

Riese, Josef; Kulgemeyer, Christoph; Zander, Simon; Borowski, Andreas; Fischer, Hans E.; Gramzow, Yvonne; Reinhold, Peter; Schecker, Horst; Tomczyszyn, Elisabeth

Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik

Blömeke, Sigrid [Hrsg.]; Zlatkin-Troitschanskaia, Olga [Hrsg.]: *Kompetenzen von Studierenden*. Weinheim u.a. : Beltz Juventa 2015, S. 55-79. - (Zeitschrift für Pädagogik, Beiheft; 61)



Quellenangabe/ Reference:

Riese, Josef; Kulgemeyer, Christoph; Zander, Simon; Borowski, Andreas; Fischer, Hans E.; Gramzow, Yvonne; Reinhold, Peter; Schecker, Horst; Tomczyszyn, Elisabeth: Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik - In: Blömeke, Sigrid [Hrsg.]; Zlatkin-Troitschanskaia, Olga [Hrsg.]: *Kompetenzen von Studierenden*. Weinheim u.a. : Beltz Juventa 2015, S. 55-79 - URN: urn:nbn:de:0111-pedocs-155037 - DOI: 10.25656/01:15503

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-155037>

<https://doi.org/10.25656/01:15503>

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

61. Beiheft

April 2015

ZEITSCHRIFT FÜR PÄDAGOGIK

**Kompetenzen
von Studierenden**

BELTZ JUVENTA

Zeitschrift für Pädagogik · 61. Beiheft

Zeitschrift für Pädagogik · 61. Beiheft

Kompetenzen von Studierenden

Herausgegeben von

Sigrid Blömeke und Olga Zlatkin-Troitschanskaia

BELTZ JUVENTA

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, bleiben dem Beltz-Verlag vorbehalten.

Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2015 Beltz Juventa · Weinheim und Basel

www.beltz.de · www.juventa.de

Herstellung: Lore Amann

Satz: text plus form, Dresden

E-Book

ISSN 0514-2717

Bestell-Nr. 443508

Inhaltsverzeichnis

<i>Sigrid Blömeke/Olga Zlatkin-Troitschanskaia</i> Kompetenzen von Studierenden. Einleitung zum Beiheft	7
--	---

<i>Lars Jenßen/Simone Dunekacke/Sigrid Blömeke</i> Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis	11
---	----

Berufsbezogene Kompetenzen

<i>Svenja Hammer/Sonja A. Carlson/Timo Ehmke/Barbara Koch-Priewe/ Anne Köker/Udo Ohm/Sonja Rosenbrock/Nina Schulze</i> Kompetenz von Lehramtsstudierenden in Deutsch als Zweitsprache: Validierung des GSL-Testinstruments	32
--	----

<i>Josef Riese/Christoph Kulgemeyer/Simon Zander/Andreas Borowski/ Hans E. Fischer/Yvonne Gramzow/Peter Reinhold/Horst Schecker/ Elisabeth Tomczyszyn</i> Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik	55
--	----

<i>Simone Dunekacke/Lars Jenßen/Sigrid Blömeke</i> Mathematikdidaktische Kompetenz von Erzieherinnen und Erziehern: Validierung des KomMa-Leistungstests durch die videogestützte Erhebung von Performanz	80
--	----

<i>Franziska Bouley/Stefanie Berger/Sabine Fritsch/Eveline Wuttke/ Jürgen Seifried/Kathleen Schnick-Vollmer/Bernhard Schmitz</i> Der Einfluss von universitären und außeruniversitären Lerngelegenheiten auf das Fachwissen und fachdidaktische Wissen von angehenden Lehrkräften an kaufmännisch-berufsbildenden Schulen	100
--	-----

<i>Olga Zlatkin-Troitschanskaia/Manuel Förster/Susanne Schmidt/ Sebastian Brückner/Klaus Beck</i> Erwerb wirtschaftswissenschaftlicher Fachkompetenz im Studium – Eine mehrbenenanalytische Betrachtung von hochschulischen und individuellen Einflussfaktoren	116
---	-----

Gabriele Kaiser

Erfassung berufsbezogener Kompetenzen von Studierenden.

Ein Kommentar 136

Forschungsbezogene Kompetenzen

Kati Trempler/Andreas Hetmanek mit Christof Wecker/Jan Kiesewetter/

Mia Wermelt/Frank Fischer/Martin Fischer/Cornelia Gräsel

Nutzung von Evidenz im Bildungsbereich – Validierung

eines Instruments zur Erfassung von Kompetenzen

der Informationsauswahl und Bewertung von Studien 144

Sandra Schladitz/Jana Groß Ophoff/Markus Wirtz

Konstruktvalidierung eines Tests zur Messung

bildungswissenschaftlicher Forschungskompetenz 167

Alexandra Winter-Hözl/Kristin Wäschle/Jörg Wittwer/

Rainer Watermann/Matthias Nückles

Entwicklung und Validierung eines Tests zur Erfassung

des Genrewissens Studierender und Promovierender

der Bildungswissenschaften 185

Gabriele Steuer/Tobias Engelschalk/Gregor Jöstl/Anne Roth/

Bastian Wimmer/Bernhard Schmitz/Barbara Schober/Christiane Spiel/

Albert Ziegler/Markus Dresel

Kompetenzen zum selbstregulierten Lernen im Studium:

Ergebnisse der Befragung von Expert(inn)en aus vier Studienbereichen 203

Johannes König

Stand der Forschung zu wissenschaftsbezogenen Kompetenzen

und weiterführende Fragen. Ein Kommentar 226

Josef Riese/Christoph Kulgemeyer/Simon Zander/Andreas Borowski/
Hans E. Fischer/Yvonne Gramzow/Peter Reinhold/Horst Schecker/
Elisabeth Tomczyszyn

Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik

Zusammenfassung: Im Hinblick auf die Verbesserung der Lehrerbildung besteht ein hohes Interesse daran, die Struktur und Genese der Kompetenzen von Lehramtsstudierenden zu beschreiben. Die Ziele des vorgestellten Projekts liegen in der Modellierung und Messung domänenspezifischer und generischer Kompetenzen, die Lehramtsstudierende der Physik im Hochschulstudium erwerben sollen. Das Modell umfasst differenzierte Facetten des physikalischen Fachwissens und des physikdidaktischen Wissens sowie das fachspezifische Erklärungswissen in unterrichtsnahen Vermittlungssituationen. Es enthält zudem Annahmen über Zusammenhänge dieser Bereiche des Professionswissens. Ausgehend vom Modell wurden schriftliche Testinstrumente und videografierte Lehr-Lern-Rollenspiele entwickelt und bei Physik-Lehramtsstudierenden eingesetzt. Die vorliegenden Ergebnisse sprechen für die Validität der genutzten Testverfahren.

Schlagerworte: Professionswissen, Lehramtsstudierende, Physik, Kompetenzmessung, Validität

1. Einleitung

1.1 Problemstellung

Um die Wirksamkeit der universitären Lehrerbildung abschätzen zu können, müssen die Struktur und die Genese derjenigen Kompetenzen beschrieben werden, die Lehramtsstudierende am Ende ihres Studiums erworben haben sollten. Allerdings sind evidenzbasierte Vorschläge zur Verbesserung des Lehramtsstudiums bisher nicht ausreichend vorhanden. Large-Scale-Studien, die den Standards nationaler und internationaler Vergleichsstudien aus dem Schulbereich entsprechen, liegen in der internationalen Lehrerbildungsforschung nicht vor (vgl. Zlatkin-Troitschanskaia & Kuhn, 2010). Das gilt auch für die naturwissenschaftlichen Fächer. Zwar sind in den letzten Jahren erste Arbeiten zur Kompetenzmodellierung und -messung in der Physik entstanden, die ausgehend von der von Baumert und Kunter (2006) vorgeschlagenen Strukturierung *Fachwissen – fachdidaktisches Wissen – allgemeines pädagogisches Wissen* Kompetenzmessungen bei Lehrkräften oder Lehramtsstudierenden der Physik vorgenommen haben (vgl. z.B. die Arbeiten von Tepner et al., 2012; Riese & Reinhold, 2012; Kröger, Neumann & Petersen, 2013). Die bisherigen Arbeiten können allerdings keine über grundlegende Beschreibungen hinausgehenden Erkenntnisse liefern, da in diesen Untersuchungen hauptsächlich Gesamtscores einzelner Wissensbereiche abgebildet werden,

wie z.B. ein globaler Kennwert für *fachdidaktisches Wissen*. Dementsprechend sind bisher lediglich globale Modellprüfungen möglich, nicht jedoch differenzierte Aussagen zu Teilbereichen einzelner Dimensionen des Professionswissens. Für eine Analyse des Professionswissens von Lehramtsstudierenden ist aber eine differenziertere Modellierung wünschenswert, die z.B. die innere Struktur von *Fachwissen* und *fachdidaktischem Wissen* und die Beziehung zwischen diesen Wissensbereichen abbilden kann, um damit z.B. die Wirkung unterschiedlicher Lerngelegenheiten und längsschnittliche Entwicklungsverläufe untersuchen zu können.

1.2 Ziele der Untersuchung

Vor dem Hintergrund des oben skizzierten Problemfelds leistet das im vorliegenden Beitrag beschriebene Projekt ProfiLe-P („*Professionswissen in der Lehramtsausbildung Physik*“, gefördert vom BMBF – FKZ 01PK11001) einen empirisch fundierten Beitrag zur Diskussion über die Wirkungen universitärer Lehrerbildung. Die Ziele von ProfiLe-P liegen in der Modellierung domänenspezifischer und generischer Wissensbereiche, die Lehramtsstudierende der Physik im Hochschulstudium als Teil ihrer professionellen Kompetenz erwerben sollten. Darüber hinaus werden die theoretischen Modelle in Messmodelle und Testinstrumente überführt, die nach mehreren Schritten der Validierung zentrale Wissensbereiche angehender Physiklehrkräfte differenzierter als bisher überprüfbar machen. Als innovatives Testformat werden praxisnahe Rollenspiele genutzt, um spezifische Anforderungen realen Lehrerhandelns in Laborsituationen abzubilden und damit, unabhängig vom empirisch schwer zugänglichen Untersuchungsfeld Schule, messbar zu machen. Mit den entwickelten Testinstrumenten wird ein Datenpool aufgebaut, der Zusammenhangsanalysen zwischen Wissensbereichen des Professionswissens sowie differenzielle Analysen ausgewählter Kompetenzen nach Lehrämtern und Studiengangstrukturen erlaubt und darüber hinaus auf der Basis von Quasi-Längsschnitt- und echten Längsschnitterhebungen Aufschlüsse über Kompetenzentwicklungsverläufe in ausgewählten Studienabschnitten liefert.

