

Plöger, Wilfried; Scholl, Daniel; Seifert, Andreas
"Und sie bewegt sich doch!". Wie spezifische Lerngelegenheiten die bildungswissenschaftlichen Kompetenzen von Lehramtsstudierenden fördern können

Zeitschrift für Pädagogik 62 (2016) 1, S. 109-130



Quellenangabe/ Reference:

Plöger, Wilfried; Scholl, Daniel; Seifert, Andreas: "Und sie bewegt sich doch!". Wie spezifische Lerngelegenheiten die bildungswissenschaftlichen Kompetenzen von Lehramtsstudierenden fördern können - In: Zeitschrift für Pädagogik 62 (2016) 1, S. 109-130 - URN: urn:nbn:de:01111-pedocs-167084 - DOI: 10.25656/01:16708

<https://nbn-resolving.org/urn:nbn:de:01111-pedocs-167084>

<https://doi.org/10.25656/01:16708>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

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

Terms of use

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

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

Kontakt / Contact:

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

Mitglied der


Leibniz-Gemeinschaft

ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 1

Januar/Februar 2016

■ *Thementeil*

Raum und Räumlichkeit in der erziehungswissenschaftlichen Forschung

■ *Allgemeiner Teil*

Fallarbeit als Angebot – fallbasiertes Lernen als Nutzung: Empirische Ergebnisse zur kognitiven Belastung, Motivation und Emotionen bei der Arbeit mit Unterrichtsfällen

„Und sie bewegt sich doch!“ – Wie spezifische Lerngelegenheiten die bildungswissenschaftlichen Kompetenzen von Lehramtsstudierenden fördern können

Differenzierung im gymnasialen Mathematik- und Deutschunterricht – vor allem bei guten Diagnostiker/innen und in heterogenen Klassen?

Inhaltsverzeichnis

Thementeil: Raum und Räumlichkeit in der erziehungswissenschaftlichen Forschung

Fabian Kessl/Sabine Reh/Edith Glaser/Werner Thole

Raum und Räumlichkeit in der erziehungswissenschaftlichen Forschung.

Eine Einführung 1

Fabian Kessl

Erziehungswissenschaftliche Forschung zu Raum und Räumlichkeit.

Eine Verortung des Thementeils „Raum und Räumlichkeit

in der erziehungswissenschaftlichen Forschung“ 5

Sebastian Dirks/Caroline Fritsche/Maike Lippelt/Christian Reutlinger

Zur pädagogischen Herstellung städtischer Räume zwischen Ort

und Klient*in. Empirische Einblicke und theoretische Rückschlüsse 20

Sabine Andresen/Susann Fegter/Nora Iranee/Elena Bütow

Doing Urban Family: Städtischer Raum und elterliche Perspektive

auf Kindheit 34

Nils Berkemeyer/Björn Hermstein/Veronika Manitius

Auf dem Weg zu einer raumorientierten Schulsystemforschung.

Was können raumsensible Sozialtheorien für empirische Analysen

regionaler Bildungsdisparitäten leisten? 48

Jeanette Böhme/Viktoria Flasche/Ina Herrmann

Die Territorialisierung des (Schul-)Pädagogischen im urbanen Wandel.

Ein Forschungsbeitrag aus der Pädagogischen Morphologie 62

Deutscher Bildungsserver

Linktipps zum Thema „Raum und Räumlichkeit
in der erziehungswissenschaftlichen Forschung“ 79

Allgemeiner Teil

Marcus Syring/Thorsten Bohl/Marc Kleinknecht/Sebastian Kuntze/

Markus Rehm/Jürgen Schneider

Fallarbeit als Angebot – fallbasiertes Lernen als Nutzung:
Empirische Ergebnisse zur kognitiven Belastung, Motivation
und Emotionen bei der Arbeit mit Unterrichtsfällen 86

Wilfried Plöger/Daniel Scholl/Andreas Seifert

„Und sie bewegt sich doch!“ – Wie spezifische Lerngelegenheiten
die bildungswissenschaftlichen Kompetenzen von Lehramtsstudierenden
fördern können 109

