

Kunina-Habenicht, Olga; Maurer, Christina; Schulze-Stocker, Franziska; Wolf, Kristin; Hein, Nora; Leutner, Detlev; Seidel, Tina; Kunter, Mareike

## Zur curricularen Validität des BilWiss 2.0-Tests zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften

*Zeitschrift für Pädagogik 65 (2019) 4, S. 542-556*



### Quellenangabe/ Reference:

Kunina-Habenicht, Olga; Maurer, Christina; Schulze-Stocker, Franziska; Wolf, Kristin; Hein, Nora; Leutner, Detlev; Seidel, Tina; Kunter, Mareike: Zur curricularen Validität des BilWiss 2.0-Tests zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften - In: Zeitschrift für Pädagogik 65 (2019) 4, S. 542-556 - URN: urn:nbn:de:0111-pedocs-239941 - DOI: 10.25656/01:23994

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-239941>

<https://doi.org/10.25656/01:23994>

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](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

# ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 4

Juli/August 2019

■ *Thementeil*

---

**Kompetenzerwerb zukünftiger LehrerInnen  
in der universitären Ausbildung**

■ *Allgemeiner Teil*

---

Der Einfluss von Lehrkräfteverbänden in der Steuerung  
von Schulsystemen: Deutschland und Frankreich  
im Vergleich

Zur Situationsspezifität des pädagogischen Ethos:  
Eine empirische Studie im Bereich der betrieblichen  
Berufsbildung

Forschendes Lernen prüfen: Hochschuldidaktische  
Gedanken zu einer Theorie des Prüfens

## Inhaltsverzeichnis

### *Thementeil: Kompetenzerwerb zukünftiger LehrerInnen in der universitären Ausbildung*

<i>Olga Zlatkin-Troitschanskaia/Hans Anand Pant/Kai Cortina</i> Kompetenzerwerb zukünftiger LehrerInnen in der universitären Ausbildung. Einführung in den Thementeil .....	469
<i>Christoph Vogelsang/Andreas Borowski/David Buschhüter/ Patrick Enkrott/Maren Kempin/Christoph Kulgemeyer/Peter Reinhold/ Josef Riese/Horst Schecker/Jan Schröder</i> Entwicklung von Professionswissen und Unterrichtsperformanz im Lehramtsstudium Physik – Analysen zu valider Testwertinterpretation .....	473
<i>Sabrina Mathesius/Moritz Krell/Annette Upmeyer zu Belzen/Dirk Krüger</i> Überprüfung eines Tests zum wissenschaftlichen Denken unter Berücksichtigung des Validitätskriteriums relations-to-other-variables ...	492
<i>Colin Jeschke/Christiane Kuhn/Anke Lindmeier/ Olga Zlatkin-Troitschanskaia/Hannes Saas/Aiso Heinze</i> What is the Relationship Between Knowledge in Mathematics and Knowledge in Economics? Investigating the professional knowledge of (pre-service) teachers trained in two subjects .....	511
<i>Lara Pohle/Georg Hosoya/Catharina Loftfield/Lars Jenßen</i> Indicators Measuring Preschool Teachers' Stimulation Quality – Theoretical background and empirical testing .....	525
<i>Olga Kunina-Habenicht/Christina Maurer/Franziska Schulze-Stocker/ Kristin Wolf/Nora Hein/Detlev Leutner/Tina Seidel/Mareike Kunter</i> Zur curricularen Validität des BilWiss 2.0-Tests zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften .....	542

*Deutscher Bildungsserver*

Linktipps zum Thema „Kompetenzerwerb zukünftiger LehrerInnen  
in der universitären Ausbildung“ ..... 557

*Allgemeiner Teil*

*Michael Dobbins/Rita Nikolai*

Der Einfluss von Lehrkräfteverbänden in der Steuerung von Schulsystemen:  
Deutschland und Frankreich im Vergleich ..... 564

*Sarah Forster-Heinzer*

Zur Situationsspezifität des pädagogischen Ethos: Eine empirische Studie  
im Bereich der betrieblichen Berufsbildung ..... 584

*Gabi Reinmann*

Forschendes Lernen prüfen: Hochschuldidaktische Gedanken  
zu einer Theorie des Prüfens ..... 608

*Dokumentation*

Erziehungswissenschaftliche Habilitationen und Promotionen 2018 ..... 627

Impressum ..... U3

## Table of Contents

### *Topic: Prospective Teachers' Competence Development in Higher Education*

<i>Olga Zlatkin-Troitschanskaia/Hans Anand Pant/Kai Cortina</i> Prospective Teachers' Competence Development in Higher Education. An introduction .....	469
<i>Christoph Vogelsang/Andreas Borowski/David Buschhüter/ Patrick Enkrott/Maren Kempin/Christoph Kulgemeyer/Peter Reinhold/ Josef Riese/Horst Schecker/Jan Schröder</i> Development of Professional Knowledge and Teaching Skills in Academic Pre-service Physics Teacher Education: Validity analyses concerning the interpretation of test scores .....	473
<i>Sabrina Mathesius/Moritz Krell/Annette Upmeier zu Belzen/Dirk Krüger</i> An Evaluation of a Scientific Reasoning Test for Pre-service Biology Teachers: Using relations-to-other-variables as source of validity evidence .....	492
<i>Colin Jeschke/Christiane Kuhn/Anke Lindmeier/ Olga Zlatkin-Troitschanskaia/Hannes Saas/Aiso Heinze</i> What is the Relationship Between Knowledge in Mathematics and Knowledge in Economics? Investigating the professional knowledge of (pre-service) teachers trained in two subjects .....	511
<i>Lara Pohle/Georg Hosoya/Catharina Loftfield/Lars Jenßen</i> Indicators Measuring Preschool Teachers' Stimulation Quality – Theoretical background and empirical testing .....	525
<i>Olga Kunina-Habenicht/Christina Maurer/Franziska Schulze-Stocker/ Kristin Wolf/Nora Hein/Detlev Leutner/Tina Seidel/Mareike Kunter</i> On the Curricular Validity of the Bilwiss-2.0-Test for the Assessment of (Pre-service) Teachers' Educational Knowledge .....	542
<i>Deutscher Bildungsserver</i> Online Ressources "Prospective Teachers' Competence Development in Higher Education" .....	557

*Articles*

*Michael Dobbins/Rita Nikolai*

The Influence of Teachers' Unions on Output- and Competition-Oriented  
School Governance: Germany and France in comparison ..... 564

*Sarah Forster-Heinzer*

On the Situational Specificity of the Pedagogical Ethos:  
An empirical study in the field of vocational workplace training ..... 584

*Gabi Reinmann*

Assessing Research-Based Learning: Thoughts about a theory  
of assessment in higher education ..... 608

Habilitation Treatises and Dissertations in Education in 2018 ..... 627

Impressum ..... U3

Olga Kunina-Habenicht/Christina Maurer/Franziska Schulze-Stocker/Kristin Wolf/Nora Hein/Detlev Leutner/Tina Seidel/Mareike Kunter<sup>1</sup>

## Zur curricularen Validität des BilWiss 2.0-Tests zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften

**Zusammenfassung:** Im bildungswissenschaftlichen Studienteil der universitären Lehrerbildung in Deutschland sollen angehende Lehrkräfte fächerübergreifendes Grundlagenwissen aufbauen, das sie zur professionellen Ausübung ihres Berufs benötigen. Im Forschungsprogramm BilWiss wurde zunächst im Kontext des Landes Nordrhein-Westfalen ein Wissenstest entwickelt, der zentrale Inhalte des bildungswissenschaftlichen Studienteils erfasst. In zwei Studien untersucht der vorliegende Beitrag, ob die Testinhalte auch bundesweit dem intendierten bildungswissenschaftlichen Curriculum in der universitären bildungswissenschaftlichen Lehrerbildung entsprechen. Sowohl länderspezifische Experteneinschätzungen als auch ein Vergleich der Testinhalte mit bundesweiten bildungswissenschaftlichen Standards für die Lehrerbildung weisen darauf hin, dass der Test bundesländerübergreifend anwendbar ist.