In diesem Beitrag wird über die Kompetenzmodellierung (Projektziel 1, siehe Abschnitt 3), die Testentwicklung und die Methodik (Projektziel 2, siehe Abschnitt 4) und die Validität berichtet (Projektziel 3, siehe Abschnitt 5). Die Erhebungen zum Aufbau des gemeinsamen Datenpools (Projektziel 4) sind zurzeit (Stand August 2014) noch nicht abgeschlossen, sodass Zusammenhangsanalysen zwischen den Wissensbereichen noch ausstehen. Im Folgenden wird zunächst der theoretische Hintergrund der Kompetenzmodellierung beschrieben.

2. Theoretischer Hintergrund

2.1 Das professionelle Wissen von Lehrkräften

Der Begriff *Professionswissen* bezeichnet in der Regel spezifisches Wissen, über das die Angehörigen einer bestimmten Profession verfügen sollten (vgl. Stichweh, 1996). In Bezug auf Lehrkräfte ist dabei jenes Wissen gemeint, das den Lehrberuf auszeichnet und das als notwendig angesehen wird, die Anforderungen dieses Berufs zu bewältigen. Dem Modell von Baumert und Kunter (2006) folgend lässt es sich auf erster Strukturierungsebene in fachliches Wissen, fachdidaktisches Wissen und nicht fachbezogenes pädagogisches Wissen unterteilen.

Fachliches Wissen (FW) bezieht sich hierbei sowohl auf Wissen über die Inhalte und Methoden des jeweiligen Fachs als auch auf epistemologisches Wissen über das Fach, in dem eine Lehrperson unterrichtet. Die historisch generierte inhaltliche und logische Struktur der Fachdisziplin (insbesondere die üblicherweise unterschiedenen Inhaltsgebiete) wird dabei in zentralen Lehrwerken und in Curricula von Studienprogrammen abgebildet. Obwohl FW sowohl eine Grundvoraussetzung für erfolgreichen Fachunterricht darstellt als auch den Erwerb von fachdidaktischem Wissen beeinflusst (vgl. z. B. Baumert & Kunter, 2006), fand es in der Untersuchung des Professionswissens von Lehramtsstudierenden oder Lehrkräften bisher wenig Berücksichtigung (vgl. Fischer, Borowski & Tepner, 2012).

Fachdidaktisches Wissen (FDW) wiederum bezeichnet einen für den Lehrberuf spezifischen Wissensbereich, der häufig in Anlehnung an das Wissenskonzept des *pedagogical content knowledge* (PCK) von Shulman (1986) definiert wird. Danach benötigen Lehrkräfte spezifisches Wissen, das sie dazu befähigt, fachliche Inhalte adressatenbezogen zu strukturieren und darzustellen. In den letzten Jahren wurden zahlreiche Modellierungen vorgenommen, in denen das ursprüngliche Konstrukt unterschiedlich ausdifferenziert wurde, allerdings existiert für FDW in den Naturwissenschaften bislang weder national noch international eine einheitliche, umfassende und auf ein Rahmenmodell zum Professionswissen abgestimmte Modellierung (vgl. Abell, 2007; Park & Chen, 2012; Fischer et al., 2012). Die meisten Untersuchungen betrachten jedoch Wissen über Schülerkonzeptionen bzw. -vorstellungen und Wissen über Lehrstrategien und Darstellungsformen als zentrale Elemente des FDW (vgl. Gramzow, Riese & Reinhold, 2013). Kulgemeyer und Schecker (2013) sprechen darüber hinaus von Erklärungswissen (EW), das Lehrpersonen für eine sachgerechte und schülergemäße Kommunikation naturwissenschaftlicher Sachverhalte benötigen. EW weist demnach starke Bezüge zum FDW und zum FW auf, wird aber von den Autoren als eigenständiger Wissensbereich betrachtet, da es sich ausschließlich auf Anforderungen in einer konkreten, unterrichtsbezogenen Handlungssituation bezieht.

Pädagogisches Wissen (PW) bezeichnet schließlich fachunabhängiges Wissen über pädagogische und allgemeindidaktische Konzepte und Inhalte. Da wir in der im Beitrag vorgestellten Untersuchung auf physikbezogenes Professionswissen fokussieren, wird PW im Folgenden nicht weiter ausgeführt.

2.2 Untersuchungen zum fachbezogenen Professionswissen von Lehrkräften

In der nationalen Literatur herrscht weitgehend Konsens im Hinblick auf die oben beschriebene Grobstruktur des Professionswissens von (angehenden) Lehrkräften. Dennoch wurden bisher häufig nur einzelne Bereiche unabhängig voneinander untersucht (Fischer et al., 2012). Auch für einzelne Fächer liegen kaum Erkenntnisse zur Struktur des fachbezogenen Professionswissens vor, die mehr als globale Zusammenhänge (wie z. B. bei Riese & Reinhold, 2012) zwischen den drei Bereichen aufklären. Allerdings ist zu vermuten, dass einige Aspekte des FDW (z. B. Wissen über typische fachliche Schüler(fehl)konzepte) ein tieferes Verständnis physikalischer Aspekte erfordern als andere (z. B. Wissen über fachdidaktische Unterrichtsmodelle), sodass globale Korrelationen die Zusammenhänge nur oberflächlich aufklären. Mangels empirisch bislang nicht trennbarer Unterfacetten des FDW können vorliegende Studien zum Zusammenhang auf Teilskalenebene jedoch keine Aussage machen (vgl. Gramzow et al., 2013). Weiter ist festzustellen, dass das Professionswissen von (angehenden) Lehrkräften bislang fast ausschließlich mit schriftlichen Leistungstests erfasst wurde, wobei unklar ist, inwieweit die erfassten Wissens Elemente die Qualität der Handlung von Lehrpersonen valide prognostizieren können (Vogelsang & Reinhold, 2013).

2.3 Konzeptionen des Professionswissens in den Naturwissenschaften

Im Vergleich zu Modellen naturwissenschaftlicher Kompetenz auf Schülerebene (wie z. B. bei Pant et al., 2013) müssen Modelle des Professionswissens von Lehrkräften deutlich umfangreicher angelegt sein, da mehr Wissensbereiche einzubeziehen sind. Im Folgenden werden typische Modellierungsansätze zum FW und FDW skizziert, die in bisherigen Studien als Ausgangsbasis für Itementwicklungen genutzt wurden.

Modelle physikalischen Fachwissens

Empirische Belege für eine bestimmte inhaltliche Ausprägung oder Dimensionierung physikalischen Fachwissens oder für die Anzahl der Facetten in einzelnen Dimensionen liegen bisher nicht vor. Häufig finden sich jedoch mehrdimensionale konzeptionelle Modelle. Üblicherweise beschreiben Autoren dabei zum einen eine inhaltliche Dimension, welche typische Themen der Physik unterscheidet (z. B. Kröger et al., 2013; Riese & Reinhold, 2012). Zum anderen werden Fachstufen unterschieden, die sich auf typische Lerngelegenheiten des Wissenserwerbs beziehen und die in den verschiedenen Untersuchungen teilweise Schwierigkeit generiert haben. Im Bereich Primarstufe unterscheidet Ohle (2010) zur Konstruktion eines Testinstruments für das Fachwissen z. B. die Fachstufen *Grundschule*, *Mittelstufe* sowie *Universität*. Daneben findet sich häufig eine dritte Teildimension, die es ermöglichen soll, Testitems a priori nach ihrer Aufgabenschwierigkeit anzuordnen und Unterschiede im Wissen stärker auf die Struktur des Wissens selbst zu beziehen. Beispielsweise unterscheiden Tepner et al. (2012) zwischen *Wissensarten* (deklaratives, prozedurales und konditionales Wissen) und Woitkowski,

Riese & Reinhold (2011) zwischen unterschiedlichen *Komplexitäten* (Fakten, Prozessbeschreibungen, lineare Kausalität, multivariate Interdependenz).

Modelle fachdidaktischen Wissens

Ebenso wie beim FW liegen in der Literatur zum FDW konzeptionelle Modelle unterschiedlicher Dimensionierung vor. Alle Modelle unterscheiden explizit fachdidaktische Inhalte, welche bezeichnet werden: als *Inhalte* (z. B. Schülerkognition, Instruktionsstrategien, Curriculum, Assessment; bei Kröger et al., 2013), als *Facetten* (Schülervorstellungen, Modelle/Konzepte, Experimente; bei Tepner et al., 2012) oder als *Kategorien* (Wissen über Schülervorstellungen, über das Curriculum und über Schwierigkeiten; vgl. Olszewski, 2010). Darüber hinaus unterscheiden einige Studien analog zu den entsprechenden Modellen des FW auch für FDW Wissensarten und Fachinhalte, um eine strukturelle Einheitlichkeit zu betonen (z. B. Tepner et al., 2012 oder Kröger et al., 2013; siehe oben).

3. Modellierung des Professionswissens von Lehramtsstudierenden der Physik

Das im Forschungsverbund ProfiLe-P entwickelte Rahmenmodell des Professionswissens angehender Physiklehrkräfte unterscheidet auf erster Strukturierungsebene zwischen physikalischem Fachwissen (FW), eher deklarativen und analytischen Aspekten fachdidaktischen Wissens (FDW) und eher prozeduralen Wissensaspekten, welche beim Erklären physikalischer Sachverhalte im Physikunterricht zum Tragen kommen. Von diesem fachspezifischen Erklärungswissen (EW) wird hypothetisch erwartet, dass es Auswirkungen auf das Erklärerhandeln in realen Vermittlungssituationen hat. Der Fokus wurde auf das Erklären physikalischer Sachverhalte gelegt, da es sich beim Erklären um eine wichtige Standardoperation im naturwissenschaftlichen Unterricht handelt (Geelan, 2012). Zudem stellt es Bezüge zwischen deklarativen und prozeduralen Wissensaspekten her. So kann beispielsweise untersucht werden, inwiefern deklarative Wissensaspekte, z. B. Wissen über typische Verständnisprobleme bei Lernenden, in einer unterrichtlichen Handlungssituation relevant werden (vgl. auch Vogelsang & Reinhold, 2013). Abbildung 1 setzt die im Vorhaben untersuchten Bereiche des Professionswissens (mit ihren jeweiligen Facetten) in Beziehung.