Andrea Westphal/Anna Gronostaj/Miriam Vock/Rico Emmrich/

Peter Harych

Differenzierung im gymnasialen Mathematik- und Deutschunterricht –
vor allem bei guten Diagnostiker/innen und in heterogenen Klassen? 131

Besprechungen

Ulrike Greiner

Ewald Terhart/Hedda Bennewitz/Martin Rothland (Hrsg.):
Handbuch der Forschung zum Lehrerberuf (2. überarbeitete
und erweiterte Auflage) 149

Beat Manz

Claude Mouchet/Raymond Bénévent: Von Freinet zu Freud:
Die institutionelle Pädagogik von Fernand Oury 153

Dokumentation

Pädagogische Neuerscheinungen 156

Impressum U3

Table of Contents

Topic: Space and Spatiality in Educational Research

| | |
|--|----|
| <i>Fabian Kessl/Sabine Reh/Edith Glaser/Werner Thole</i> Space and Spatiality in Educational Research. An introduction | 1 |
| <i>Fabian Kessl</i> Educational Research on Space and Spatiality – A localization of the topic “Space and Spatiality in Educational Research” | 5 |
| <i>Sebastian Dirks/Caroline Fritsche/Maike Lippelt/Christian Reutlinger</i> On the Pedagogical Construction of Urban Spaces between Location and Client – Empirical insights and theoretical conclusions | 20 |
| <i>Sabine Andresen/Susann Fegter/Nora Iranee/Elena Bütow</i> Doing Urban Family: Urban space and parental perspectives on childhood | 34 |
| <i>Nils Berkemeyer/Björn Hermstein/Veronika Manitius</i> On the Way to Spatially Oriented Research on the School System – In what way can spatially aware social theories contribute to empirical analyses of regional educational disparities? | 48 |
| <i>Jeanette Böhme/Viktoria Flasche/Ina Herrmann</i> The Territorialization of the (School-) Pedagogical in the Context of Urban Change – A research contribution by pedagogical morphology | 62 |
| <i>Deutscher Bildungsserver</i> Tips of links relating to the topic of “Space and Spatiality in Educational Research” | 79 |
| Contributions | |
| <i>Marcus Syring/Thorsten Bohl/Marc Kleinknecht/Sebastian Kuntze/ Markus Rehm/Jürgen Schneider</i> Casework as Offer – Case-based Learning as Use: Empirical results on cognitive stress, motivation, and emotions when working with teaching cases | 86 |

| | |
|---|-----|
| <i>Wilfried Plöger/Daniel Scholl/Andreas Seifert</i> | |
| “And yet it moves!” – How specific learning opportunities can promote educational-scientific competencies of student teachers | 109 |
| <i>Andrea Westphal/Anna Gronostaj/Miriam Vock/Rico Emmrich/ Peter Harych</i> | |
| Differentiation in Mathematics and in German Lessons on Grammar School Level – Is it to be found mainly among good diagnosticians and in heterogeneous classes? | 131 |
| Book Reviews | 149 |
| New Books | 156 |
| Impressum | U3 |

Wilfried Plöger/Daniel Scholl/Andreas Seifert

„Und sie bewegt sich doch!“

Wie spezifische Lerngelegenheiten die bildungswissenschaftlichen Kompetenzen von Lehramtsstudierenden fördern können

Zusammenfassung: Derzeit scheint es in der ersten Phase der Lehrerbildung nicht hinreichend zu gelingen, bildungswissenschaftliches Wissen systematisch aufzubauen. Als Schlüssel zur Verbesserung dieser Situation wird weniger die Quantität der jeweiligen Angebote, sondern deren *Qualität* gesehen. Diese wird entscheidend durch die inhaltliche und methodische Gestaltung von *spezifischen Lerngelegenheiten* geprägt. Im vorliegenden Beitrag wird darüber berichtet, wie solche Lerngelegenheiten curricular konzipiert und über eine entsprechende Lehrveranstaltung lernwirksam realisiert werden können. Die Evaluation erfolgt über ein Solomon-Viergruppen-Design. Die sich in der Experimentalgruppe zeigenden signifikanten und hohen Effekte bezüglich der Leistungsentwicklung sprechen für die Bedeutung *spezifischer Lerngelegenheiten* beim Erwerb bildungswissenschaftlichen Wissens und für die Notwendigkeit, weitere Modelle entsprechender Lernangebote zu entwickeln und empirisch zu evaluieren.