**Schlagnote:** Wissenstest, Lehrerbildung, Bildungswissenschaften, Validität, länderübergreifend

### 1. Die Bildungswissenschaften im Lehramtsstudium

Die erste Phase der Lehrerbildung in Deutschland umfasst verschiedene Studienanteile wie Fachwissenschaften, Fachdidaktiken, die sogenannten *Bildungswissenschaften* sowie Praxisphasen. Die Bildungswissenschaften fassen dabei den fächerübergreifenden Studienteil zusammen, innerhalb dessen Lehramtsstudierende Wissen in den Bereichen Psychologie, Erziehungswissenschaft, Soziologie und verwandten Wissenschaften aufbauen sollen. Dementsprechend wird dieses sowohl auf den Unterricht bezogene als auch über den Unterricht hinausgehende fächerunabhängige Wissen von Lehrkräften als *bildungswissenschaftliches Wissen* bezeichnet und als Aspekt der professionellen Kompetenz von Lehrkräften verstanden (vgl. Linninger et al., 2015; Kunter et al., 2017).

Mit den 2004 von der Ständigen Konferenz der Kultusminister der Länder (KMK) verabschiedeten und 2014 aktualisierten bildungswissenschaftlichen Standards wurden curriculare Ziele formuliert, die gemeinsam mit den Lehrerausbildungsgesetzen der Länder die Lernziele des bildungswissenschaftlichen Teils der Lehrerbildung, also dessen intendiertes Curriculum abbilden (UNESCO International Bureau of Education, 2013). Sie beziehen sich auf die für den Lehrerberuf zentralen Kompetenzbereiche *Un-*

---

1 Olga Kunina-Habenicht und Christina Maurer sind beiderseits die Erstautorinnen des Beitrags.

*terrachten, Erziehen, Beurteilen* und *Innovieren* (KMK, 2014). Die KMK-Standards weisen (mittlerweile) eine hohe Akzeptanz innerhalb des fachlichen Diskurses auf (vgl. Lohmann, Seidel & Terhart, 2011; Terhart, 2014).

Doch nicht nur die Ebene des intendierten Curriculums ist entscheidend für die Entwicklung bildungswissenschaftlichen Wissens bei angehenden Lehrkräften (vgl. Schulze-Stocker, Holzberger & Lohse-Bossenz, 2017). Die Realisierung dieser Vorgaben erfolgt an Universitäten in Form des potentiellen Curriculums (universitätsspezifische Modulhandbücher, Prüfungs- und Studienordnungen), das z. B. festlegt, welche Inhalte und Kompetenzen im Rahmen des Lehramtsstudiums vermittelt werden sollen. Beim Vergleich bildungswissenschaftlicher Curricula an deutschen Hochschulen zeigte sich auf dieser Ebene bereits eine zunehmende Anpassung an die KMK-Standards (vgl. Hohenstein, Zimmermann, Kleickmann, Köller & Möller, 2014; Schulze-Stocker, 2016).

Die Umsetzung des potentiellen Curriculums in konkrete Lehrveranstaltungen wird als implementiertes Curriculum bezeichnet und die daraus resultierenden kurzfristigen und längerfristigen Wirkungen auf Seiten der Lernenden, wie z. B. deren tatsächliche Wissensentwicklungen, als erreichtes Curriculum (vgl. UNESCO International Bureau of Education, 2013). Bildungswissenschaftliche Studienanteile im Lehramtsstudium wurden in der Vergangenheit mehrfach für fehlende Kohärenz und Praxisnähe kritisiert (vgl. Merzyn, 2002; Veenman, 1984). Dieser Kritik entgegen steht die Hypothese, dass eine theoretische Wissensbasis als Teil der professionellen Kompetenz von Lehrkräften eine notwendige Grundlage für späteres professionelles Verhalten im Beruf darstellt und dabei auch die fächerübergreifenden Inhalte der Bildungswissenschaften umfasst (vgl. Baumert & Kunter, 2006; Shulman, 1986).

## **2. Messung des bildungswissenschaftlichen Wissens**

In den letzten zehn Jahren wurden mehrere Verfahren zur Erfassung des unterrichtsbezogenen, pädagogisch-psychologischen Wissens von Lehrkräften entwickelt (vgl. u. a. König & Blömeke, 2009; Lenske, Thillmann, Wirth, Dicke & Leutner, 2015; Seifert, Hilligus & Schaper, 2009; Voss, Kunter & Baumert, 2011; für eine Übersicht vgl. Voss, Kunina-Habenicht, Hoehne & Kunter, 2015). Doch neben den Inhalten, die einen unmittelbaren Bezug zum Unterrichtsgeschehen aufweisen, sind auch Wissensbestände relevant, die über den Unterrichtskontext hinausgehen, wie bspw. Wissen über Beratung, Schulorganisation oder Bildungstheorie (vgl. Shulman, 1986; Voss et al., 2015).

Im Forschungsprogramm BilWiss<sup>2</sup> wurde ein Testverfahren zur Erfassung dieses konzeptuell breiter angelegten bildungswissenschaftlichen Wissens bei angehenden und

---

2 Das Forschungsprogramm wurde gefördert vom Bundesministerium für Bildung und Forschung: 1. Förderphase „BilWiss“ (2009–2013; FKZ 01JH0910), 2. Förderphase „BilWiss-Beruf“ (2012–2016; FKZ 01PK11007), 3. Förderphase „BilWiss-UV“ (2016–2019; FKZ 01PK15007).

praktizierenden Lehrkräften (zunächst) in Nordrhein-Westfalen (NRW) entwickelt (vgl. Kunter et al., 2017; Schulze-Stocker et al., 2017). Als eine zentrale Grundlage für die Testentwicklung wurden in einer Delphi-Studie durch Konsensfindung über 100 relevante bildungswissenschaftliche Themen aus den Bereichen Unterricht, Lernen, Entwicklung, Sozialisation, Diagnostik und Evaluation, Heterogenität und soziale Konflikte, Bildungstheorie, Bildungssystem und Schulorganisation sowie Lehrerberuf als zentrale bildungswissenschaftliche Wissensinhalte identifiziert (vgl. Kunina-Habenicht et al., 2012). An der Delphi-Studie nahmen 49 Expertinnen und Experten aus der ersten und zweiten Phase der Lehrerbildung in NRW teil, wobei in der ersten Phase alle lehrerbildenden Universitäten in NRW vertreten waren, sodass die Ergebnisse der Studie repräsentativ für erste Phase Lehrerbildung in NRW sind. Anschließend wurden für die ermittelten relevanten Delphi-Themen konkrete Wissensitems entwickelt und erprobt (vgl. Linninger et al., 2015). Nach einer mehrjährigen Erprobungsphase des Tests mit unterschiedlichen Pilotierungen und Validierungsstudien umfasst der aktuelle Wissenstest („BilWiss-2.0-Test“) 70 bildungswissenschaftliche Themen und besteht aus 119 Multiple-Choice-Items, die sich auf die sechs Inhaltsbereiche *Unterrichtsgestaltung*, *Lernen und Entwicklung*, *Diagnostik und Evaluation*, *Bildungstheorie*, *Schule als Bildungsinstitution* und *Lehrerberuf als Profession* verteilen (vgl. Kunina-Habenicht et al., 2019). Die Inhaltsbereiche des Wissenstests mit den von ihnen abgedeckten bildungswissenschaftlichen Themen sind in Tabelle 1 dargestellt.