In allen drei Wissensbereichen konzentrieren wir uns entsprechend der Leitlinie des BMBF-Rahmenprogramms „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“ (vgl. Blömeke & Zlatkin-Troitschanskaia, 2013) auf universitär erwerbbares physikbezogenes Professionswissen. Als normativ-präskriptive Orientierungspunkte bei der Entwicklung der jeweiligen Kompetenzmodelle wurden u. a. die KMK-Standards für die Lehrerbildung (KMK, 2008; Fachprofil Physik), das Kerncurriculum der Gesellschaft für Fachdidaktik (GFD, 2004), der Fachdidaktikteil der Studie der Deutschen Physikalischen Gesellschaft zur Ausbildung für das Lehramt Physik (DPG, 2014), aber auch die Fachphysik- und Physikdidaktikcurricula der Universitä-

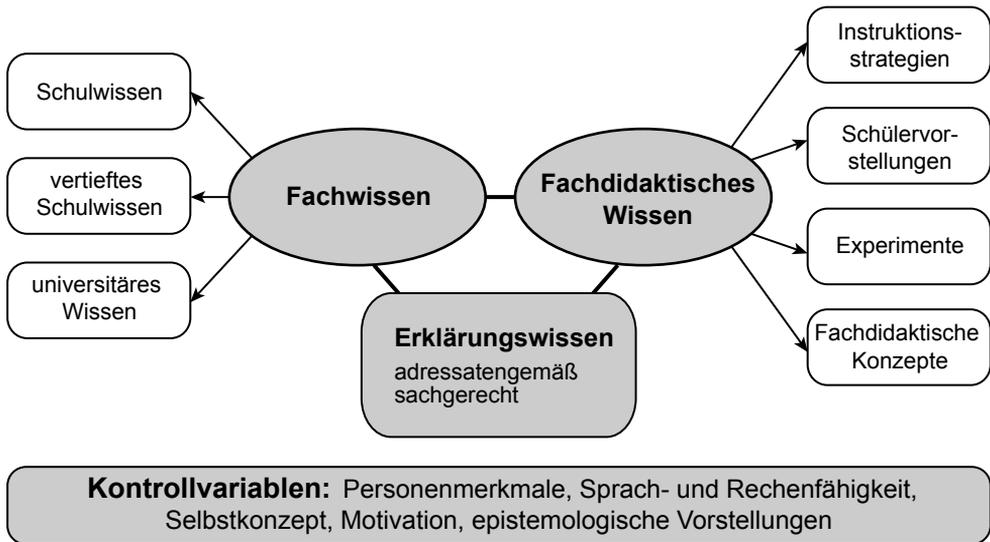


Abb. 1: Mögliche Zusammenhänge zwischen unterschiedlichen Bereichen des Professionswissens

ten des Verbunds sowie Erkenntnisse zum Eingangswissen von Studierenden (Riese & Reinhold, 2012) herangezogen.

Inhaltlich beziehen sich alle Modellierungen auf den Bereich Mechanik, da die Erhebungen mit dem ersten Studiensemester beginnen, in dem im Wesentlichen Mechanik gelehrt wird. Zudem sind Schülvorstellungen und Schülerlernprozesse in der Mechanik gut erforscht (z. B. Schecker, 1985) und daher für die Entwicklung von Items zum FDW und zum EW von Studierenden gut geeignet. Schließlich korreliert FDW in Mechanik nach Borowski & Riese (2010) hoch mit FDW in Elektrizitätslehre.

Im Folgenden werden die Binnenstrukturen der drei Modelle vorgestellt.

3.1 Fachwissen

Das Modell für das Fachwissen unterscheidet die Fachstufen *Schulwissen*, *vertieftes Schulwissen* und *universitäres Wissen* (Abb. 1, vgl. auch z. B. COACTIV-Studie). *Schulwissen* umfasst Inhalte der Richtlinien und Lehrpläne der an diesem Projekt teilnehmenden Bundesländer, mit Beschränkung auf den Grundkurs der Oberstufen. Das *vertiefte Schulwissen* modelliert folgende Fähigkeiten: (1) verschiedene Wege zur Lösung einer Aufgabe identifizieren und anwenden, (2) Lösung einer Aufgabe aus theoretischer Sicht planen, (3) Randbedingungen einer Schulaufgabe erkennen, (4) Aufgaben zielgruppengerecht vereinfachen, (5) Zusammenhänge, Gemeinsamkeiten und Unterschiede physikalischer Phänomene erkennen. *Universitäres Wissen* enthält Inhalte, die an Universitäten für Abschlussprüfungen im Fach vorausgesetzt werden. Bei diesen drei

Fachstufen handelt es sich um qualitativ unterschiedliche Bereiche, von denen erwartet wird, dass sie empirisch zu trennen sind und über mittlere Korrelationen zusammenhängen. Um zu gewährleisten, dass Aufgaben unterschiedlicher Schwierigkeit erstellt werden, wurde die Komplexität (Schoppmeier, Borowski & Fischer, 2013) als weitere Teildimension bei der Entwicklung berücksichtigt, wobei zwischen *Fakten benennen*, *Zusammenhänge zwischen mehreren Fakten herstellen* und dem *Umgang mit übergeordneten physikalischen Konzepten* unterschieden wird. Bei Aufgaben zu übergeordneten physikalischen Konzepten muss ein grundlegendes physikalisches Prinzip (z. B. Basiskonzept Energie) verstanden, in einer gegebenen Situation erkannt oder auf neue Situationen angewandt werden.

3.2 Fachdidaktisches Wissen

Ausgehend von einem fachdidaktischen Modell, das unter Nutzung allgemeiner Standards und normativer Leitbilder der Physiklehrerbildung (z. B. DPG, 2014; Korneck, Lamprecht, Wodzinski & Schecker, 2010) entwickelt wurde und den gesamten Bereich des FDW umspannt (detailliert in Gramzow et al., 2013), wurden vier Facetten für ein separates Modell zur Itementwicklung ausgewählt, um den Testumfang auf ein praktikables Maß zu begrenzen. Hierbei handelt es sich um die fachdidaktischen *Facetten Instruktionsstrategien, Schülervorstellungen, Experimente und Vermittlung eines angemessenen Wissenschaftsverständnisses* sowie *Fachdidaktische Konzepte* (vgl. Abb. 1), die im Sinne einer empirisch zu prüfenden Binnenstruktur des FDW zu verstehen sind. Es wird erwartet, dass diese Facetten für das Erklärungswissen (EW) besonders relevant sind (siehe auch 3.3). Als weitere Teildimension wurden *Kognitive Anforderungen* in den Stufen „Reproduzieren“, „Analysieren“ und „Anwenden“ modelliert (vgl. Anderson & Krathwohl, 2001), um das Spektrum der Anforderungen breit erfassen zu können.

3.3 Erklärungswissen

Das Modell des Erklärungswissens bezieht sich auf unterrichtsnahe, dialogische Erklärungssituationen von Physik (in Anlehnung an Kulgemeyer & Schecker, 2013). Es berücksichtigt als grundlegende Anforderungen die Adressatengerechtheit und die Sachgemäßheit einer Erklärung. Dabei werden folgende Variablen unterschieden: Mathematisierungsgrad, gewählte Beispiele bzw. Analogien, Sprachniveau und Darstellungsformen. Diese Variablen sind nicht im Sinne einer empirisch zu prüfenden Binnenstruktur des EW zu verstehen, sondern als Gestaltungsmittel von Erklärungen, die für die Analyse der Güte von Erklärungen verwendet werden.

In dem grundlegenden Modell wird davon ausgegangen, dass eine Erklärung ein sprachlicher Akt ist und *Erklären* den Prozess der adressatengemäßen und sachgerechten Erstellung und Modifikation von *Erklärungen* bedeutet. Der Prozess läuft im ein-

fachsten Falle wie folgt ab: Es wird eine initiale erklärende Darstellung des Sachverhalts gewählt, die nach Verständnismrückmeldung (verbal oder nonverbal) durch den Adressaten entsprechend modifiziert werden muss. Dabei sind es gerade die in 3.2 genannten vier Facetten des FDW, die eine adressatengemäße und sachgerechte Erstellung bzw. Modifikation der Variablen ermöglichen. So kann beispielsweise das Sprachniveau von der Fach- zur Alltagssprache wechseln oder die Darstellungsform von einem logischen Diagramm zu einem realen Foto. Analogien bzw. Beispiele können sich auf unterschiedliche Gegenstandsbereiche beziehen, z. B. auf fachnahe oder alltagsnahe. Der Mathematisierungsgrad kann wechseln, indem physikalische Größengleichungen expliziert oder verbal umschrieben werden. Gütekriterien für Erklärungen werden z. B. von Brown (2006) und Wellenreuther (2005) beschrieben, etwa „Evaluation des Verständnisses“. Das Modell ist die Adaption eines Modells, das empirisch erfolgreich verwendet wird, um physikbezogene Kommunikation auf Schülerebene zu beschreiben (detailliert in Kulgemeyer & Schecker, 2013).

4. Testentwicklung und Methodik

4.1 Messinstrumente

Auf der Basis der drei oben skizzierten Teilmodelle wurden sowohl schriftliche Leistungstests als auch ein qualitatives, videobasiertes Instrument entwickelt, mit dem die Ausprägung des Erklärungswissens (EW) gemessen wird. Die Inhalte der Tests zur Erhebung von FW und FDW sind dabei auf die Thematiken des Tests zum EW abgestimmt. Alle Tests adressieren den Fachinhalt Mechanik (vgl. Abschnitt 3). Dabei stand bei allen Tests das Ziel im Vordergrund, Strukturen der einzelnen Konstrukte und Zusammenhänge zu Bedingungsfaktoren (z. B. Lerngelegenheiten) untersuchen zu können, nicht aber die Ermöglichung von Individualdiagnostik.

Fachwissen

Für den FW-Test wurden, ausgehend vom in 3.1 beschriebenen Modell, insgesamt 143 geschlossene Aufgaben im Multiple-Choice-Format (Single Select) entwickelt (Bsp. siehe Abb. 2), sodass für die finale Testzusammenstellung noch ausreichend Kürzungsspielraum vorhanden war. Dabei wurden Items zu allen Subdimensionen des Modells erstellt, um das Konstrukt bestmöglich abzudecken.

Der Test wurde in einer ersten Feldstudie im Multimatrixdesign mit 395 Studierenden des Lehramts Physik ($N = 120$), des Physik-Monobachelors ($N = 213$; Vergleichsgruppe) und sonstigen Physikstudierenden ($N = 62$; z. B. Mathematikstudierende mit Physik als Nebenfach) eingesetzt (EAP/PV-Reliabilität .76; Varianz 0.93; bei 127 der 143 Aufgaben gilt $0.8 < \text{gewichteter MNSQ} < 1.2$ und $-1.9 < T < 1.9$). Auf dieser Datengrundlage wurden 40 Aufgaben für den weiteren Einsatz ausgewählt, die in der Pilotierung möglichst gute Kennwerte aufwiesen als auch einen möglichst großen Ausschnitt der Fähigkeitsskala abdeckten, um zu gewährleisten, dass der finale Test während des

Betrachten Sie ein Pendel (z. B. ein Federpendel). Damit dieses Pendel harmonisch schwingt, muss die zurücktreibende Kraft einer ganz bestimmten Gesetzmäßigkeit gehorchen.