Schlagnote: Lehrerbildung, Lerngelegenheiten, Entwicklung bildungswissenschaftlicher Kompetenzen, Unterrichtsqualität, Evaluation

1. Problemaufriss

Kritik am jeweiligen Status quo ist ein wichtiger Motor für soziale und politische Entwicklungsprozesse. Das gilt auch für die Lehrerbildung in Deutschland. Mit Blick auf größere Zeiträume lassen sich die darin erkennbar werdenden Fortschritte unter dem Strich durchaus als Erfolgsgeschichte interpretieren. Nimmt man dagegen kürzere Zeitspannen in Augenschein, kann das Ausmaß der Veränderung – trotz vehementer Kritik – sehr viel geringer ausfallen. Für die letzten 15 bis 20 Jahre scheint dies insbesondere für die Gestaltung und Wirksamkeit des erziehungswissenschaftlichen Anteils¹ im Lehramtsstudium der Fall zu sein. Denn obwohl die für diesen Zeitraum wahrnehmbaren Stimmen unermüdlich gleiche oder zumindest ähnliche Defizite anmahnen, kann der Eindruck entstehen, „dass sich überhaupt nichts ändert!“ (Terhart, 2007, S. 10). Die diesbezügliche Kritik kann hier nicht im Detail wiedergegeben werden, sondern sei exemplarisch auf drei Aspekte reduziert:

1 Wir benutzen die frühere Bezeichnung *erziehungswissenschaftlich* und die heutige Benennung *bildungswissenschaftlich* synonym, weil für diesen Bericht der rückwärtsgewandte kritische Blick auf Lehrerbildung und *zugleich* künftige Fragen der Gestaltung von Lehrerbildung eine zentrale Rolle spielen. Zur Verwendung des Begriffes Bildungswissenschaften s. im Detail Terhart (2012).

- 1) *Wahlfreiheit/Beliebigkeit des Lehrangebotes*: Als Folge der starken Ausdifferenzierung der Erziehungswissenschaft als eigenständige wissenschaftliche Disziplin hat sich auch ihr Lehrangebot stark erweitert. Für Studierende bedeutet dieser Sachverhalt große Wahlfreiheit, für kritische Stimmen ist er eher ein Beleg für die *Beliebigkeit* der Inhalte (Terhart, 2009; Cramer, 2013). Neuere Untersuchungen zu den curricularen Angeboten verschiedener Hochschulen zeichnen weiterhin – zumindest für den gymnasialen Lehramtsbereich – „ein sehr heterogenes Bild bildungswissenschaftlicher Studienangebote und zwar sowohl in struktureller als auch in inhaltlicher Hinsicht“ (Bauer, Diercks, Rösler, Möller & Prenzel, 2012, S. 106).
- 2) *Starke Zersplitterung*: Lehramtsstudierende belegen zwei bis drei Fächer samt zugehöriger Fachdidaktik. Hinzu kommt das erziehungs- bzw. bildungswissenschaftliche Studium, dessen Umfang je nach Bundesland und Schulformbezug des jeweiligen Studiengangs allerdings sehr unterschiedlich ausfällt. Die Koordination dieser Studienelemente ist den Studierenden weitgehend selbst überlassen mit der Folge, dass fachwissenschaftliche Studienanteile Priorität haben und erziehungswissenschaftliche eher als Lückenfüller im individuellen Stundenplan verwertet werden (Lohmann, Seidel & Terhart, 2012).
- 3) *Mangelnder Berufsfeldbezug*: Kontrovers wird die Frage diskutiert, ob die inhaltliche Gestaltung der Lehrangebote in hinreichendem Maße berufsfeldbezogen erfolgt. Diese Diskussion betrifft sowohl die fachwissenschaftlichen und fachdidaktischen Angebote als auch das erziehungswissenschaftliche Studium, wobei von diesem behauptet wird, dass es viel zu wenig auf die spezifischen Anforderungen von Unterricht und Erziehung im schulischen Kontext ausgelegt sei (s. z. B. Wissenschaftsrat, 2001; Nolle, 2004; Czerwenka & Nölle, 2011; Kunina-Habenicht et al., 2013).