Mit dem BilWiss-2.0-Test sollen insbesondere Wissensbestände über bildungswissenschaftliche Inhalte erfasst werden, die im Rahmen der universitären Lehrerbildung thematisiert werden. Dieser Test soll zum einen als Forschungsinstrument zur Beantwortung von Forschungsfragen dienen bspw. in Bezug auf Unterschiede zwischen verschiedenen Gruppen, Veränderung des Wissens oder auch bzgl. der Bedeutsamkeit des theoretischen Wissens für praktisches Handeln. Bisherige Befunde zur Messgüte im Sinne der aktuell geltenden Standards psychologischen Testens (vgl. AERA, APA & NCME, 2014) weisen auf eine valide Erfassung bildungswissenschaftlichen Wissens mithilfe der Testitems hin. Zum einen bestätigen kognitive Interviews, dass zum Lösen der Testitems universitäres Wissen genutzt wird, zum anderen erzielen Absolventinnen und Absolventen eines Lehramtsstudiums höhere Testwerte als Quereinsteigerinnen und Quereinsteiger bzw. als Lehramtsstudierende am Beginn ihres Studiums (vgl. Kunina-Habenicht et al., 2013; Linninger et al., 2015). Zudem spiegeln sich bildungswissenschaftliche Studienfortschritte und Studienschwerpunkte in höheren Testwerten wider (vgl. Kunina-Habenicht et al., 2019; Schulze-Stocker, Holzberger, Kunina-Habenicht, Terhart & Kunter, 2016). Auch Zusammenhänge mit Berufserfolgskriterien als Hinweise für kriteriale Validität der Testwerte werden sichtbar: Mit dem Test gemessenes Wissen hängt positiv mit einer positiveren Beurteilung des eigenen Unterrichts (vgl. Lohse-Bossenz, Kunina-Habenicht, Dicke, Leutner & Kunter, 2015) zusammen und kann Beeinträchtigungen des beruflichen Wohlbefindens abpuffern (vgl. Dicke et al., 2015). Für eine ausführliche Beschreibung des Testinstruments und dessen Messgütekriterien wird auf Kunina-Habenicht u. a. (2019) verwiesen.

**Unterrichtsgestaltung****A) Lernwirksame Merkmale im Unterricht**

- Wissen um Klassenführung und Forschung zur Klassenführung
- Modelle der Lehrer-Schüler-Interaktion/Konflikte und Kooperation im Unterricht/Unterricht als soziale Situation
- Transparenz von Zielen und Anforderungen
- Kognitive Aktivierung
- Partizipation der Schülerinnen u. Schüler
- Allgemeine und fachliche Lernvoraussetzungen
- Unterstützung bei Lernprozessen
- Transfer auf neue/andere Aufgabenbereiche
- Konsolidierung des Gelernten
- Positives Lernklima/anregende Lernumgebung
- Konstruktiver Umgang mit Fehlern/Analyse von Schülerfehlern

**B) Didaktische und methodische Gestaltung von Unterricht**

- Allgemeine Didaktik: Lehr-lern-theoretische Modelle
- Allgemeine Didaktik: Konstruktivistische Modelle
- Methodische Formen des Unterrichts
- Methodische Organisation und Sozialformen
- Forschung zu Unterrichtsmethoden
- Vorbereitung variabler/differenzierter Unterrichtsverläufe
- Gesprächsführung und Frageformate
- Gestaltung von Rückmeldungen an Schülerinnen u. Schüler

**Lernen und Entwicklung****A) Entwicklungsprozesse**

- Kognitive Entwicklung
- Entwicklung der Motivation, Emotion- und Handlungsregulation
- Bedeutung von Familie, Gleichaltrigen und Schule für psychosoziale Entwicklung
- Entwicklung der Sozialen Kognition
- Persönlichkeit- und Selbstkonzeptentwicklung
- Moralische Entwicklung

**B) Lernprozesse**

- Lernmotivation
- Modelllernen
- Metakognition und selbstreguliertes Lernen
- Attribution
- Wissenserwerb
- Selbstwirksamkeit, Akademisches Selbstkonzept
- Gedächtnismodelle
- Verhaltensorientierte Theorien
- Motivationale Störungen und Interventionsansätze

**C) Sozialisationsprozesse**

- Sozialisation und Identitätsentwicklung
- Sozialisation und Jugend
- Sozialisation und Geschlecht
- Heterogenität bzgl. des Geschlechts
- Heterogenität bzgl. des sozialen Hintergrunds

**Diagnostik und Evaluation**

- Methoden und Verfahren
- Leistungs- und Verhaltensbeurteilung
- Grundbegriffe Gütekriterien
- Interpretation von Testwerten
- Grundbegriffe der Diagnostik
- Deskriptive Statistik
- Qualitative Methoden
- Grundbegriffe empirischer Studien
- Evaluationsdesigns

**Bildungstheorie**

- Bildungsphilosophische Betrachtung aktueller Bildungsfragen
- Anthropologische und philosophische Grundfragen von Erziehung und Bildung
- Strukturprobleme des deutschen Bildungssystems
- Normen und Normenbegründung in Erziehung und Bildung
- Gleichheit und Gerechtigkeit in Bildungskontexten
- Ethik pädagogischer Berufe
- Sozialer Wandel

**Schule als Bildungsinstitution**

- Neue Steuerung/Educational Governance
- Ökologie von Schule, Schulkultur, Schulleben
- Effektivität der Einzelschule
- Struktur des Bildungssystems
- Gegenstände von Schultheorien
- Rechtsstellung von Lehrkräften
- Heterogenität des kulturellen Hintergrunds
- Konzepte zum Umgang mit kultureller Heterogenität

**Lehrerberuf als Profession**

- Pädagogische Professionalität/ Professionelles Lehrerhandeln
- Theorien und Forschungsansätze zum Lehrerberuf
- Aufgabenbereiche/formale Anforderungen an den Lehrerberuf
- Forschung zu Belastung und Beanspruchung
- Stressmodelle/Techniken des Stressmanagements
- Wissen über Beratung und Beratungssituationen
- Beratung von Schülerinnen, Schülern und Eltern
- Grundprinzipien kooperativen Arbeitens
- Konfliktarten und deren Bedingungen