Im Folgenden bezeichnen x die Auslenkung des Pendels, \dot{x} die Geschwindigkeit des Pendels, k eine Konstante mit $k < 0$ und F die zurücktreibende Kraft.

Welche Aussage ist richtig?

Es muss gelten:

- 1) $F = k$
- 2) $F = k \cdot x$
- 3) $F = \frac{1}{2} k \cdot x^2$
- 4) $F = k \cdot \dot{x}$

Abb. 2: Beispielitem FW zum Schulwissen, einen Fakt benennen (Lösung: 2)

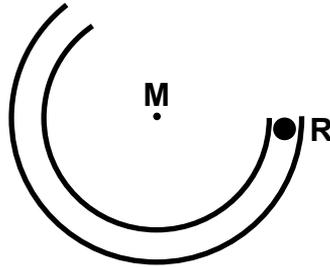
kompletten Studiums eingesetzt werden kann. Weitere Kriterien für die Auswahl waren, dass zentrale für die Mechanik relevante Themenbereiche (z. B. Energie/Arbeit/Leistung) erfasst werden und die Inhalte das notwendige FW zur Beantwortung der fachdidaktischen Aufgaben sowie der Erklärungssituationen abdecken. Zusätzlich wurde berücksichtigt, dass die drei Fachstufen der theoretisch angenommenen inneren Struktur des Fachwissens (vgl. Abschnitt 3.1) getestet werden können (Schulwissen: $N = 17$ Items; universitäres Wissen: $N = 12$ Items; vertieftes Schulwissen: $N = 11$ Items).

Insgesamt haben die Studierenden für die Bearbeitung des finalen Tests 60 Minuten Zeit, wobei der Rechenaufwand zur Auswahl der richtigen Lösungen ohne Taschenrechner zu leisten ist. Zur Erfassung des Einflusses der mathematischen Fähigkeiten auf das Lösen der Physikaufgaben wurde zudem ein Rechentest entwickelt, der die mathematischen Anforderungen der Physikfachwissen-Items separat darstellt. Hierbei wurde auf die mathematischen Inhaltelemente Terme, Trigonometrie, Vektoren und Matrizen, Differenzieren und Integrieren fokussiert.

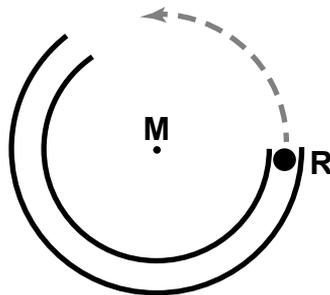
Fachdidaktisches Wissen

Im Zuge der Entwicklung des FDW-Tests wurden zunächst die Subdimensionen des Modells aus 3.2 im Sinne einer deduktiven Testentwicklung ausdifferenziert, um Kerninhalte für einzelne Items zu generieren, sodass eine bestmögliche Abbildung des Konstrukts gewährleistet wird. Davon ausgehend wurden zielgerichtet Items konstruiert, die bestimmte Kerninhalte adressieren sollen. In diesem Zusammenhang konnten einige Items, sofern sie eine Passung zum Modell aufwiesen, von Riese und Reinhold (2012) adaptiert werden. Alle Items wurden anschließend im gesamten Projektteam intensiv diskutiert und überarbeitet. Insgesamt wurden so 50 Aufgaben mit 91 sowohl offenen als auch geschlossenen (Multiple-Choice-Format, Multiple Select) Items konstruiert (Schülervorstellungen: 16 Aufgaben; Instruktionsstrategien: 7 Aufgaben; Experimente

Schüler sollen folgende Situation betrachten: Ein Ball rollt in der dargestellten Rinne (Draufsicht) und verlässt diese am Punkt R.



a) Schüler A zeichnet folgende Bahn, die der Ball nach Verlassen der Rinne beschreiben soll:



Lösung von Schüler A

Angenommen, der Schüler versteht die Zeichnung korrekt als Draufsicht: Welche fachlich nicht korrekte Vorstellung des Schülers A liegt bei der gezeichneten Bahnkurve zugrunde?

Abb. 3: Beispielitem FDW zu Schülervorstellungen zur Kreisbewegung

und Vermittlung eines angemessenen Wissenschaftsverständnisses: 12 Aufgaben; Fachdidaktische Konzepte: 15 Aufgaben). Diese wurden auf zwei Testhefte für jeweils 90 Minuten Bearbeitungszeit (inklusive Angaben zur Demografie) aufgeteilt, wobei 14 Aufgaben in beiden Testheften enthalten sind (Multi-Matrix-Design).

Abbildung 3 zeigt ein Item, welches sich auf die kognitive Anforderung „Analysieren“ und die Facette „Schülervorstellungen“ bezieht. Eine (falsche) Schülerantwort muss dahingehend analysiert werden, welche Schüler(wohl)vorstellung der Antwort zugrunde liegt. Die Idee zu diesem Item lieferte der Force Concept Inventory Test (vgl. Hestenes, Wells & Swackhamer, 1992).

Da ein Testheft mit einer Bearbeitungsdauer von 90 Minuten für den Einsatz im Verbund mit den anderen Tests zu umfangreich war, blieb Spielraum, Items mit der besten Inhaltsvalidität auszuwählen (entsprechende Analysen siehe 5.2). Für die Hauptstudie wurden schließlich 26 Aufgaben (43 Items) übernommen, die in einem Testheft mit einer Bearbeitungszeit von 75 Minuten (inklusive Demografie) zusammengefasst wurden.

Erklärungswissen

Das EW wird in einer standardisierten, interaktiven Testsituation erhoben, bei der die Probanden ein vorgegebenes physikalisches Phänomen einer Schülerin bzw. einem Schüler aus der zehnten Klasse erklären sollen, z. B. das Gefühl der Schwerelosigkeit bei der Achterbahnfahrt (vgl. Kulgemeyer & Schecker, 2013). In ihrer zehnminütigen Vorbereitungszeit werden den Erklärenden Anschauungsmaterialien zur Verfügung gestellt, die sie in der Erklärung nutzen können, beispielsweise Diagramme, Bilder oder einschlägige Formeln (siehe Abb. 4).

Die Erklärungen (10 Min. Dauer) werden dann videografiert. Den Probandinnen und Probanden ist nicht bekannt, dass die Schülerinnen und Schüler vorher darauf trainiert worden sind, sich in allen Erklärungssituationen in einer bestimmten Weise zu verhalten, insbesondere gleichartige Fragen zu stellen, z. B. „Gibt es dafür ein Beispiel?“.

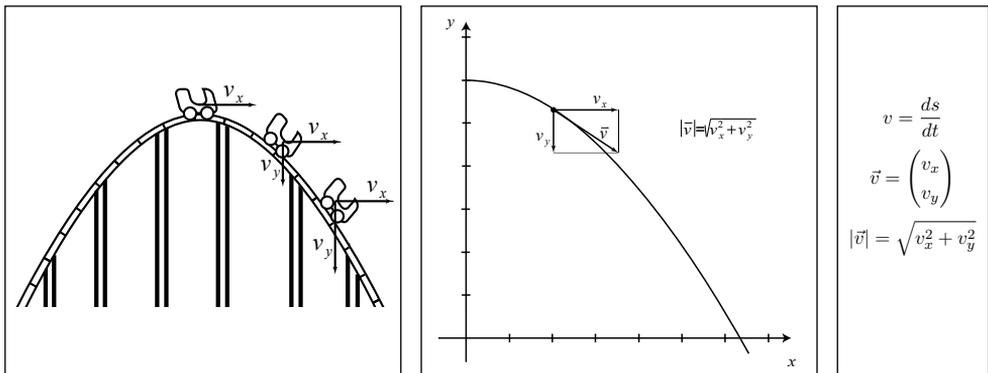


Abb. 4: Beispiele für zur Verfügung stehende Darstellungsformen zum Thema „Schwerelosigkeit in der Achterbahn“

Durch solche Prompts werden in allen Erklärungssituationen vergleichbare Anforderungen an die Variation der Variablen Mathematisierungsgrad, Beispiele, Sprachniveau und Darstellungsformen aus dem oben genannten Modell gestellt. Der Einsatz dieser Prompts wird mit den Schülerinnen oder Schülern in nachgestellten Erklärungsszenen geübt und anhand von Videoaufzeichnungen besprochen, bis das angestrebte Verhalten zur Routine geworden ist.

Zur weiteren Auswertung wurde ein Kategoriensystem mit 45 mittel- bis niedrig-inferent einzuschätzenden Kategorien mit Ankerbeispielen entwickelt, z. B. „Erklärer gibt ein Zahlenbeispiel“. Diese Kategorien wurden mit qualitativer Inhaltsanalyse (Mayring, 2003) gewonnen. Ziel war dabei das ökonomische Kodieren am Videomaterial ohne vorherige Transkribierung. Das Verfahren war dabei wie folgt: Ein Startsatz an Überkategorien ergab sich deduktiv aus dem Modell für EW, indem Variationen in den im Modell angenommenen Variablen kategorisiert wurden (z. B. Variationen im Sprachniveau, siehe 3.3). Diese Überkategorien wurden am Material induktiv in Unterkategorien verfeinert, wozu Ankerbeispiele festgehalten und daraus Generalisierungen erstellt wurden, die zur Unterkategorie weiterentwickelt wurden. So gab es unterschiedliche Arten, das Sprachniveau zu variieren, z. B. indem auf lexikalischer Ebene Fachbegriffe direkt nach Nennung erläutert oder ganz vermieden wurden und eine Erläuterung auf alltagssprachlichem Niveau erfolgte. Zudem wurde das Auftreten von aus der Literatur bekannten Qualitätskriterien für Erklärungen deduktiv mitkodiert (z. B. Evaluation des Verständnisses, siehe 3.3). Auch diese Kategorien dienten wieder als Oberkategorien, die in Unterkategorien differenziert wurden. Die Evaluation des Verständnisses konnte z. B. erfolgen, indem direkte Fragen oder Aufgaben gestellt wurden, die den erklärten Inhalt thematisierten. Es zeigte sich bei der Analyse der Videos durch einen Rater, dass die Kategorien und damit das Modell erschöpfend alle Variationen beschreiben, die die Probanden in ihren Erklärungen vornehmen, um deren Verständlichkeit zu verbessern. Alle Veränderungen konnten durch den Rater Kategorien zugeordnet werden. Zur Überprüfung der Reliabilität der Einstufungen wurden Interraterreliabilitäten bei den Kategorien berechnet. Dabei werden bei zwei Ratern bislang Übereinstimmungen zwischen 73 % und 97 % erreicht. Durch anschließende Kommunikation konnte in allen Fällen Einigkeit erzielt werden, weitere Berechnungen laufen.