Derartige Kritik wird im Wesentlichen in hochschulpolitischen und -didaktischen Diskursen formuliert. Sie müsste allerdings mit den realen Wirkungen konfrontiert werden, insbesondere mit den über objektive Erfassungsmethoden erhobenen tatsächlich erworbenen Kompetenzen angehender Lehrpersonen. Das ist in den letzten zehn Jahren im Rahmen national wie international ausgerichteter Forschungsprogramme zur Lehrerbildung auch geschehen, so etwa in den Projekten COACTIV (*Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz*; Kunter, Baumert et al., 2011), TEDS-M (*Teacher Education and Development Study in Mathematics*; Blömeke, Kaiser & Lehmann, 2010) oder LEK (*Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden*; König & Seifert, 2012). Hinsichtlich der in diesen Studien untersuchten Aspekte des erziehungs- bzw. bildungswissenschaftlichen Wissens zeigte sich durchaus ein entsprechender Zuwachs im Verlauf des Lehramtsstudiums (s. z. B. König & Seifert, 2012), zugleich wurden aber auch große interindividuelle Unterschiede bei der Kompetenzentwicklung erkennbar.

Bei der Suche nach Erklärungen für diese Unterschiede steht die Frage nach der Bedeutung von *Lerngelegenheiten* und deren *individueller Nutzung* im Vordergrund, wobei zwischen informellen und formellen Lerngelegenheiten (s. z. B. Eraut, 2004; Kunter,

Kleickmann, Klusmann & Richter, 2011) differenziert wird. *Informelle* Lerngelegenheiten sind nicht in institutionelle Systeme eingebettet, werden daher kaum durch entsprechende Vorgaben gesteuert und führen in der Regel nicht zu anerkannten Abschlüssen. *Formelle* Lerngelegenheiten stehen in der Verantwortung ausgewiesener Bildungseinrichtungen, sind durch strukturell definierte Lehr-Lernangebote gekennzeichnet und ermöglichen qualifizierende Abschlüsse. Zur Einschätzung der *Wirksamkeit* formeller Lerngelegenheiten unterscheiden Kunina-Habenicht et al. (2013) *drei Ebenen*: (1) generelle Wirksamkeit des Lehramtsstudiums, (2) Studienstrukturen, die zum einen die Art des Studiengangs (grundständig vs. modularisiert) und zum anderen die schulform- bzw. schulstufenspezifischen Lehrangebote bestimmen, und (3) Auswirkungen von institutionellen Unterschieden.

Empirische Belege für die Wirksamkeit solcher Lerngelegenheiten sind derzeit noch rar (Lohmann et al., 2012). Für die generelle Wirksamkeit des Lehramtsstudiums (Ebene 1) spricht etwa der Vergleich von Lehramtsabsolvent(inn)en mit Seiten- bzw. Quereinsteiger(inne)n, wobei Letztere in der Regel keine oder nur wenige erziehungswissenschaftliche Veranstaltungen belegt und daher auch erwartungsgemäß entsprechend geringeres Wissen vorzuweisen haben (s. z.B. Kleickmann & Anders, 2011). Auch für das Lehramtsstudium selbst lässt sich im Verlauf eine Zunahme des Wissens feststellen (König & Seifert, 2012).

Mit Blick auf Ebene 2 zeigt z.B. die LEK-Studie, dass Studierende in grundständigen wie modularisierten Studiengängen eine ähnliche Entwicklung hinsichtlich des erziehungswissenschaftlichen Wissens nehmen (König & Seifert, 2012). Nach den Ergebnissen der TEDS-M-Studie (Blömeke et al., 2010) erreichen Studierende des Primarstufenlehramts die höchsten Leistungen im pädagogischen Wissen, gefolgt von Studierenden des Lehramts für Sekundarstufe II und – weiter abgeschlagen – des Lehramts für die Sekundarstufe I.