Tab. 1: Bildungswissenschaftliche Themen, die der BilWiss-2.0-Test abdeckt

Der BilWiss-2.0-Test soll darüber hinaus auch als Instrument für Bildungsmonitoring bzw. Evaluationszwecke eingesetzt werden, bspw. zur Überprüfung der Wirksamkeit der institutionellen Lehrerbildung in verschiedenen Bundesländern. Da der Test im Kontext der Lehrerbildung im Land NRW entwickelt wurde, ist bislang noch ungeklärt, wie gut die Testinhalte auch die bildungswissenschaftlichen Inhalte der universitären Lehrerbildung der anderen Bundesländer repräsentieren. Bevor der Test also auch außerhalb von NRW eingesetzt werden kann, muss zunächst geprüft werden, ob die Testwertinterpretationen auch länderübergreifend curricular (im Sinne eines intendierten Curriculums) valide sind. An dieser Stelle knüpft der vorliegende Beitrag mit dem Ziel an, Hinweise auf die curriculare Validität der Testwerte – als einen wichtigen Aspekt der Inhaltsvalidität – im Kontext der länderspezifischen universitären Lehrerbildung zu gewinnen und damit Aussagen zur bundesweiten Anwendbarkeit des Tests treffen zu können.

### 3. Fragestellungen dieses Beitrags

Insgesamt geht es in diesem Beitrag also darum zu prüfen, inwieweit die im Test behandelten Inhalte das intendierte Curriculum für den bildungswissenschaftlichen Studienanteil auch außerhalb des Landes NRW widerspiegeln. Hierzu wurden zwei Studien durchgeführt. Studie 1 überprüft auf der Ebene der bildungswissenschaftlichen Delphi-Themen, welche die theoretische Basis für die konstruierten Testitems darstellten, die Relevanz der Inhalte auch außerhalb des Landes NRW. Dazu schätzten Vertretungen der Kultusministerien der Bundesrepublik Deutschland die Relevanz dieser bildungswissenschaftlichen Themen für die universitäre Lehrerbildung in ihrem Bundesland ein. Mithilfe dieser Studie sollten zwei Fragestellungen beantwortet werden:

- 1a) Gibt es einen Konsens der Landesvertretungen über die Relevanz der einzelnen Themen, oder weichen die Einschätzungen zwischen einzelnen Bundesländern stark voneinander ab?
- 1b) Wie relevant werden die Themen der sechs Inhaltsbereiche für die universitäre Lehrerbildung außerhalb des Landes NRW beurteilt?

Studie 2 prüft, ob auch auf der Ebene der einzelnen Testitems eine bundesweite Relevanz der bildungswissenschaftlichen Testinhalte festgestellt werden kann. Dazu erfolgte in einer Dokumentenanalyse die Zuordnung der Inhalte der im BilWiss-2.0-Test enthaltenen Items zu den bundesweit geltenden KMK-Standards für die Bildungswissenschaften in der Lehrerbildung (für die theoretischen Ausbildungsabschnitte). Das Ziel war hier die Beantwortung der folgenden Fragestellungen:

- 2a) Inwieweit entsprechen die Testinhalte den bundesweit geltenden KMK-Standards?
- 2b) Wie verteilt sich die Passung der Testitems auf die KMK-Bereiche *Unterrichten*, *Erziehen*, *Beurteilen* und *Innovieren*?

## 4. Studie 1: Umfrage der ministerialen Vertretungen zur Relevanz der bildungswissenschaftlichen Themen

### 4.1 Methode

In Studie 1 bewerteten ministeriale Vertretungen aus 13 Bundesländern<sup>3</sup> die Relevanz der bildungswissenschaftlichen Themen, die dem BilWiss-2.0-Test zugrunde liegen, für die universitäre Lehrerbildung in ihrem Bundesland. Die Rekrutierung der Expertinnen und Experten erfolgte durch Anschreiben an die für die Lehrerbildung in den Ministerien zuständigen Stellen, vermittelt über das Sekretariat der KMK. Diese Stellen haben dann einen oder zwei einschlägige, kompetente und unmittelbar im Bereich der Lehrerbildung tätige Mitarbeitende in der jeweiligen Institution beauftragt, an der schriftlichen Befragung teilzunehmen. Alle Landesvertretungen schätzten für jedes der 70 Themen, gegliedert nach den Inhaltsbereichen des Tests (siehe Tab. 1), die Bedeutung auf einer 4-stufigen Skala von 1 („nicht relevant“) bis 4 („uneingeschränkt relevant“) ein. Die genaue Instruktion lautete „Ihre Aufgabe ist es, die vorgestellten Themen dahingehend zu beurteilen, ob sie aus Ihrer Sicht für die universitäre Lehrer/innenbildung in Ihrem Bundesland relevant sind.“ In drei Bundesländern gaben aufgrund landesinterner unterschiedlicher Zuständigkeitsbereiche jeweils zwei Personen ihre Einschätzungen ab. Vor weiteren Analysen wurden die Mittelwerte der Angaben der beiden Vertretungen gebildet, um kein Bundesland stärker zu gewichten. Aus allen anderen Ländern lagen Bewertungen durch jeweils eine Person vor.

Um den Konsens der Einschätzungen der verschiedenen Landesvertretungen zu ermitteln, wurde für jedes der 70 Themen der für kleine Gruppen adjustierte *Average Deviation Index* ( $ADM_{adj}$ ; vgl. Lohse-Bossenz, Kunina-Habenicht & Kunter, 2014) über die 13 Bundesländer hinweg berechnet. Dieser Kennwert gibt die mittlere Abweichung innerhalb einer Gruppe von (mehr als zwei) Beurteilenden wieder, wobei hohe ADM-Werte eine niedrige Übereinstimmung indizieren. Eine gute Beurteilungsübereinstimmung liegt vor, wenn der berechnete Index unterhalb des kritischen Werts  $ADM_{crit} = c/6$  liegt, wobei  $c$  die Anzahl der Beurteilungskategorien darstellt (vgl. Burke & Dunlap, 2002). In der vorliegenden Studie ergibt sich mit  $c = 4$  dementsprechend 0.67 als kritischer Wert.

Im ersten Schritt wurde die Beurteilungsübereinstimmung für jedes einzelne Thema bestimmt. Im zweiten Schritt wurden diese Werte für jeden Inhaltsbereich des Wissens-tests aggregiert, indem die mittlere Beurteilungsübereinstimmung und Streuung über die betreffenden Themen im jeweiligen Inhaltsbereich hinweg berechnet wurden. In Ta-

<sup>3</sup> Zwei Bundesländer nahmen keine Bewertungen vor, da von deren Vertretungen rückgemeldet wurde, dass sich die formalen Vorgaben zu Inhalten der Lehrerbildung momentan in Bearbeitung befinden und daher zum Zeitpunkt der Erhebung keine Aussage möglich sei. Von NRW lagen die Einschätzungen durch die Delphi-Studie (vgl. Kunina-Habenicht et al., 2012) bereits vor, daher wurde in diesem Land nicht erneut um eine Einschätzung gebeten.

belle 2 wurden die Relevanzeinschätzungen der einzelnen Themen zwecks einer übersichtlicheren Ergebnisdarstellung für jeden Inhaltsbereich ebenfalls aggregiert.