In einer ersten Vorstudie wurden fünf physikalische Themen zur Erklärung eingesetzt. Hiervon wurden drei im Anschluss weiter verwendet. Kriterien für die Auswahl waren neben curricularer Validität (siehe 5.3) vor allem ein gemessen am fachlich-physikalischen Inhalt vergleichbarer Schwierigkeitsgrad. Dazu wurden die Videos gesichtet und Themen, bei denen gehäuft Probleme auftraten, nach einvernehmlicher Diskussion von drei Personen im Projektteam gestrichen.

Miterhoben werden zudem Erklärungserfahrung, erklärungsbezogene Selbstwirksamkeitserwartung und Interesse am Erklären. Dazu wurden eigene Fragebögen mit Likert-Skalen entwickelt, für die Selbstwirksamkeitserwartung ergibt sich ebenso eine reliable Skala (Cronbachs $\alpha = 0.71$) wie für das Interesse am Erklären (Cronbachs $\alpha = 0.79$).

4.2 Datenerhebungen

Die beschriebenen Testinstrumente wurden für den Einsatz bei Studierenden der Lehrämter an Gymnasien/Gesamtschulen und Haupt-/Real-/Gesamtschulen entwickelt. In einer Begleiterhebung zum FW, FDW und EW werden neben allgemeinen Personenmerkmalen (PM) die Rechenfähigkeit, kognitiv-sprachliche Fähigkeiten, der Studienfortschritt (SWS und Credit Points) als Maß für bisherige Lerngelegenheiten sowie fachbezogenes Interesse, fachbezogenes Selbstkonzept, die Fähigkeit zur Perspektivenübernahme und epistemologische Überzeugungen erhoben. Die Daten werden für die Aufklärung bestimmter Fähigkeitsausprägungen in den Bereichen des FW, FDW und EW genutzt.

In einer ersten Erhebungswelle, die als Querschnittserhebung über alle Studiensemester primär zur Generierung von Validitätsargumenten durchgeführt wurde, wurden die jeweiligen (noch zu umfangreichen) Tests zunächst in Stichproben eingesetzt, die nur zum Teil zusammenhängen, um die Testzeit der Probanden im erträglichen Rahmen zu halten. Die Probandenzahlen der einzelnen Teiluntersuchungen sind in den entsprechenden Teilen von Abschnitt 5 aufgeführt.

In der laufenden Haupterhebung, bei der alle Instrumente in einer gemeinsamen Stichprobe eingesetzt werden, werden ca. 500 Testpersonen unterschiedlicher Studiensemester an den vier Universitäten des Forschungsverbands jeweils schriftlich zum FW und FDW getestet, davon sollen ca. 200 in etwa gleich verteilt über die verschiedenen Studiensemester zusätzlich zum EW getestet werden. Der Datenpool ermöglicht somit bereichsübergreifende Zusammenhangsanalysen. Die Datenerhebung erfolgt dabei mittels einer Kombination aus Quasi-Längsschnitt und echtem Längsschnitt (Erhebungen vom 1. bis 3. Semester, Erhebungen vom 3. bis 5. Semester) sowie eines Querschnitts bei fortgeschrittenen Studierenden.

4.3 Testtheorie

Um einen Zusammenhang zwischen den Rohdaten und den daraus resultierenden Leistungsscores herzustellen, werden in ProfiLe-P verschiedene testtheoretische Ansätze verwendet. Für die Tests zur Erfassung des FW und des FDW werden Verfahren der probabilistischen Testtheorie genutzt, da diese gegenüber der klassischen Testtheorie mehr Möglichkeiten bieten (z. B. Stichprobenunabhängigkeit, Ermöglichung einer großen Itemzahl zur angemessenen Abdeckung der Kompetenzmodelle im Multi-Matrix-Design; vgl. Rost, 2004; Moosbrugger & Kelava, 2012). Konkret werden im Falle von FW und FDW ein- und mehrdimensionale Rasch-Modelle verwendet, um Zusammenhänge zwischen der Lösungswahrscheinlichkeit der Testitems und den verschiedenen Parametern herzustellen. Zur Überprüfung von Hypothesen zur Dimensionalität der erfassten Konstrukte werden unterschiedliche Modelle in Bezug auf die Passung zum Datensatz mittels der Informationskoeffizienten AIC und BIC sowie mithilfe des Chi-Quadrat-Differenztests verglichen (ebd.).

Auf eine Verwendung des 2-pl-Birnbaum-Modells wird aufgrund der Stichprobengröße in den hier berichteten Studien zunächst verzichtet, da in der Literatur eine Verwendung nur dann als sinnvoll erachtet wird, wenn eine große Stichprobe zur Verfügung steht, die das gesamte Fähigkeitsspektrum hinreichend abdeckt (vgl. Moosbrugger & Kelava, 2012). Diese Bedingungen sind jedoch für die berichteten Studien zur Sammlung von Validitätsargumenten nicht gegeben, da z. B. mit dem FDW-Test nur gut 200 Probanden befragt werden konnten, wobei nicht genug Personen im oberen Fähigkeitsbereich zu verorten sind.

Im Hinblick auf die deutlich aufwendiger zu generierenden Daten des Tests zur Erfassung des EW mittels Videografie (vgl. 4.2) stehen bisher nicht genug Probanden zur Verfügung, um Methoden der probabilistischen Testtheorie nutzen zu können. Dementsprechend erfolgt für diesen Fall die Nutzung der klassischen Testtheorie, um die Fähigkeitsparameter zu generieren.

5. Nachweis von Validität der Erhebungsinstrumente

Zur Validierung verfolgt ProfiLe-P einen Argument-based-Ansatz, der sich auf die Interpretation eines Testwerts bezieht, wie ihn Jenßen, Dunekacke und Blömeke (2015) im Einleitungsbeitrag dieses Beihefts vorschlagen (vgl. auch Kane, 2013). Dabei ist zu beachten, dass die Validität einer Testwertinterpretation – z. B. bzgl. der Verallgemeinerung eines Testergebnisses über die spezifischen Testinhalte hinaus – zwar durch beigebrachte Validitätsargumente unterstützt oder zurückgewiesen, nicht jedoch abschließend belegt werden kann (vgl. ebd.). In diesem Sinne wurden mehrere Teiluntersuchungen zur Generierung von Validitätsargumenten durchgeführt, die im Folgenden erläutert werden. Dabei sollen die entwickelten Testinstrumente primär eine Testwertinterpretation legitimieren, die auf die Erklärung von Leistungen abzielt (angelehnt an die Unterscheidung von Hartig, Frey & Jude, 2012).

Die in ProfiLe-P durchgeführten Validierungsstudien werden im Folgenden anhand der methodischen Kriterien *Inhaltsvalidität*, *Konstruktvalidität* und *Kriteriumsvalidität* gegliedert, welche im Bereich der Leistungsmessung üblicherweise unterschieden werden (vgl. Einleitungsbeitrag von Jenßen et al., 2015). Um die Probanden nicht über Gebühr zu beanspruchen, wurden die unterschiedlichen Validierungsstudien nicht in einer gemeinsamen Stichprobe, sondern für die Teilprojekte getrennt durchgeführt.

5.1 Fachwissen

Bezogen auf die Modelldimension „Fachstufen“ (Schulwissen, vertieftes Schulwissen, universitäres Wissen) erfolgte die curriculare Validierung des FW-Tests auf der Basis von Lehrbuch-Analysen für Schulen und Universitäten, Analysen der Vorlesungsinhalte für die Einführungsvorlesungen und unter Berücksichtigung der Richtlinien und Lehrpläne an weiterführenden Schulen. Es wurden dazu die Anforderungen und Inhalte aller

Einführungsvorlesungen der beteiligten Universitäten analysiert, die Unterschiede der Konzepte (z. B. die Energieerhaltung) auf den verschiedenen Fachstufen herausgearbeitet und die Aufgaben mit Fachphysikern und Fachdidaktikern diskutiert. Zudem wurde zur Inhaltsvalidierung mithilfe einer Konsenskodierung durch Lehramtsstudierende der Physik höherer Semester noch einmal geprüft, ob die Aufgaben als prototypisch zu den intendierten Teildimensionen des Modells konstruiert wurden. Hierbei wurde festgestellt, dass für die Fachstufe *vertieftes Schulwissen* von den in 3.1 benannten Fähigkeiten nur Aufgaben zu den Fähigkeiten 3 (Randbedingungen einer Schulaufgabe erkennen) und 5 (Zusammenhänge, Gemeinsamkeiten und Unterschiede physikalischer Phänomene erkennen) konstruiert wurden.

Um Hinweise zur Konstruktvalidität zu erhalten, wurde erstens die Dimensionalität der hypothetisch angenommenen inneren Struktur des Modells zum Fachwissen (wie in 3.1 beschrieben) geprüft und zweitens untersucht, in welcher Form theoretisch angenommene Zusammenhänge mit anderen Variablen vorliegen (Verortung des Konstrukts in einem nomologischen Netzwerk mittels Korrelationsanalysen; vgl. auch Einleitungsbeitrag von Jenßen et al., 2015). Dazu wurden Erhebungen an bislang vier Standorten mit insgesamt 469 Studierenden durchgeführt, davon 186 Lehramtsstudierende, 222 Physik-Monobachelorstudierende und 61 sonstige Studierende der ersten vier Semester.

Zunächst zeigt sich hier beim Vergleich des 1D-Modells der Raschanalyse (EAP/PV-Reliabilität: 0.83; Varianz: 0.90) mit dem 3D-Modell, dass das 3D-Modell zu bevorzugen ist. Die EAP/PV-Reliabilitäten für die drei Fachstufen sind gut, die Varianzen akzeptabel (vgl. Tab. 1), und die Unterschiede zwischen den Modellen sind signifikant (siehe Abb. 5).

Weiter wurde untersucht, welcher Zusammenhang zwischen der Personenfähigkeit im FW-Test mit Rechenfähigkeiten und Zeugnisnoten besteht. Dabei zeigte sich eine hohe Korrelation zwischen FW-Test und Rechentest (siehe Tab. 2), wobei die Korrelation vom Schulwissen über das vertiefte Schulwissen zum universitären Wissen hin abnimmt. Einschränkend muss angemerkt werden, dass die Korrelationen nur für die Hälfte der Stichprobe berechnet werden konnten, da die Codes der Studierenden, die sowohl am Fachwissens- als auch am Rechentest (zusammen mit dem FDW-Test zu einem anderen Zeitpunkt) teilgenommen haben, oft nicht übereinstimmten.