Einige dieser Befunde werden neuerdings durch die *BilWiss-Studie* bestätigt (Kunina-Habenicht et al., 2013), mit der in NRW das bildungswissenschaftliche Wissen von Lehramtsabsolvent(inn)en zu Beginn des Referendariats (Vollerhebung mit N = 3298) ermittelt wurde. Darin zeigen allerdings die Absolvent(inn)en des gymnasialen Lehramtsstudiums bessere Testleistungen als die nicht-gymnasialer Studiengänge. Darüber hinaus liefert die BilWiss-Studie wertvolle Informationen zum Einfluss der Institutionen (Ebene 3). Negativ fällt dabei vor allem der Sachverhalt auf, dass „weniger als 2% der Varianz im bildungswissenschaftlichen Wissen der Absolvent(inn)en auf Unterschiede zwischen den Universitäten zurückgeführt werden kann“ (Kunina-Habenicht et al., 2013, S. 16).

Dieses Faktum war nicht mit der individuellen Nutzung der Lerngelegenheiten zu erklären, weil sich keine signifikanten Zusammenhänge zwischen der *Anzahl* belegter Veranstaltungen und den Testleistungen zeigten. Deshalb hegen die Autor(inn)en die Vermutung, dass weniger die Quantität, sondern die *Qualität* der Lehrangebote eine entscheidende Rolle spielen könnte (S. 18).

Zur Berücksichtigung dieses *Qualitätsaspekts* müsste man das Drei-Ebenen-Modell dann allerdings um eine *vierte* Ebene erweitern, die für das Angebot *spezifischer Lern-*

gelegenheiten steht. Bisher gibt es jedoch nur wenige empirische Belege für die Bedeutung solcher Lerngelegenheiten. König und Blömeke (2012) konnten z. B. zeigen, dass hinreichende Gelegenheiten zu *eigenständigem Unterrichten* in der zweiten Ausbildungsphase, gekoppelt mit entsprechender mentorieller Unterstützung, die Entwicklung des pädagogischen Wissens positiv beeinflussen. Hinsichtlich der in Veranstaltungen des Studiums erfahrenen *Lehrmethoden* – eine weitere Ausprägung *spezifischer Lerngelegenheiten* – ergeben sich dagegen keine nennenswerten Effekte (Blömeke & König, 2011; König & Seifert, 2012).

An diesem Forschungsdesiderat setzt der vorliegende Beitrag an, indem exemplarisch gezeigt wird, wie solche spezifischen Lerngelegenheiten konzipiert und empirisch evaluiert werden können. Von einer Definition des Begriffes *spezifische Lerngelegenheiten* ausgehend wird das Curriculum einer Lehrveranstaltung² erläutert, in deren Rahmen Lehramtsstudierende *didaktisch-methodisches Wissen* erwerben und zur Analyse von Unterricht *anwenden* sollen. Die Evaluation des Konzepts selbst erfolgte über ein Solomon-Viergruppensdesign. Im Ergebnis zeigen sich in der Experimentalgruppe *signifikante Effekte*, die für die *Bedeutung spezifischer Lerngelegenheiten* sprechen.

2. Was sind spezifische Lerngelegenheiten?

Unter spezifischen Lerngelegenheiten werden hier Lehr-Lernarrangements verstanden, die *mindestens* diesen Kriterien genügen sollten:

- 1) Spezifische Lerngelegenheiten werden durch Lehrveranstaltungen realisiert, die hinsichtlich ihrer curricularen Inhalte und ihrer methodischen Realisierung klar umrissen sind. Solche Konzeptionen bilden im Sinne McDonnells (1995) das jeweils *implementierte Curriculum*, das im günstigsten Fall verschriftlicht vorliegt. Je eindeutiger die normativ-curricularen Aspekte beschrieben sind, desto eher ist ihre unterstellte Wirkung (das *erreichte Curriculum*) empirisch überprüfbar.
- 2) Im Zentrum spezifischer Lerngelegenheiten sollte (im Rahmen der ersten Phase der Lehrerbildung) der „Erwerb konzeptuell-analytischer Kompetenz“ (Expertenkommission, 2007, S. 17) stehen. Mit der Aneignung des entsprechenden theoretischen Wissens sollte für die Lernenden aber zugleich ein deutlicher *Bezug zum Berufsfeld* erkennbar sein.
- 3) Für die empirische Überprüfung der Wirksamkeit spezifischer Lerngelegenheiten sind „Standards“³ eine notwendige Voraussetzung (Terhart, 2007). Das könnten z. B.

