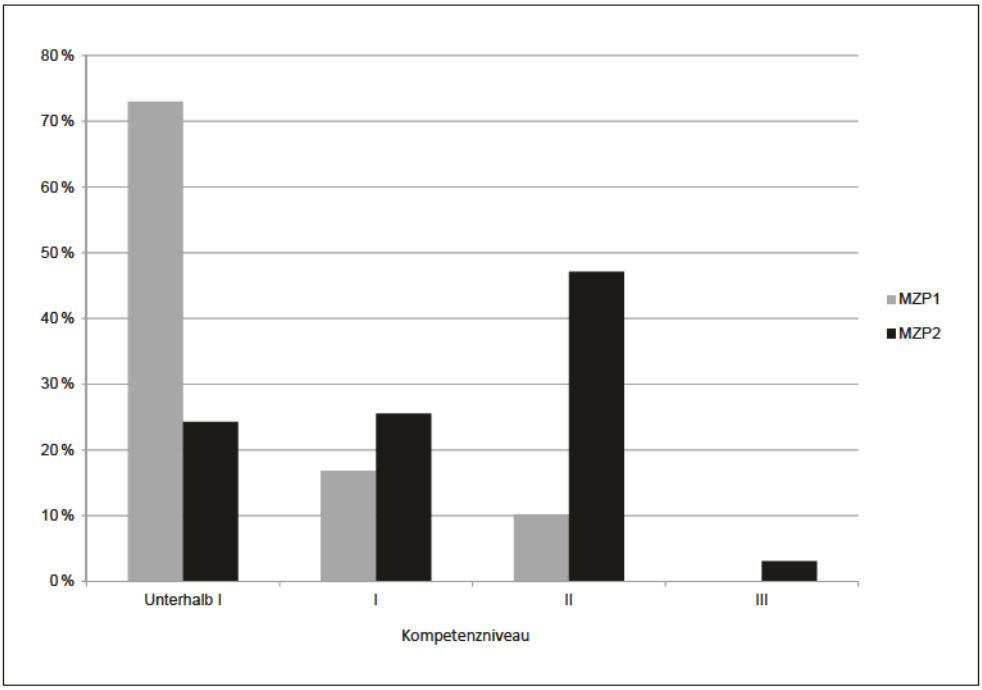

Abb. 3: Prozentuale Häufigkeitsverteilung der Testpersonen je Kompetenzniveau nach Messzeitpunkt

des Studiums erreicht keine Testperson das höchste Niveau. Im letzten Jahr ihres Bachelorstudiums erreicht hingegen rund die Hälfte der Studierenden das zweite Niveau. Jeweils etwa ein Viertel verbleibt unterhalb des ersten bzw. auf dem ersten Niveau. Das dritte Niveau erreichen rund 3 % der Studierenden im 5. Semester.

Mit Blick auf die dritte Fragestellung wurde geprüft, ob die Schulabschlussnoten der Studierenden geringe Effekte auf die Erreichung von höheren Kompetenzniveaus aufweisen. Hierzu wurden einfaktorielle Varianzanalysen mit dem Kompetenzniveau als Faktor getrennt nach Land durchgeführt und Effektstärken berechnet (Deutschland (DE): Abiturnote; Österreich (AT): Durchschnittliche Abschlussnoten des Maturitätszeugnisses). Über beide MZP hinweg sind die Mittelwertunterschiede (siehe Tab. 5) der Abschlusszeugnisnoten zwar signifikant, es zeigen sich jedoch nur äußerst geringe Effektstärken (DE: $F(3,1491) = 5{,}247$, $p < 0{,}01$, $\eta^2 = 0{,}01$; AT: $F(3,1218) = 7{,}899$, $p < 0{,}01$, $\eta^2 = 0{,}02$). Lediglich 1 % (DE) bzw. 2 % (AT) der Varianz in den Schulabschlussnoten lässt sich auf die Zugehörigkeit zu den Kompetenzniveaus zurückführen was als ein Indiz für diskriminante Validität gedeutet werden kann.