Bei den Noten zeigt sich eine ähnlich hohe Korrelation des Mechaniktests zur Abitur- und Mathematiknote, gefolgt von der Physiknote. Der Zusammenhang zur Deutschnote ist gering. Es ist auffällig, dass die Korrelationen zum Schulwissen höher als zu den beiden anderen Fachstufen sind. Dies ist plausibel, da dieses Wissen zu Beginn des Studiums wenig ausgeprägt ist und geraten werden muss. Ein vergleichbarer Zusammenhang der Abitur-, Mathematik- und Physiknote mit Fachwissenstests zeigt sich auch in anderen Studien (z. B. Schoppmeier et al., 2013), wobei der Zusammenhang dort höher ist. Momentan wird eine Konstruktvalidierung mit dem ähnlich konstruierten FW-Test von Woitkowski et al. (2011) durchgeführt.

Fachstufe	EAP/PV-Reliabilität	Varianz
Schulwissen	.83	1.43
Vertieftes Schulwissen	.78	1.05
Universitäres Wissen	.81	0.98

Tab. 1: EAP/PV-Reliabilitäten und Varianzen der Fachstufen beim FW

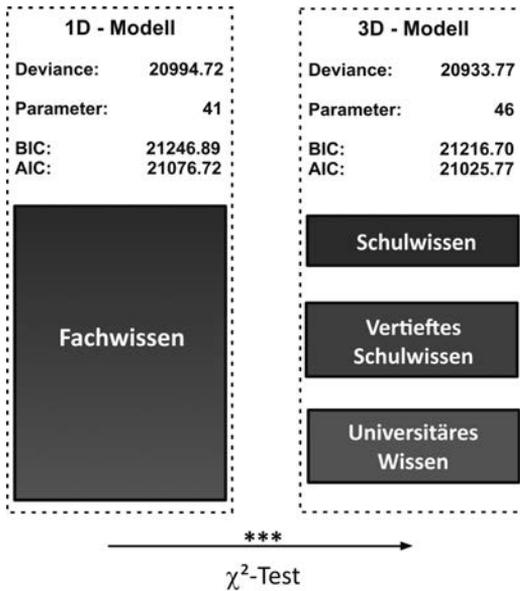


Abb. 5: Vergleich von 1D- und 3D-Modell bzgl. Fachstufen

	Schulwissen			Vertieftes Schulwissen			Universitäres Wissen		
	df	r	p	df	r	p	df	r	p
Rechentest	274	.67	<.001	274	.51	<.001	274	.41	<.001
Abiturnote	456	.46	<.001	456	.32	<.001	456	.36	<.001
Mathenote	447	.44	<.001	447	.27	<.001	447	.32	<.001
Physiknote	425	.35	<.001	425	.25	<.001	425	.26	<.001
Deutschnote	429	.15	.002	429	.10	.04	429	.12	.01

Tab. 2: Zusammenhang zwischen den drei Fachstufen und Kontrollvariablen

5.2 Fachdidaktisches Wissen

Zunächst wurden Analysen zur Inhaltsvalidierung durchgeführt, wobei die curriculare Validität als Teil der Inhaltsvalidität betrachtet wird (vgl. Jenßen et al., 2015, in diesem Beiheft). Hierzu wurden Curriculum-Analysen durchgeführt, indem die Lehrenden an den Erhebungsstandorten zur Passung der Testinhalte befragt wurden. Mithilfe dieser Daten konnte jedem Item ein Wert für die „curriculare Passung“ zugewiesen werden. Auf einer Skala von 1 (sehr gute Passung) bis 5 (gar keine Passung) erzielten die Items Werte von 1.0 bis 3.2. Der Mittelwert aller Items lag bei 2.1 (Standardabweichung 0.6).

Zur weiteren Inhaltsvalidierung wurde eine „Think-aloud“-Studie mit 15 Studierenden (5 Lehramt HR, 5 Lehramt GyGe, 5 Fachwissenschaftler) durchgeführt. Sie sollte Rückschlüsse auf die Wissensbestände geben, die zur erfolgreichen Bearbeitung der Testitems nötig sind. Sofern beispielsweise ermittelt wurde, dass ein Item ausschließlich unter Nutzung von physikalischem FW beantwortet werden konnte, wurde es von der weiteren Verwendung ausgeschlossen. Schließlich wurden im Zusammenhang mit der Inhaltsvalidierung drei Experten (Physikdidaktiker) gebeten, die jeweiligen Testitems in das zugrunde liegende Modell (4 fachdidaktische Facetten und 3 kognitive Anforderungen) einzuordnen, um sicherzustellen, dass die Items als prototypisch für den intendierten Anforderungsbereich angesehen werden können. Dabei zeigt sich eine gute Übereinstimmung mit Cohens Kappa von .83 bis .89 für die Zuordnung der Items zu den Facetten und Kappa von .63 bis .81 für die Zuordnung zu den kognitiven Anforderungen.

Um weiter im Sinne einer Konstruktvalidierung die hypothetisch angenommene Struktur des FDW-Modells (vgl. 3.2) mit Methoden der probabilistischen Testtheorie zu prüfen, wurden 216 Lehramtsstudierende der Physik (davon 127 Gymnasialbereich und 71 Haupt-Realschulbereich) aus 12 Standorten getestet. Dabei wurde ein 1D-Rasch-Modell (EAP/PV-Reliabilität: .84; Varianz: .39) mit einem 4D-Rasch-Modell verglichen, das sich gemäß dem zugrunde liegenden Modell in die fachdidaktischen Facetten *Instruktionsstrategien*, *Schülervorstellungen*, *Experimente* und *Vermittlung eines angemessenen Wissenschaftsverständnisses* sowie *Fachdidaktische Konzepte* aufspaltet. Dabei passt das 4D-Modell hochsignifikant besser als das 1D-Modell (Chi-Quadrat-Test, vgl. Abb. 6). Wie beim 1D-Modell sind die Varianzen der einzelnen Facetten gering und ihre EAP/PV-Reliabilitäten sind akzeptabel (vgl. Tab. 3).

Eine Trennung der vier theoretisch angenommenen Facetten ist somit auch statistisch vertretbar, was als Indiz für die Konstruktvalidität gewertet werden kann. Im Zusammenhang dieser Analysen wurden Items mit schlechtem Fit aus dem Test entfernt (falls nicht $0.8 < \text{gewichteter MNSQ} < 1.2$ und $-1.9 < T < 1.9$). Letztlich wurden unter Nutzung aller Erkenntnisse aus den berichteten Untersuchungen 43 von 91 Items für die weiteren Erhebungen zugelassen, womit gleichzeitig die Testzeit auf 60 Minuten begrenzt werden konnte.

Schließlich wurden theoretisch angenommene Zusammenhänge des mit dem FDW-Test ermittelten Konstrukts zu anderen Variablen überprüft (Prüfung eines nomologischen Netzes; vgl. Jenßen et al., 2015, in diesem Beiheft). Konkret wird erwartet, dass

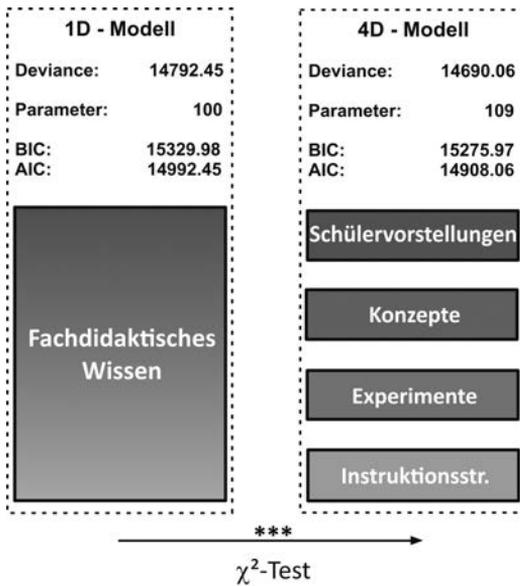


Abb. 6: Vergleich von 1D- und 4D-Modell fachdidaktischer Facetten

Facette	EAP/PV-Reliabilität	Varianz
Schülervorstellungen	.69	.52
Fachdidaktische Konzepte	.76	.78
Experimente	.74	.48
Instruktionsstrategien	.62	.49

Tab. 3: EAP/PV-Reliabilitäten und Varianzen der Facetten beim FDW

die Fähigkeitswerte des 4D-Rasch-Modells eine hohe Korrelation mit der Anzahl der Credit Points (CP) aufweisen, die in physikdidaktischen Seminaren erworben wurden (Indikator für die bisherigen Lerngelegenheiten), während geringere Korrelationen mit der Abiturnote (Indikator für allgemeine kognitive Leistungsfähigkeit), der letzten Physiknote (Indikator für das fachbezogene Vorwissen) und der letzten Deutschnote (Indikator für sprachliche Fähigkeiten) erwartet werden. Wie Tabelle 4 zeigt, sind erwartungskonform die höchsten und durchweg signifikanten Korrelationen mit den erworbenen CP in Physikdidaktik zu beobachten, was darauf hindeutet, dass der FDW-Test Wissens-elemente erfasst, die tatsächlich in fachdidaktischen Seminaren erworben werden. Die restlichen Korrelationen sind geringer, aber unterschiedlich für die einzelnen Facetten, was wiederum die Annahme stützt, dass die vier Facetten unterschiedliche Konstrukte darstellen. Insgesamt können die Korrelationsanalysen als weiteres Argument für die

	Schüler- vorstellungen			Fachdidaktische Konzepte			Experimente			Instruktions- strategien		
	df	r	p	df	r	p	df	r	p	df	r	p
CP Physikdidaktik	194	.26	<.001	194	.20	.005	194	.36	<.001	194	.27	<.001
Abiturnote	205	.16	.020	205	.10	n. s.	205	.25	<.001	205	.12	n. s.
Physiknote	200	.06	n. s.	200	.05	n. s.	200	.24	<.001	200	.15	.032
Deutschnote	201	.07	n. s.	201	.12	n. s.	201	.20	.004	201	.12	n. s.

Tab. 4: Zusammenhang zwischen den vier Facetten und Kontrollvariablen

Konstruktvalidierung verstanden werden, da sie Testwertinterpretationen stützen, die auf die Erklärung von Leistungen abzielen.