## 4.2 Ergebnisse

Die Übereinstimmung der Landesvertretungen pro Inhaltsbereich (Tab. 2) lag mit Werten von 0.50 bis 0.59 in fünf von sechs Inhaltsbereichen zufriedenstellend unter dem kritischen Wert  $ADM_{crit} = 0.67$ . Das zeigt, dass generell ein hoher Konsens bezüglich der Relevanz der bildungswissenschaftlichen Themen besteht. Lediglich bezüglich der Einschätzungen des Inhaltsbereichs *Bildungstheorie* gingen die Meinungen der Landesvertretungen weiter auseinander; dort lag die Beurteilungsübereinstimmung bei fünf von sieben Themen über dem kritischen Wert und damit auch die mittlere Beurteilungsübereinstimmung mit 0.71 leicht darüber. Die Relevanzeinschätzungen der jeweiligen Landesvertretungen sind in Abbildung 1 dargestellt, jeweils über die einzelnen Themen jedes Inhaltsbereichs gemittelt. Für die fünf Inhaltsbereiche mit guter Übereinstimmung über die Landesvertretungen hinweg ist außerdem ein Gesamtmittelwert der Einschätzungen pro Inhaltsbereich dargestellt. Für den Inhaltsbereich *Bildungstheorie* wird wegen der unzureichenden Beurteilungsübereinstimmung von einer Mittelung über alle Landesvertretungen abgesehen.

Aus Abbildung 1 ist ersichtlich, dass die Relevanzeinschätzungen vorwiegend im Bereich von 2.5 bis 4.0 liegen, d. h. die Themen aller Inhaltsbereiche werden im Mittel zwischen „eher relevant“ und „uneingeschränkt relevant“ eingeschätzt. Eine Ausnahme davon bilden die Einschätzungen einer einzelnen Landesvertretung. Hier werden vier von sechs Inhaltsbereichen unterhalb des theoretischen Skalenmittelwerts eingeschätzt, also als weniger relevant. Diese vergleichsweise niedrigen Bewertungen wurden von der Vertretung mit einer höheren Relevanz dieser Themen erst in späteren Lehrerbildungsphasen begründet. Zwei weitere Landesvertretungen schätzen die Themen des Inhaltsbereichs *Bildungstheorie* im Mittel ebenfalls als weniger relevant ein.

Betrachtet man darüber hinaus die Relevanzbewertungen auf der Ebene der einzelnen Themen (vgl. Tab. 1) gemittelt über die Landesvertretungen, ist das einzige Thema,

$ADM_{adj}$	Unterrichtsgestaltung	Lernen und Entwicklung	Diagnostik und Evaluation	Lehrerberuf als Profession	Schule als Bildungsinstitution	Bildungstheorie
<i>M</i>	0.57	0.58	0.50	0.59	0.55	0.71
<i>SD</i>	0.06	0.07	0.09	0.07	0.11	0.11

Werte kleiner als .67 indizieren eine gute Beurteilerübereinstimmung

Tab. 2: Mittlere Beurteilungsübereinstimmungen zwischen den Bundesländern nach Inhaltsbereichen.  $ADM_{adj}$  = adjusted Average Deviation Index; *M* = Mittelwert; *SD* = Standardabweichung

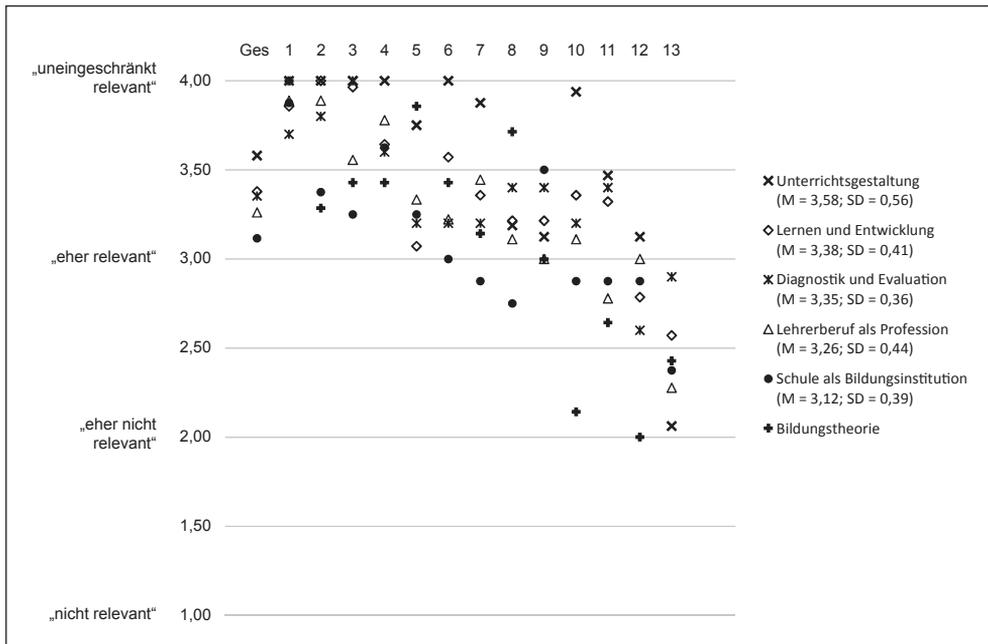


Abb. 1: Relevanzeinschätzungen der Landesvertretungen 1 bis 13. Ges = Gesamtmittelwert über die Landesvertretungen hinweg berechnet, aufgrund unzureichender Beurteilungsübereinstimmung jedoch aber nicht für den Inhaltsbereich Bildungstheorie

dessen mittlere Relevanz mit 2.4 ( $SD = 0.7$ ) unterhalb des theoretischen Skalenmittelwerts von 2.5 bewertet wird, das Thema „Struktur des Bildungssystems in Deutschland, Besonderheiten im internationalen Vergleich, Entwicklungstendenzen“ des Inhaltsbereichs *Schule als Bildungsinstitution*. Die höchste Relevanzbewertung ( $M = 3.8$ ;  $SD = 0.4$ ) erhält das Thema „Heterogenität bezüglich des kulturellen Hintergrunds, Chancengleichheit“ aus dem gleichen Inhaltsbereich. Insgesamt sprechen die Ergebnisse von Studie 1 für eine große Einigkeit zwischen den Bundesländern und eine zufriedenstellende Bedeutsamkeit der dem Test zugrundeliegenden bildungswissenschaftlichen Themen.