	Deutschland		Österreich	
Kompetenzniveau	MW	SD	MW	SD
Unterhalb I	2,388	0,592	2,479	0,741
I	2,414	0,560	2,360	0,780
II	2,302	0,561	2,336	0,775
III	2,056	0,635	1,667	0,488

Tab. 5: Mittelwerte (MW) und Standardabweichungen (SD) der Schulabschlussnoten nach Land und Kompetenzniveau

7. Zusammenfassung und Diskussion

Das Anliegen des Beitrags war die Konstruktion und Prüfung eines Kompetenzniveaumodells im pädagogischen Wissen angehender Lehrkräfte. Bezüglich der ersten Fragestellung bestätigen die Ergebnisse, dass die empirischen Aufgabenschwierigkeiten durch die theoretisch angenommenen Aufgabenmerkmale erwartungskonform vorhergesagt werden können. Das Modell klärt mit 43 % einen bedeutsamen Anteil der Varianz der Itemschwierigkeiten auf. Die Varianzaufklärung fungiert als Indikator dafür, wie gut die auf Basis des Modells angenommenen Aufgabenschwierigkeiten die empirischen Aufgabenschwierigkeiten vorhersagen können (vgl. Hartig, 2007). Unter der Annahme, dass die Abstufungen der Aufgabenmerkmale die interindividuellen Unterschiede in den kognitiven Strukturen abbilden, können die Befunde hinsichtlich der Konstruktrepräsentation bzw. -validität demnach positiv gewertet werden: Die theoretischen Annahmen darüber wie das Konstrukt in kognitiver Hinsicht repräsentiert ist, können auch empirisch gestützt werden (vgl. Hartig & Frey, 2012; Winther, 2010). In diesem Zusammenhang sollte jedoch auch erwähnt werden, dass mehr als die Hälfte der Varianz nicht mithilfe der systematischen Beschreibung erklärt werden kann. Eine Erklärung dafür könnte sein, dass verschiedene schwierigkeitsbestimmende Merkmale nicht bereits bei der Konstruktion des Messinstruments umfassend berücksichtigt wurden. Eine theorie- und regelgeleite Konstruktion von Items auf Basis der hier angewendeten und ggf. weiteren Ansätzen, die *a priori* stattfindet, wird vermutlich zu einer höheren Varianzaufklärung und genaueren Beschreibungen der Kompetenzniveaus führen (vgl. Winther, 2010). Es könnte in Erwägung gezogen werden, die Ergebnisse aus der vorliegenden Analyse zur systematischen Weiterentwicklung und Optimierung von Testitems zu verwenden. Hierbei ließen sich bspw. Aufgaben generieren, die eine Vernetzung zwischen Wissenselementen detaillierter erfassen können (vgl. Kauertz et al., 2010).

Hinsichtlich der zweiten Fragestellung wird deutlich, dass das konstruierte Modell verschiedene Ausbildungsstadien erwartungskonform abbilden kann. Dies verweist auf eine hohe Änderungssensitivität des Kompetenzniveaumodells und deutet auf die Wirksamkeit der Lehramtsausbildung hinsichtlich eines Leistungszuwachses der Stu-

dierenden hin. Der Anteil Studierender im ersten Semester unterhalb des ersten Kompetenzniveaus ist unerwartet groß, allerdings lässt die Verteilung der Studierenden zum zweiten Messzeitpunkt annehmen, dass die im Studium behandelten Inhalte für den Kompetenzzuwachs bedeutsam sind. Diese Annahme wird dadurch bestärkt, dass etwa 50 % der in ihrem Studium fortgeschrittenen Studierenden mindestens das zweite Kompetenzniveau erreichen, welches durch fachsprachliche bzw. wissenschaftssprachliche Terminologie charakterisiert ist. Fortgeschrittene Studierende können dementsprechend eher Aufgaben lösen, die einen deutlichen Bezug zu wissenschaftlichen Theorien und Ansätzen aufweisen. Diese Inhalte spielen im Curriculum der akademischen Lehrer*innenausbildung eine wesentliche Rolle (vgl. König, 2012).

Im Rahmen der dritten Fragestellung wurde das Kompetenzmodell mit Blick auf ein diskriminantes Validitätskriterium untersucht. Es konnten keine bedeutsamen Effekte der Schulabschlussnoten auf die Kompetenzniveaus festgestellt werden. Eine Konfundierung des Konstrukts des pädagogischen Wissens mit dem Intelligenzkonzept ist demnach als gering einzustufen. Die Befunde bestätigen bisherige Ergebnisse (vgl. König, 2012) und können als Indizien für diskriminante Validität gewertet werden.