5.3 Erklärungswissen

Um das EW in seiner Ausprägung auswerten zu können, braucht es einen validen Gütemaßstab. Zunächst wurden deshalb Einblicke in die Inhaltsvalidität gewonnen, anschließend wurde auf Basis der Kategorien ein solcher Maßstab entwickelt und mit einer Expertenerhebung validiert – in diesem Falle ist die Expertenbefragung zur Diskussion von Kriteriumsvalidität genutzt worden (vgl. Jenßen et al., 2015, in diesem Beiheft).

Die curriculare Validität der zu den Erklärungen ausgewählten Themen wurde unter anderem durch den Vergleich der Themen mit den Inhalten von gängigen Lehrbüchern der Experimentalphysik sowie von verbreiteten Physikschulbüchern sichergestellt. Alle Themen bilden dort Schwerpunkte.

Das diagnostische Verfahren selbst entspricht einer berufsbezogenen Anforderung, nämlich dem Erklären – dies spricht für Inhaltsvalidität („triviale Validität“; vgl. Jenßen et al., 2015, in diesem Beiheft). In einer begleitenden Interviewstudie wurden zusätzlich sieben Testpersonen zu den Erklärungen befragt. Es wurden besonders solche Testpersonen ausgewählt, die nicht auf die vorhergesehene Art auf die standardisierten Schülerfragen geantwortet haben. Die Probanden gaben auf direkte Nachfrage in Nachinterviews einheitlich an, dass sie die Settings als authentische Erklärungssituationen empfunden haben. Insbesondere haben sie nicht bemerkt, dass die Schülerinnen und Schüler trainiert waren. Zudem wurden die Testpersonen anhand ihrer Erklärungsvideos zu den Gründen ihrer von Ratern als problematisch eingeschätzten Reaktionen befragt (Video-Stimulated Recall). Aus diesen Interviews ging hervor, dass es sich bei den Gründen für das möglicherweise problematische Verhalten in den Erklärungen nicht um fachlich-physikalische Defizite, sondern um Probleme mit der Art des Erklärens handelte. Beispielsweise gaben zwei der Testpersonen an, dass sie Probleme mit der Erkundung des Vorwissens der Schüler hatten. Keine der Testpersonen nannte den fachlichen

Schwierigkeitsgrad als Grund. Die Probleme liegen also im intendierten Zielbereich, dies kann als weiteres Argument für Inhaltsvalidität gesehen werden.

Zur Vorbereitung der Entwicklung eines quantitativen Gütemaßes aus den qualitativen Daten für die Erklärungen wurde ein Expertenrating durchgeführt, das der konkurrenten Validierung dienen soll (Kriteriumsvalidität). Das naheliegende Außenkriterium zu erheben – nämlich das Verständnis der Schülerinnen und Schüler, die als Adressaten fungieren – konnte hier nicht angewendet werden. Die Schülerinnen und Schüler sind schließlich im Sinne der Methode zu standardisiertem Verhalten trainiert worden und kannten die Themen. Deshalb wurden sieben Experten (Fachdidaktiker und Lehrkräfte mit hoher Lehrerfahrung) jeweils Paare von insgesamt 16 Erklärungsvideos gezeigt. Die Experten mussten hochinferent entscheiden, in welchem Video die bessere Erklärung vorlag. Jedes Paar wurde von fünf Experten beurteilt. Diese Validierungsstrategie durch Paarvergleiche wurde bei komplexen Merkmalen schon häufiger gewählt (Kulgemeyer & Schecker, 2013), gerade bei hoher Interraterreliabilität mehrerer Rater wird ihr Aussagekraft zugemessen. Es ergab sich hier eine sehr gute Beurteilerübereinstimmung über die bessere Erklärung (95%, Fleiss' $\kappa = .80$). Diese Expertenentscheidung kann zur späteren konvergenten Validierung eines Bewertungsmaßes für Erklärungswissen verwendet werden.

Anschließend wurde schrittweise ein Bewertungsmaßstab für EW entwickelt. Aus den bis dahin vorliegenden 45 Kategorien bzgl. guter Erklärungen wurden diejenigen ausgewählt, die nach der Theorie positiv für Erklärungsqualität sind (z. B. „Evaluation des Verständnisses“, siehe 3.3). Ein weiteres Auswahlkriterium war das gehäufte Auftreten in solchen Erklärungen, die von den Experten als besonders gut oder als besonders schlecht eingeschätzt wurden. Die Kategorien, die in diesen beiden „Polen“ von Erklärungen besonders häufig vorkommen, wurden für die weiteren Analysen berücksichtigt. Über das Vorkommen der verbleibenden 26 positiven und negativen Kategorien in einer Erklärung wurde analog zu dichotomen Testitems summiert (0/1), wobei die negativen Kategorien umgekehrt gepolt berücksichtigt wurden. Es ergibt sich eine Skala für EW mit zufriedenstellender Reliabilität (Cronbachs $\alpha = .74$).

In nächsten Schritt sollte unter dem Aspekt der Auswertungsökonomie die Anzahl der Kategorien reduziert werden, ohne Validität oder Reliabilität der Skala zu verringern. Dazu wurden in einem Wechselspiel der Prüfung der Auswirkungen auf Reliabilität (Item-Skala-Korrelation, Änderungen in Cronbachs Alpha) und Prognosefähigkeit (Übereinstimmung mit dem Expertenrating, Cohens Kappa) schrittweise Kategorien gestrichen. Es sollten dabei Kategorien gestrichen werden, ohne diese beiden Kriterien zu verschlechtern. Die verbleibenden zwölf Kategorien sagen die Expertenentscheidung über die bessere Erklärung im Paarvergleich zweier Videomittschnitte voraus. Jeder Bereich des Kommunikationsmodells (Sprache, Beispiele, Darstellungsform, Mathematisierung) ist dabei im Maß mit mindestens einer Kategorie vertreten (z. B. „Gibt Zahlenbeispiele“ für Mathematisierung). Die Übereinstimmung zwischen den mit diesem schlankeren Maß getroffenen Entscheidungen über die bessere Erklärung im Paarvergleich und den Ergebnissen aus dem Expertenrating beträgt Cohens $\kappa = .75$ (87.5%). Die Reliabilität der Skala steigt sogar (Cronbachs $\alpha = .77$) und ist damit im guten Bereich.

6. Diskussion und Ausblick

Im Projektverbund ProfiLe-P wurde ein Ansatz zur Messung des Professionswissens von Physiklehrkräften entwickelt und validiert, der zentrale Wissens Elemente in einem gemeinsamen Rahmenmodell differenziert abbildet und damit bereichsübergreifend überprüfbar macht. Ausgehend von drei bereichsspezifischen Kompetenzmodellen können so physikalisches Fachwissen (FW), physikbezogenes fachdidaktisches Wissen (FDW) und die Fähigkeit zum Erklären von Physik (EW) zueinander in Beziehung gesetzt werden. Damit können erstmals im Rahmen von Large-Scale-Assessments in den Naturwissenschaften Zusammenhänge zwischen dem Professionswissen und Elementen des Handelns von angehenden Lehrkräften verhältnismäßig ökonomisch untersucht werden, indem zeitlich begrenzte (10 Min.) Lehr-Lern-Rollenspiele mit trainierten „Schülerinnen und Schülern“ in Laborsituationen videografiert werden.

Wie für alle empirischen Studien stellt sich die Frage, inwieweit dem fundamentalen und zugleich komplexesten Gütekriterium, der Validität, Rechnung getragen wird. Dementsprechend wurde in ProfiLe-P eine Reihe von Untersuchungen durchgeführt, um Argumente zur Inhaltsvalidität, zur Konstruktvalidität und zur Kriteriumsvalidität zu erhalten. Insgesamt deuten die durchgeführten Studien in der Summe darauf hin, dass die gewählten Testverfahren valide Testwertinterpretationen im Hinblick auf die Erklärung von Leistungen zulassen. Bislang liegt jedoch noch keine ausreichende Evidenz vor, inwieweit die Tests Interpretationen zulassen, die auf eine (individuelle) Bewertung von Leistung abzielen, wenn die Leistungen einer Person mit den Leistungen einer anderen Person verglichen werden (vgl. Hartig et al., 2012).

Darüber hinaus bleiben für die laufenden Erhebungen noch einige offene Fragen. So werden Analysen zur Trennbarkeit des vertieften Schulwissens vom universitären Wissen möglicherweise dadurch verzerrt, dass nicht alle spezifischen Fähigkeiten im FW-Test berücksichtigt werden. Die Facetten des dem FDW-Test zugrunde liegenden Modells wiederum lassen sich zwar empirisch trennen und ermöglichen so eine differenziertere Messung als vergleichbare Tests bisheriger Studien, allerdings ist der Test insgesamt etwas zu schwer. Damit sind Aussagen zu leistungsschwachen Gruppen nur bedingt möglich. Weiterhin findet die fachliche Korrektheit der Erklärung beim Maß für das EW wenig Beachtung, bislang wird vor allem die Adressatengemäßheit operationalisiert. Daran wird derzeit gearbeitet, um den dialogischen Aspekt des Erklärens stärker hervorzuheben und differenziertere Aussagen zu ermöglichen. Wenn die angestrebte Stichprobe von $N = 200$ erreicht ist, könnte auch im EW mit Methoden der probabilistischen Testtheorie eine Modellierung geprüft werden. Wir gehen davon aus, dass die Daten aus den laufenden Erhebungen weitere Aufschlüsse ermöglichen. Zudem hat sich die Kodierung der offenen Items des FDW-Tests als sehr aufwendig erwiesen, da nicht immer eindeutig zu entscheiden ist, ob eine Antwort als falsch oder richtig zu beurteilen ist. Auch die Kodierung der videografierten Erklärungen des Tests zur Erfassung des EW ist bislang noch sehr aufwendig, sodass hier für Folgestudien zu prüfen ist, inwieweit einzelne Anforderungen mit geschlossenen oder online-basierten Items erfasst werden können.