## 5. Studie 2: Überprüfung der Passung von Testitems und KMK-Standards

### 5.1 Methode

Im Gegensatz zur Studie 1, bei der eine Befragung zur Relevanz der Delphi-Themen erfolgte, wurde in Studie 2 eine Dokumentenanalyse der Testitems im BilWiss-Wissenstest auf der Basis der KMK-Standards vorgenommen, um die Relevanz der Bil-

Wiss-2.0-Testinhalte für die Bildungswissenschaften in der bundesweiten universitären Lehrerbildung auch auf der Ebene der einzelnen Testitems zu überprüfen. Nach einer ausführlichen mündlichen Einweisung bezüglich des Kodierauftrags waren an der Dokumentenanalyse insgesamt drei Beurteilerinnen (Projektmitarbeiterinnen) beteiligt. Dabei übernahm Beurteilerin A alle Inhaltsbereiche und die Beurteilerinnen B bzw. C übernahmen jeweils die Zweitbeurteilung für drei Inhaltsbereiche. Aufgabe der Beurteilerinnen war es, für jedes Testitem festzustellen, ob es möglich ist, dieses Item einem oder mehreren KMK-Standards zuzuordnen. Somit war es möglich, dass sich sowohl Passungen eines Items mit mehreren Standards ergeben als auch Passungen eines Standards mit mehreren Items. Die Beurteilerübereinstimmung der Item-Standard-Passungen wurde über den Koeffizienten  $\kappa_n$  berechnet (vgl. Brennan & Prediger, 1981). Die Übereinstimmung zwischen den Beurteilerinnen war mit  $\kappa_n = .95$  bzw.  $\kappa_n = .96$  in beiden Fällen sehr gut. Die Beurteilerinnen überarbeiteten Abweichungen zwischen der Erst- und Zweitbeurteilung anschließend in einer gemeinsamen Konsensfindung.

## 5.2 Ergebnisse

Es wurden für 115 von 119 Testitems (96.6%) Item-Standard-Passungen ermittelt: 80 Items stimmen inhaltlich mit genau einem Standard überein, 33 Items mit zwei Standards und jeweils ein Item mit drei bzw. vier Standards. Bei den Inhaltsbereichen *Lernen und Entwicklung*, *Diagnostik und Evaluation* sowie *Schule als Bildungsinstitution* gab es Passungen für jeweils alle Items, bei *Unterrichtsgestaltung* sind es 22 von 23 Items, bei *Lehrerberuf als Profession* 17 von 18 Items und bei *Bildungstheorie* zwölf von 14 Items. In Tabelle 3 ist dargestellt, wie sich alle für die Inhaltsbereiche des Tests ermittelten Item-Standard-Passungen absolut und prozentual auf die KMK-Kompetenzbereiche verteilen.

Umgekehrt decken die Item-Standard-Passungen den Großteil der KMK-Standards ab, d. h. elf von zwölf Standards im Kompetenzbereich *Unterrichten*, acht von zwölf Standards im Kompetenzbereich *Erziehen* und jeweils acht von zehn Standards in den Kompetenzbereichen *Beurteilen* und *Innovieren*.<sup>4</sup> Insgesamt ist zu erkennen, dass es eine hohe inhaltliche Übereinstimmung zwischen den KMK-Standards und den Items des BilWiss-2.0-Tests gibt.

4 Nicht abgedeckt sind im Einzelnen die Standards 1.4, 5.2–5.4, 6.3, 7.7, 8.3, 10.2–10.3. Da der Test nicht zur vollständigen Abdeckung der KMK-Standards entwickelt wurde, wird hier auf diese Themen nicht näher eingegangen.

	Kompetenzbereiche der KMK-Standards				Gesamt
	Unterrichten	Erziehen	Beurteilen	Innovieren	
<b>Inhaltsbereiche BilWiss-2.0-Test</b>					
<i>Unterrichtsgestaltung</i>	16 (66.7%)	3 (12.5%)	5 (20.8%)	0 (0.0%)	<b>24 (100%)</b>
<i>Lernen und Entwicklung</i>	14 (50.0%)	10 (35.7%)	4 (14.3%)	0 (0.0%)	<b>28 (100%)</b>
<i>Diagnostik und Evaluation</i>	1 (3.0%)	0 (0.0%)	28 (84.8%)	4 (12.1%)	<b>33 (100%)</b>
<i>Lehrerberuf</i>	4 (23.5%)	2 (11.8%)	4 (23.5%)	7 (41.2%)	<b>17 (100%)</b>
<i>Schule als Bildungsinstitution</i>	0 (0.0%)	5 (18.5%)	2 (7.4%)	20 (74.1%)	<b>27 (100%)</b>
<i>Bildungstheorie</i>	2 (13.3%)	6 (40.0%)	0 (0.0%)	7 (46.7%)	<b>15 (100%)</b>
<b>Gesamt</b>	<b>37 (25.7%)</b>	<b>26 (18.1%)</b>	<b>43 (29.9%)</b>	<b>38 (26,4%)</b>	<b>144 (100%)</b>

Tab. 3: Absolute Anzahl und prozentuale Verteilung der Item-Standard-Passungen. Die Zeilensummen ergeben die Anzahl der Item-Standard-Passungen pro Inhaltsbereich des BilWiss-2.0-Tests, die Spaltensummen ergeben die Anzahl der Item-Standard-Passungen pro Kompetenzbereich der KMK-Standards. Die Gesamtzahl der Item-Standard-Passungen ist höher als die Gesamtanzahl der 119 Testitems, da einige Items inhaltlich mit mehreren Standards übereinstimmen

## 6. Diskussion

Ziel dieses Beitrags war die Überprüfung der bundesländerübergreifenden Einsetzbarkeit des BilWiss-2.0-Tests zur Messung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften. Im Rahmen von zwei Studien wurden Hinweise auf die curriculare Validität (im Sinne eines intendierten Curriculums) von Interpretationen der BilWiss-2.0-Testwerte hinsichtlich des bildungswissenschaftlichen Wissensstands gewonnen.

In Studie 1 standen die beiden Fragen im Fokus, wie einig sich Vertretungen der Länder der Bundesrepublik Deutschland in ihren Relevanzeinschätzungen der dem BilWiss-2.0-Test zugrunde liegenden bildungswissenschaftlichen Themen sind und wie relevant die in diesem Wissenstest behandelten Themen für die universitäre Lehrerbildung in den einzelnen Bundesländern eingeschätzt werden. Zur Beantwortung dieser Fragen wurde eine Befragung von Expertinnen und Experten aus Schulministerien von 13 Bundesländern durchgeführt.<sup>5</sup> Die Ergebnisse von Studie 1 weisen darauf hin, dass es zwischen den Landesvertretungen eine hohe Übereinstimmung der Relevanzbewertungen in fünf der sechs Inhaltsbereiche gibt. Lediglich die Relevanz der Themen des Inhaltsbereichs *Bildungstheorie* wird kontrovers gesehen. Die über die Landesvertretungen gemittelten Einschätzungen liegen vorwiegend im Skalenbereich, der

5 Wir bedanken uns an dieser Stelle herzlich bei allen ministerialen Vertreterinnen und Vertretern der 13 Länder für ihre Teilnahme an der Expertenstudie.

als relevant interpretiert werden kann. Auch die Bewertungen der einzelnen Bundesländer bewegen sich vorwiegend im positiv relevanten Bereich. Lediglich für ein Bundesland fallen die Relevanzeinschätzungen geringer aus, da hier ein Teil der Themen als vorwiegend relevant für die zweite Phase der Lehrerbildung gesehen wurde. Da nicht auszuschließen ist, dass die Einschätzungen der Expertinnen und Experten trotz eindeutiger Instruktion von subjektiven Bewertungstendenzen beeinflusst waren, ist eine solch starke Abweichung einer einzelnen Landesvertretung in ihrer Gesamtbedeutung zu relativieren. Insgesamt sprechen die Ergebnisse von Studie 1 für eine landesübergreifende Gültigkeit der bildungswissenschaftlichen Themen, die dem Wissenstest zugrunde liegen.