Hinsichtlich einer praktischen Verwertbarkeit des Kompetenzniveaumodells für die Lehrer*innenbildung rücken zwei Aspekte in den Fokus. (1) Angesichts eines Viertels der Proband*innen, die zum zweiten MZP noch unterhalb des ersten Niveaus verbleiben, ist eine Analyse naheliegend, ob es sich um eine Art „Risikogruppe“ handeln könnte. Die EMW-Studie bietet die Möglichkeit, diese Gruppe hinsichtlich motivationaler Eigenschaften und der Inanspruchnahme von Lerngelegenheiten genauer zu untersuchen. (2) Durch die Bildung von Kompetenzniveaus wurde deutlich, dass die Lösung von komplexen Aufgaben in hohem Maße anspruchsvoll für angehende Lehrer*innen ist und unter Umständen Förderungsbedarf besteht. Wir vermuten jedoch, dass im Studium weiter fortgeschrittene Studierende Aufgaben des höchsten Komplexitätsniveaus in größerer Zahl lösen können.

Zusammenfassend lässt sich festhalten, dass die Analysen des vorliegenden Beitrags die Ergebnisse von König (2009) in Grundzügen bestätigen. Auch wenn das in diesem Artikel aufgestellte und empirisch geprüfte Kompetenzniveaumodell grundsätzlich nicht als Verallgemeinerung für andere Testaufgaben bzw. -instrumente verstanden werden kann, so ist es dennoch von Interesse, ob die schwierigkeitsbestimmenden Aufgabenmerkmale auch bei anderen Testinstrumenten zum GPK herangezogen werden können. Im Hinblick auf zukünftige Forschung erscheint es empfehlenswert, den Bereich der Kompetenzniveaumodellierung auch in der Lehrer*innenbildung verstärkt in den Blick zu nehmen und differenzierte theoretische Modelle sowie deren Prüfung zu fokussieren.

Literatur

Altrichter, H., & Posch, P. (2007). *Lehrerinnen und Lehrer erforschen ihren Unterricht: Unterrichtsentwicklung und Unterrichtsevaluation durch Aktionsforschung*. Bad Heilbrunn: Klinkhardt.

Anderson, J. R. (1982). Acquisition of Cognitive Skills. *Psychological Review, 89*(4), 369–406.

Bastian, J., & Helsper, W. (2000). Professionalisierung im Lehrberuf – Bilanzierung und Perspektiven. In J. Bastian, W. Helsper, S. Reh & C. Schelle (Hrsg.), *Professionalisierung im Lehrerberuf: Von der Kritik der Lehrerrolle zur pädagogischen Professionalität* (S. 167–192). Opladen: Leske + Budrich.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520.

Beaton, E., & Allen, N. (1992). Interpreting Scales Through Scale Anchoring. *Journal of Educational Statistics, 17*(2), 191–204.

Berliner, D. C. (2004). Describing the Behavior and Documenting the Accomplishments of Expert Teachers. *Bulletin of Science, Technology & Society, 24*(3), 200–212.

Bieri, J., Atkins, A. L., Briar, S., Leaman, R. L., Miller, H., & Tripodi, T. (1966). *Clinical and Social Judgment*. New York: Wiley.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences*. Mahawah, NJ: Lawrence Erlbaum Associates.

Bromme, R. (1992). *Der Lehrer als Experte: zur Psychologie des professionellen Wissens*. Bern: Huber.

Combe, A., & Kolbe, F.-U. (2008). Lehrerprofessionalität: Wissen, Können, Handeln. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 857–875). Wiesbaden: VS Verlag.

Embretson, S. E. (1983). Construct Validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*(1), 179–197.

Gorin, J. S., & Embretson, S. E. (2006). Item Difficulty Modeling of Paragraph Comprehension Items. *Applied Psychological Measurement, 30*(5), 394–411.

Gruber, H., & Stöger, H. (2011). Experten-Novizen-Paradigma. In E. Kiel & K. Zierer (Hrsg.), *Basiswissen Unterrichtsgestaltung. Band 2: Unterrichtsgestaltung als Gegenstand der Wissenschaft* (S. 247–264). Baltmannsweiler: Schneider Verlag Hohengehren.

Harsch, C., & Hartig, J. (2011). Modellbasierte Definition von fremdsprachlichen Kompetenzniveaus am Beispiel der Bildungsstandards Englisch. *Zeitschrift für Interkulturelle Fremdsprachenforschung, 16*(2), 6–17.

Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Berlin: Springer.

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., & Frey, A. (2012). Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten. *Psychologische Rundschau, 63*(1), 43–49.

Hartig, J., & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17–36). Bonn/Berlin: BMBF.

Helsper, W. (2002). Wissen, Können, Nicht-Wissen-Können. Wissensformen des Lehrens und Konsequenzen für die Lehrerbildung. In Zentrum für Schulforschung und Fragen der Lehrerbildung (Hrsg.), *Die Lehrerbildung der Zukunft. Eine Streitschrift* (S. 67–86). Opladen: Leske & Budrich.

Huber, G. (1982). Kognitive Komplexität als Bedingung politischen Lernens. In W. Hagemann (Hrsg.), *Kognition und Moralität in politischen Lernprozessen* (S. 15–34). Opladen: Leske & Budrich.

Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden.* (61. Beiheft der Zeitschrift für Pädagogik, S. 11–31). Weinheim: Beltz.

Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften, 16,* 135–154.

Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms bei der DFG. *Zeitschrift für Pädagogik, 52*(6), 876–903.

KMK = Kultusministerkonferenz (2004). *Standards für die Lehrerbildung: Bildungswissenschaften* (Beschluss der Kultusministerkonferenz vom 16. 12. 2004). http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf [01. 02. 2019].

König, J. (2009). Zur Bildung von Kompetenzniveaus im Pädagogischen Wissen von Lehramtsstudierenden: Terminologie und Komplexität kognitiver Bearbeitungsprozesse als Anforderungsmerkmale von Testaufgaben? *Lehrerbildung auf dem Prüfstand, 2*(2), 244–262.

König, J. (2010). Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden (LEK): Theoretischer Rahmen, Fragestellungen, Untersuchungsanlage und erste Ergebnisse zu Lernvoraussetzungen von angehenden Lehrkräften. *Lehrerbildung auf dem Prüfstand, 3*(1), 56–83.

König, J. (2012). Die Entwicklung von pädagogischem Unterrichtswissen: Theoretischer Rahmen, Testinstrument, Skalierung und Ergebnisse. In J. König & Seifert, A. (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung* (S. 143–182). Münster: Waxmann.

König, J., & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerausbildung. *Zeitschrift für Erziehungswissenschaft, 12*(3), 499–527.

König, J., & Blömeke, S. (2010a). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich* (S. 253–273). Münster: Waxmann.

König, J., & Blömeke, S. (2010b). *Pädagogisches Unterrichtswissen (PUW). Dokumentation der Kurzfassung des TEDS-M-Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung.* Berlin: Humboldt-Universität.

König, J., Blömeke, S., Paine, L., Schmidt, B., & Hsieh, F.-J. (2011). General Pedagogical Knowledge of Future Middle School Teachers. On the complex ecology of teacher education in the United States, Germany, and Taiwan. *Journal of Teacher Education, 62*(2), 188–201.

König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M., Strauß, S., Bremerich-Vos, A., Fladung, I., & Kaiser, G. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? Struktur des professionellen Wissens bei angehenden Deutsch-, Englisch- und Mathematiklehrkräften im Studium. *Zeitschrift für Erziehungswissenschaft, 21*(3), 1–38.

König, J., & Klemenz, S. (2015). Der Erwerb von pädagogischem Wissen bei angehenden Lehrkräften in unterschiedlichen Ausbildungskontexten: Zur Wirksamkeit der Lehrerausbildung in Deutschland und Österreich. *Zeitschrift für Erziehungswissenschaft, 18*(2), 247–277.

König, J., Tachtsoglou, S., & Seifert, A. (2012). Individuelle Voraussetzungen, Lerngelegenheiten und der Erwerb von pädagogischem Professionswissen. In J. König & Seifert, A. (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung* (S. 243–283). Münster: Waxman.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV.* Münster: Waxmann.

Lüders, M. (2012). „Pädagogisches Wissen" – Eine Testkritik. *Zeitschrift für Erziehungswissenschaft, 15*(4), 775–791.

Mandl, H., & Huber, G. L. (1978). *Kognitive Komplexität.* Göttingen: Hogrefe.