Grundsätzlich ist das der Itementwicklung zugrunde liegende Modell konzeptionell anchlussfähig an die aktuelle Diskussion zur Lehrerbildungsforschung, da es eine Weiterentwicklung bzw. Adaption von vorliegenden Modellen darstellt. Daher werden auf Basis der Auswertung der Haupterhebung, in der alle Instrumente in einer gemeinsamen Stichprobe zum Einsatz kommen, vergleichende Auswertungen und Interpretationen mit Studien in anderen Fachdomänen, bei denen Bezüge zwischen Bereichen des Professionswissens einerseits und Aspekten der Unterrichtsqualität oder des Lehrerhandelns in realitätsnahen Situationen andererseits hergestellt werden (für Mathematiklehrkräfte z. B. Kunter et al., 2013), grundsätzlich ermöglicht. Einschränkend muss dabei angemerkt werden, dass sich die Ergebnisse der Haupterhebung nur auf die Erhebungen an vier Lehrerbildungsstandorten in drei Bundesländern beziehen (werden), was die Generalisierbarkeit der Ergebnisse zum Teil begrenzt. Dies ist vor allem der Tatsache geschuldet, dass die Probanden zeitlich sehr stark in Anspruch genommen werden (mehrere Testhefte an den jeweiligen Messzeitpunkten), was die Probandenakquise an „fremden“ Lehrerbildungsstandorten deutlich erschwert. Auf der anderen Seite wurden Pilotierungen, Begleituntersuchungen und curriculare Analysen für einzelne Testteile an insgesamt zwölf Lehrerbildungsstandorten durchgeführt und es wurden allgemeine Standards der Lehrerbildung für die Entwicklung der Tests herangezogen, sodass zumindest im Zusammenhang mit der Testentwicklung bestmöglich gewährleistet sein sollte, dass die Tests bundesweit einsetzbar sind.

Schließlich verfolgen wir u. a. die folgenden bereichsverbindenden Fragestellungen im Zusammenhang mit der noch laufenden Haupterhebung, bei der alle Instrumente in einer gemeinsamen Stichprobe eingesetzt werden:

- Wie viel Varianz im Erklärungswissen lässt sich durch das Fachwissen und das fachdidaktische Wissen aufklären?
- Gibt es über globale Zusammenhänge zwischen den untersuchten drei Wissensbereichen hinausgehende spezifische Zusammenhänge auf der Ebene jeweiliger Unterfacetten?
- Wie entwickeln sich die verschiedenen Teilbereiche des Professionswissens im Längsschnitt des ersten und zweiten Studienjahres?

Perspektivisch bieten die entwickelten Testinstrumente die Möglichkeit, Kompetenzstandenserhebung in zentralen Abschnitten der universitären Lehramtsausbildung durchführen zu können, womit Voraussetzungen sowohl für die Durchführung von Interventionsstudien zur Wirkung in der fachlichen und fachdidaktischen Ausbildung als auch für die Evaluation von Studienprogrammen im Lehramtsstudium Physik gegeben sind. Dabei erscheint ein Transfer der genutzten Methoden im Zusammenhang mit der Entwicklung des Kompetenzmodells und der Testinstrumente in andere Fächer grundsätzlich möglich.

Literatur

- Abell, S. K. (2007). Research on Science Teacher Knowledge. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (S. 1105–1149). Mahwah: Lawrence Erlbaum.
- Anderson, L. W., & Krathwohl, D. R. (Hrsg.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520.
- Blömeke, S., & Zlatkin-Troitschanskaia, O. (2013). *Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor: Ziele, theoretischer Rahmen, Design und Herausforderungen des BMBF-Forschungsprogramms KoKoHs* (KoKoHs Working Papers, 1). Berlin/Mainz: Humboldt-Universität/Johannes Gutenberg-Universität.
- Borowski, A., & Riese, J. (2010). Physikalisch-fachdidaktisches Wissen – Was kommt in der Praxis an? *Praxis der Naturwissenschaften – Physik in der Schule*, 59(5), 5–8.
- Brown, G. (2006). Explaining. In O. Hargie (Hrsg.), *The handbook of communication skills* (S. 195–228). East Sussex: Taylor & Francis.
- Deutsche Physikalische Gesellschaft (DPG) (2014). *Zur fachlichen und fachdidaktischen Ausbildung für das Lehramt Physik*. Bad Honnef: Deutsche Physikalische Gesellschaft.
- Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional Knowledge of Science Teachers. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second international handbook of science education* (S. 435–448). Dordrecht: Springer.
- Geelan, D. (2012). Teacher Explanations. In B. Fraser, K. Tobin & C. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 987–999). Dordrecht: Springer.
- Gesellschaft für Fachdidaktik (GFD) (2004). *Kerncurriculum Fachdidaktik. Orientierungsrahmen für alle Fachdidaktiken*. Entwurf des Arbeitskreises „Kerncurriculum“ Fachdidaktik der Gesellschaft für Fachdidaktik e. V.
- Gramzow, Y., Riese, J., & Reinhold, P. (2013). Modellierung fachdidaktischen Wissens angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 31–49.
- Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–173). Berlin: Springer.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141–158.
- Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. *Zeitschrift für Pädagogik*, 61. Beiheft, 11–31.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Korneck, F., Lamprecht, J., Wodzinski, R., & Schecker, H. (2010). *Quereinsteiger in das Lehramt Physik – Lage und Perspektiven der Physiklehrrausbildung in Deutschland*. Bad Honnef: Deutsche Physikalische Gesellschaft.
- Kröger, J., Neumann, K., & Petersen, S. (2013). Messung professioneller Kompetenz im Fach Physik. In S. Bernholt (Hrsg.), *Inquiry-based Learning – Forschendes Lernen*. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Hannover 2012 (S. 533–535). Kiel: IPN.
- Kulgemeyer, Ch., & Schecker, H. (2013). Students explaining science – assessment of science communication competence. *Research in Science Education*, 43, 2235–2256.
- Kultusministerkonferenz (KMK) (2008). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerinnen- und Lehrerbildung* (Beschluss der Kultusministerkonferenz vom 16. Oktober 2008). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.

- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., et al. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim/Basel: Beltz.
- Moosbrugger, H., & Kelava, A. (Hrsg.) (2012). *Testtheorie und Fragebogenkonstruktion* (2. Aufl.). Berlin: Springer.
- Ohle, A. (2010). *Primary school teachers' content knowledge in physics and its impact on teaching and students' achievement*. Berlin: Logos.
- Olszewski, J. (2010). *The impact of physics teachers' Pedagogical content knowledge on teacher action and student outcomes*. Berlin: Logos.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T., & Pöhlmann, C. (Hrsg.) (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe*. Münster: Waxmann.
- Park, S., & Chen, Y. C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK) – Examples from high school biology classrooms. *Journal of Research in Science Teaching, 49*(7), 922–941.
- Riese, J., & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen – Empirische Hinweise für eine Verbesserung des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft, 15*(1), 111–143.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Hans Huber.
- Schecker, H. (1985). *Das Schilervorverständnis zur Mechanik – eine Untersuchung in der Sekundarstufe II unter Einbeziehung historischer und wissenschaftstheoretischer Aspekte* (Dissertation). Bremen: Universität.
- Schoppmeier, F., Borowski, A., & Fischer, H. E. (2013). Validierung eines Kompetenzmodells für Physik in der Sekundarstufe II. In S. Bernholt (Hrsg.), *Inquiry-based Learning – Forschendes Lernen*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Hannover 2012 (S. 206–208). Kiel: IPN.
- Shulman, L. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher, 15*, 4–14.
- Stichweh, R. (1996). Professionen in einer funktional differenzierten Gesellschaft. In A. Combe & W. Helsper (Hrsg.), *Pädagogische Professionalität – Untersuchungen zum Typus pädagogischen Handelns* (S. 49–69). Frankfurt a. M.: Suhrkamp.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., et al. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften, 18*, 7–28.
- Vogelsang, C., & Reinhold, P. (2013). Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften, 19*, 103–128.
- Wellenreuther, M. (2005). *Lehren und Lernen – aber wie? Empirisch-experimentelle Forschung zum Lehren und Lernen im Unterricht*. Baltmannsweiler: Schneider.
- Woitkowski, D., Riese, J., & Reinhold, P. (2011). Modellierung fachwissenschaftlicher Kompetenz angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften, 17*(1), 289–313.
- Zlatkin-Troitschanskaia, O., & Kuhn, C. (2010). Messung akademisch vermittelter Fertigkeiten und Kenntnisse von Studierenden bzw. Hochschulabsolventen – Analysen zum Forschungsstand. In *Arbeitspapiere WP Heft 56*. <http://www.wipaed.uni-mainz.de/lis/382.php> [11. 04. 2014].

Abstract: For improving teacher education, there has been an increasing interest in describing teachers' professional competencies and their development in the course of implementing educational programs. The focus of the present project is on modeling and measuring domain-specific and generic competencies that future physics teachers acquire during their university studies. The model comprises characteristics and relationships between physics content knowledge, pedagogical content knowledge, and skills for explaining physics phenomena. Based on the model, assessment instruments were developed and applied as paper-and-pencil-tests and videotaped expert-novice dialogues for measuring the competencies in a large sample of physics student teachers. Trials and validation suggest that our instruments are valid in terms of content and construct validities.

Keywords: Professional Knowledge, Student Teachers, Physics, Competencies, Validity

Anschrift der Autor(inn)en

Dr. Josef Riese, Universität Paderborn, AG Didaktik der Physik,
Warburger Straße 100, 33098 Paderborn, Deutschland
E-Mail: josef.riese@upb.de

Dr. Christoph Kulgemeyer, Universität Bremen, Institut für Didaktik der Naturwissenschaften,
Abteilung Physikdidaktik, Otto-Hahn-Allee 1, 28359 Bremen, Deutschland
E-Mail: kulgemeyer@physik.uni-bremen.de

Simon Zander, Universität Duisburg-Essen, Physikdidaktik,
Universitätsstraße 2, 45141 Essen, Deutschland
E-Mail: simon.zander@uni-due.de

Prof. Dr. Andreas Borowski, Universität Potsdam, AG Didaktik der Physik,
Karl-Liebknecht Straße 24/25, 14476 Potsdam, Deutschland
E-Mail: andreas.borowski@uni-potsdam.de

Prof. Dr. Hans E. Fischer, Universität Duisburg-Essen, Lehrstuhl für Didaktik
der Physik, Universitätsstraße 2, 45141 Essen, Deutschland
E-Mail: hans.fischer@uni-due.de

Yvonne Gramzow, Universität Paderborn, AG Didaktik der Physik,
Warburger Straße 100, 33098 Paderborn, Deutschland
E-Mail: ygramzow@upb.de

Prof. Dr. Peter Reinhold, Universität Paderborn, AG Didaktik der Physik,
Warburger Straße 100, 33098 Paderborn, Deutschland
E-Mail: peter.reinhold@upb.de

Prof. Dr. Horst Schecker, Universität Bremen, Institut für Didaktik der Naturwissenschaften,
Abteilung Physikdidaktik, Otto-Hahn-Allee 1, 28359 Bremen, Deutschland
E-Mail: schecker@physik.uni-bremen.de

Elisabeth Tomczyszyn, Universität Bremen, Institut für Didaktik der Naturwissenschaften,
Abteilung Physikdidaktik, Otto-Hahn-Allee 1, 28359 Bremen, Deutschland
E-Mail: e.tomczyszyn@uni-bremen.de