Zu diesen Befunden passen auch die Ergebnisse der Studie 2, in der die Frage im Fokus stand, inwieweit die Testinhalte auf der Ebene der einzelnen Items den bundesweit geltenden KMK-Standards für die Bildungswissenschaften in der universitären Lehrerbildung (2014) entsprechen. Die Ergebnisse der Dokumentenanalyse zeigen, dass 96.6%, also fast alle Testitems, inhaltlich mit mindestens einem KMK-Standard übereinstimmen und dass auch umgekehrt der Großteil der KMK-Kompetenzen durch Items im BilWiss-2.0-Test abgedeckt ist. Dies weist auf eine hohe Entsprechung der Testinhalte und den bundesweiten Standards über bedeutsame bildungswissenschaftliche Inhalte der universitären Lehrerbildung hin. Die Item-Standard-Passungen verteilen sich annähernd gleichmäßig auf alle vier KMK-Bereiche. Für Untersuchungen, die speziell Kompetenzentwicklungen im Sinne der KMK-Standards prüfen wollen, ist der BilWiss-2.0-Test theoretisch einsetzbar, jedoch ist zu beachten, dass nicht alle KMK-Kompetenzen von den Testinhalten abgedeckt werden (siehe Fußnote 3). Für eine vollständige Überprüfung der den KMK-Standards entsprechenden Kompetenzen stellt der Essener SBW-Test eine Alternative dar, dessen Entwicklung auf den KMK-Standards basiert (vgl. Müser, Fleischer & Leutner, 2018), allerdings zunächst spezifisch für die Untersuchung der Kompetenzentwicklung an der Universität Duisburg-Essen konzipiert wurde. Eine Übertragbarkeit auf andere Kontexte wäre hier noch zu prüfen.

Die Ergebnisse beider Studien liefern erste Hinweise auf die curriculare Validität der Testwertinterpretationen auf einer vergleichsweise distalen Ebene des intendierten Curriculums. Die berichteten Studien erlauben jedoch keine Aussagen bezüglich der implementierten und realisierten Curricula, da in Bezug auf die Lehrangebote in den Bildungswissenschaften und deren Nutzung seitens der Studierenden zwischen und innerhalb der lehrerbildenden Hochschulen große Unterschiede bestehen. Eine weitere Limitation der Studie besteht darin, dass pro Bundesland nur eine oder zwei Personen befragt wurden und dass die Vertretungen zweier Bundesländer in Studie 1 nicht vertreten waren. Der hohe Konsens der Landesvertretungen mit überwiegend zufriedenstellenden Relevanzeinschätzungen aus Studie 1 sowie die in Studie 2 ermittelte große inhaltliche Übereinstimmung der Testitems mit bundesweiten bildungswissenschaftlichen Standards geben allerdings keinen Anlass zur Annahme einer entscheidenden bundeslandspezifischen Abweichung. Als eine weitere Limitation wurden beim Abgleich der Testitems mit den KMK-Standards nur die Kompetenzen für die theoretischen Ausbildungsabschnitte betrachtet und die praktischen Ausbildungsabschnitte ausgeblendet.

Zudem bleibt die Frage der Situationspezifität des bildungswissenschaftlichen Wissens offen.

Die beiden hier vorgestellten Studien konnten hilfreiche Hinweise zur Beantwortung der Fragestellungen nach der länderübergreifenden Anwendbarkeit des Tests liefern. Die Ergebnisse beider Studien sprechen unter Berücksichtigung der genannten Limitationen im Großen und Ganzen für eine bundesweite Repräsentation relevanter bildungswissenschaftlicher Inhalte und damit für einen länderübergreifenden Einsatz des Wissenstests. Nichtsdestotrotz wird empfohlen, bei einem Testeinsatz sorgfältig zu prüfen, inwieweit die mit dem Test erfassbaren Inhaltsbereiche bildungswissenschaftlichen Wissens eine geeignete Operationalisierung des interessierenden Konstrukts darstellen, damit die Interpretation der Testwerte inhaltlich valide ausfallen kann.

## Literatur

- AERA = American Educational Research Organization, APA = American Psychological Association & NCME = National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing* (2. Ed.). Washington, DC: Authors.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687–699.
- Burke, M. J., & Dunlap, W. P. (2002). Estimating Interrater Agreement with the Average Deviation Index: A user's guide. *Organizational Research Methods*, 5(2), 159–172.
- Dicke, T., Parker, P. D., Holzberger, D., Kunina-Habenicht, O., Kunter, M., & Leutner, D. (2015). Beginning Teachers' Efficacy And Emotional Exhaustion: Latent changes, reciprocity, and the influence of professional knowledge. *Contemporary Educational Psychology*, 41, 62–72.
- Hohenstein, F., Zimmermann, F., Kleickmann, T., Köller, O., & Möller, J. (2014). Sind die bildungswissenschaftlichen Standards für die Lehramtsausbildung in den Curricula der Hochschulen angekommen? *Zeitschrift für Erziehungswissenschaft*, 17(3), 497–507.
- 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.
- KMK = Kultusministerkonferenz (2014). *Standards für die Lehrerbildung: Bildungswissenschaften* (Beschluss der Kultusministerkonferenz vom 16. 12. 2004 i. d. F. vom 12. 06. 2014). Berlin: KMK.
- Kunina-Habenicht, O., Lohse-Bossenz, H., Kunter, M., Dicke, T., Förster, D., Göbbling, J., Schulze-Stocker, F., Schmeck, A., Baumert, J., Leutner, D., & Terhart, E. (2012). Welche bildungswissenschaftlichen Inhalte sind wichtig in der Lehrerbildung? Ergebnisse einer Delphi-Studie. *Zeitschrift für Erziehungswissenschaft*, 15(4), 649–682.
- Kunina-Habenicht, O., Schulze-Stocker, F., Kunter, M., Baumert, J., Leutner, D., Förster, D., Lohse-Bossenz, H., & Terhart, E. (2013). Die Bedeutung der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens. *Zeitschrift für Pädagogik*, 59(1), 1–23.
- Kunina-Habenicht, O., Maurer, C., Wolf, K., Holzberger, D., Schmidt, M., Dicke, T., Teuber, Z., Koc-Januchta, M., Lohse-Bossenz, H., Leutner, D., Seidel, T., & Kunter, M. (2019). *Der BilWiss-2.0-Test: Ein revidierter Test zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften*. Eingereichtes Manuskript, derzeit im Peer-Review-Verfahren einer Fachzeitschrift.