Messick, S. (1995). Validity of Psychological Assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

OECD (2016). *Low-Performing Students: Why they fall behind and wow to help them succeed.* Paris: PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264250246-en [02. 02. 2019].

Peterson, C., & Scott, W.A. (1983). Toward Fundamental Measurement of Dimensionality. *British Journal of Social Psychology, 22,* 197–202.

Prenzel, M., Häußler, P., Rost, J., & Senkbeil, M. (2002). Der PISA Naturwissenschaftstest. Lassen sich die Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft, 30*(2), 120–135.

Rost, J. (2004a). *Lehrbuch Testtheorie Testkonstruktion.* Bern: Huber.

Rost, J. (2004b). Psychometrische Modelle zur Überprüfung von Bildungsstandards anhand von Kompetenzmodellen. *Zeitschrift für Pädagogik, 50*(5), 662–678.

Schaper, N. (2009). Aufgabenfelder und Perspektiven bei der Kompetenzmodellierung und -messung in der Lehrerbildung. *Lehrerbildung auf dem Prüfstand, 2*(1), 166–199.

Schön, D. (1983). The Reflective Practitioner. How professionals think in action. New York: Basic Books.

Scott, W.A. (1962). Cognitive Complexity and Cognitive Flexibility. *Sociometry, 25*(4), 405–414.

Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the new reform. *Harvard Educational Research, 57*(1), 1–22.

Terhart, E. (1991). Pädagogisches Wissen. Überlegungen zu seiner Vielfalt, Funktion und sprachlichen Form am Beispiel des Lehrerwissens. In J. Oelkers & H.-E. Tenorth (Hrsg.), *Pädagogisches Wissen* (27. Beiheft der Zeitschrift für Pädagogik S. 129–141). Weinheim: Beltz.

Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz.* Münster: Universität Münster.

Terhart, E. (2012). Wie wirkt Lehrerbildung? *Zeitschrift für Bildungsforschung, 2*(3), 3–21.

Terhart, E. (2014). Standards für die Lehrerbildung: Bildungswissenschaften – nach zehn Jahren. *Die deutsche Schule, 106*(4), 300–323.

von Eye, A. (1999). Kognitive Komplexität. Messung und Validität. Zeitschrift für Differentielle und Diagnostische Psychologie, (20)2, 81–96.

Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft, 18*(2), 187–223.

Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung.* Bielefeld: wbv.

Wipperfürth, M. (2015). *Professional Vision in Lehrernetzwerken: Berufssprache als ein Weg und ein Ziel von Lehrerprofessionalisierung.* Münster: Waxmann.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software* [computer program]. Camberwell: Australian Council for Educational Research.

Wyss, C. (2013). *Unterricht und Reflexion. Eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften.* Münster: Waxmann.

Abstract: The modelling of proficiency levels using criterion-based interpretations is rarely used in current teacher education research although it is considered to be more appropriate to describe outcomes than normative statements. However, standard-based teacher education requires information about the precise abilities of pre-service teachers. Only then can concrete statements be made about which persons have which competences. This article investigates how proficiency levels of pre-service teachers' pedagogical knowledge can be modelled. Pedagogical knowledge is assessed via a standardized paper-and-pencil test that was developed in the TEDS-M study. We examine a sample of 1451 future teachers from German and Austrian universities and teacher training colleges. The description of proficiency levels follows a method that predicts item difficulty by item attribute. The item attributes complexity of cognitive processes and terminological requirements are integrated into the model. They explain a large part of the variance in item difficulties. A frequency distribution of pre-service teachers differentiated by period of training supports our expectations. We found evidence supporting the construct validity of the model and its sensitivity to change.

Keywords: Professional Knowledge, Teacher Education, Proficiency Levels, Longitudinal Study, Pedagogical Knowledge

Anschrift der Autor_innen

Stefan Klemenz, Universität zu Köln,
Humanwissenschaftliche Fakultät, Institut für Allgemeine Didaktik und Schulforschung,
Gronewaldstr. 2a, Gebäude 214, 50931 Köln, Deutschland
E-Mail: stefan.klemenz@uni-koeln.de

Prof. Dr. Johannes König, Universität zu Köln,
Humanwissenschaftliche Fakultät, Institut für Allgemeine Didaktik und Schulforschung,
Gronewaldstr. 2a, Gebäude 214, 50931 Köln, Deutschland
E-Mail: johannes.koenig@uni-koeln.de