- Kunter, M., Kunina-Habenicht, O., Baumert, J., Dicke, T., Holzberger, D., Lohse-Bossenz, H., Leutner, D., Schulze-Stocker, F., & Terhart, E. (2017). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung: Ergebnisse des Projekts BilWiss. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 37–54). Wiesbaden: Springer-Online.
- Lenske, G., Thillmann, H., Wirth, J., Dicke, T., & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests. *Zeitschrift für Erziehungswissenschaft, 18*(2), 1–21.
- Linninger, C., Kunina-Habenicht, O., Emmenlauer, S., Dicke, T., Schulze-Stocker, F., Leutner, D., Seidel, T., Terhart, E., & Kunter, M. (2015). Assessing Teachers' Educational Knowledge: Construct specification and validation using mixed methods. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47*(2), 72–83.
- Lohse-Bossenz, H., Kunina-Habenicht, O., Dicke, T., Leutner, D., & Kunter, M. (2015). Teachers' Knowledge About Psychology: Development and validation of a test measuring theoretical foundations for teaching and its relation to instructional behavior. *Studies in Educational Evaluation, 44*, 36–49.
- Lohse-Bossenz, H., Kunina-Habenicht, O., & Kunter, M. (2014). Estimating Within-group Agreement in Small Groups: A proposed adjustment for the average deviation index. *European Journal of Work and Organizational Psychology, 23*(3), 456–468.
- Lohmann, V., Seidel, V., & Terhart, E. (2011). Bildungswissenschaften in der universitären Lehrerbildung: Curriculare Strukturen und Verbindlichkeiten. Eine Analyse aktueller Studienordnungen an nordrhein-westfälischen Universitäten. *Lehrerbildung auf dem Prüfstand, 4*(2), 271–302.
- Merzyn, G. (2002). *Stimmen zur Lehrerausbildung: Ein Überblick über die Diskussion zur Gymnasiallehrausbildung, basierend vor allem auf Stellungnahmen von Wissenschafts- und Bildungsgremien sowie auf Erfahrungen von Referendaren und Lehrern*. Baltmannsweiler: Schneider Hohengehren.
- Müser, S., Fleischer, J., & Leutner, D. (2018). Die KMK-Standards in der Lehrerbildung. Konzeption und Validierung eines Messinstruments (Vortrag auf der *Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung AEPF*, Lüneburg, 24.–26.09.2018).
- Schulze-Stocker, F. (2016). *Die Normierung der Bildungswissenschaften in den Lehramtsstudiengängen durch gesetzliche Vorgaben: Wie (re)organisieren Universitäten ihr bildungswissenschaftliches Lehrangebot? Eine empirische Untersuchung von Studienprogrammen an Universitäten in Nordrhein-Westfalen*. Münster: Monsenstein & Vannerdat.
- Schulze-Stocker, F., Holzberger, D., Kunina-Habenicht, O., Terhart, E., & Kunter, M. (2016). Spielen Studienschwerpunkte wirklich eine Rolle? Zum Zusammenhang von bildungswissenschaftlichen Studienschwerpunkten, selbst eingeschätzten Kenntnissen und gemessenem Wissen am Ende eines Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft, 19*(3), 559–623.
- Schulze-Stocker, F., Holzberger, D., & Lohse-Bossenz, H. (2017). Das bildungswissenschaftliche Curriculum – Zentrale Ergebnisse des BilWiss-Programms. *Das Hochschulwesen, 65*(4+5), 134–138.
- Shulman, L. S. (1986). Those Who Understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–21.
- Seifert, A., Hilligus, A. H., & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstruments zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand, 2*(1), 82–103.
- Terhart, E. (2014). Standards für die Lehrerbildung: Bildungswissenschaften – nach zehn Jahren. *Die Deutsche Schule, 106*(4), 300–323.

- UNESCO International Bureau of Education (2013). *Glossary of Curriculum Terminology*. Genf: UNESCO IBE. <http://www.ibe.unesco.org/sites/default/files/resources/ibe-glossary-curriculum.pdf> [17.04.2019].
- Veenman, S. (1984). Perceived Problems of Beginning Teachers. *Review of Educational Research*, 54(2), 143–178.
- 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.
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing Teacher Candidates' General Pedagogical and Psychological Knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969.

**Abstract:** In their studies on educational foundations as a part of pre-service teacher training in Germany, pre-service teachers are supposed to develop the generic knowledge needed for professional behavior in their job as a teacher. The research program BilWiss developed a test on educational knowledge within the context of the federal state of North-Rhine Westphalia (NRW) covering relevant content of educational foundations in pre-service teacher training. In two studies, the present paper investigates whether the test content corresponds to the intended academic teacher training curriculum at federal level. Both state delegates' expert ratings and the comparison of the test content with federal standards on the subject unspecific, educational part of pre-service teacher training indicate that the test can be applied across federal states.

**Keywords:** Knowledge Test, Teacher Education, Educational Foundations, Validity, at Federal Level

### **Anschrift der Autor\_innen**

Jun. Prof. Dr. Olga Kunina-Habenicht, Pädagogische Hochschule Karlsruhe,  
Fakultät für Geistes- und Humanwissenschaften,  
Institut für bildungswissenschaftliche Forschungsmethoden,  
Bismarckstraße 10, 76133 Karlsruhe, Deutschland  
E-Mail: [olga.kunina-habenicht@ph-karlsruhe.de](mailto:olga.kunina-habenicht@ph-karlsruhe.de)

Dr. Christina Maurer, Goethe-Universität Frankfurt am Main,  
Fachbereich 05 Psychologie & Sportwissenschaften, Institut für Psychologie,  
Theodor-W.-Adorno-Platz 6, 60629 Frankfurt a. M., Deutschland  
E-Mail: [maurer@psych.uni-frankfurt.de](mailto:maurer@psych.uni-frankfurt.de)

Dr. Franziska Schulze-Stocker, Technische Universität Dresden,  
Zentrum für Qualitätsanalyse,  
Chemnitzer Str. 48a, 01187 Dresden, Deutschland  
E-Mail: [franziska.schulze-stocker@tu-dresden.de](mailto:franziska.schulze-stocker@tu-dresden.de)

Kristin Wolf, Goethe-Universität Frankfurt am Main,  
Fachbereich 05 Psychologie & Sportwissenschaften, Institut für Psychologie,  
Theodor-W.-Adorno-Platz 6, 60629 Frankfurt a. M., Deutschland  
E-Mail: [wolf@paed.psych.uni-frankfurt.de](mailto:wolf@paed.psych.uni-frankfurt.de)

Nora Hein, Universität Duisburg-Essen,  
Fakultät für Bildungswissenschaften, Lehrstuhl für Lehr-Lernpsychologie,  
Universitätsstr. 2, 45141 Essen, Deutschland  
E-Mail: [nora.hein@uni-due.de](mailto:nora.hein@uni-due.de)

Prof. Dr. Dr. h. c. Detlev Leutner, Universität Duisburg-Essen,  
Fakultät für Bildungswissenschaften, Lehrstuhl für Lehr-Lernpsychologie,  
Universitätsstr. 2, 45141 Essen, Deutschland  
E-Mail: [detlev.leutner@uni-due.de](mailto:detlev.leutner@uni-due.de)

Prof. Dr. Tina Seidel, Technische Universität München (TUM),  
TUM School of Education, Friedl Schöller-Stiftungslehrstuhl für Pädagogische Psychologie,  
Arcisstraße 21, 80333 München, Deutschland  
E-Mail: [tina.seidel@tum.de](mailto:tina.seidel@tum.de)

Prof. Dr. Mareike Kunter, Goethe-Universität Frankfurt am Main,  
Fachbereich 05 Psychologie & Sportwissenschaften, Institut für Psychologie,  
Theodor-W.-Adorno-Platz 6, 60629 Frankfurt a. M., Deutschland  
E-Mail: [kunter@paed.psych.uni-frankfurt.de](mailto:kunter@paed.psych.uni-frankfurt.de)