2 Dieses hochschuldidaktische Konzept wurde im Rahmen des von der DFG geförderten Projekts „Analysekompetenz von Lehrpersonen – Diagnose und Entwicklung“ (Kennzeichen: PL 272/3-1) realisiert.

3 Wir verzichten hier auf eine Definition der Begriffe *Standards* und *Kompetenzen*, weil wir annehmen, dass auch heute noch das Urteil Terharts über die Entwicklung von Standards zutrifft: „Die gegenwärtige Forschungssituation zur Lehrarbeit und zur Lehrerbildung erlaubt es nicht, alle Dimensionen, Kompetenzebenen, Skalierungen vollständig auszuarbei-

die von der KMK formulierten Standards für Bildungswissenschaften sein (2004). Als Alternative könnten aber auch die im Rahmen des BilWiss-Projekts über eine Delphi-Studie (Kunina-Habenicht et al., 2012) ermittelten Themen dienen, die für den *Konsens* von Expert(inn)en der ersten und zweiten Phase der Lehrerbildung (N = 36 bzw. 13) über insgesamt neun Themenbereiche des bildungswissenschaftlichen Studiums stehen. Für den speziellen Bereich „Unterricht“, um den es im hier zu beschreibenden Projekt im Wesentlichen geht, werden dort 34 als relevant erachtete Themen benannt. Schließlich könnte sich die inhaltliche Gestaltung spezifischer Lernangebote auch an jenen Wissensfacetten orientieren, die im Rahmen *standardisierter* Testverfahren (COACTIV, TEDS-M, LEK) zur Erfassung professioneller Lehrer(innen)kompetenzen eine zentrale Rolle spielen. Diese drei Optionen stellen „Standards“ in unterschiedlichen Kontexten dar, inhaltlich bilden sie aber eine große Schnittmenge wichtiger Themen der Lehrerbildung.

- 4) In der Festlegung und Realisierung spezifischer Lerngelegenheiten spiegeln sich die individuellen Schwerpunktsetzungen von Dozent(inn)en, die im Hinblick auf die besonderen Ausbildungsumstände vor Ort zweifelsohne notwendig sind. Darüber hinaus sollte aber auch durch Urteile mehrerer Expert(inn)en gewährleistet werden, dass die Transformation von „Standards“ in spezifische Lerngelegenheiten zu einer akzeptablen *inhaltlichen und methodischen* Qualität der Lehrveranstaltungen führt. Im Idealfall wären die Urteile der Expert(inn)en kompatibel mit der durch Studierende wahrgenommenen Qualität und Akzeptanz der Lehrveranstaltung (Kunina-Habenicht et al., 2012, S. 654).

3. Theoretische Rahmung und Begründung des Curriculums

3.1 Didaktisch-methodisches Wissen – Merkmale von Unterrichtsqualität

Für das von uns konzipierte und evaluierte Curriculum ist mit Bezug auf gegenwärtige Projekte zur Lehrerbildungsforschung (z. B. COACTIV, TEDS-M, LEK) die Annahme leitend, dass die *Wissensbasis* von Lehrpersonen einen *zentralen Faktor* in der Entwicklung professioneller Kompetenzen darstellt, Wissen also – neben Überzeugungen und motivational-selbstregulativen Merkmalen – „die Grundlage für effektives Handeln von Lehrkräften“ (Baumert & Kunter, 2011, S. 46) bildet. Professionelles Wissen wird in diesen Projekten durch die Unterscheidung von pädagogisch-psychologischem Wissen, Fachwissen und fachdidaktischem Wissen definiert (s. S. 34).

Die Konzeption des Curriculums bezieht sich auf das (fächerübergreifende) pädagogisch-psychologische Wissen. Da das Curriculum in Form einer einsemestrigen Vorlesung (13 Termine á 1.5 Stunden), also in einem engen Zeitfenster, realisiert wird, kann

ten“ (Terhart, 2005, S. 278). Die folgenden Überlegungen in Abschnitt 3 (theoretischer Hintergrund) gehen darauf ein, inwiefern die Erziehungswissenschaft „auf dem Weg“ zur Etablierung von Standards ist und wie wir unser Evaluationsprojekt in dieser Tendenz verorten.

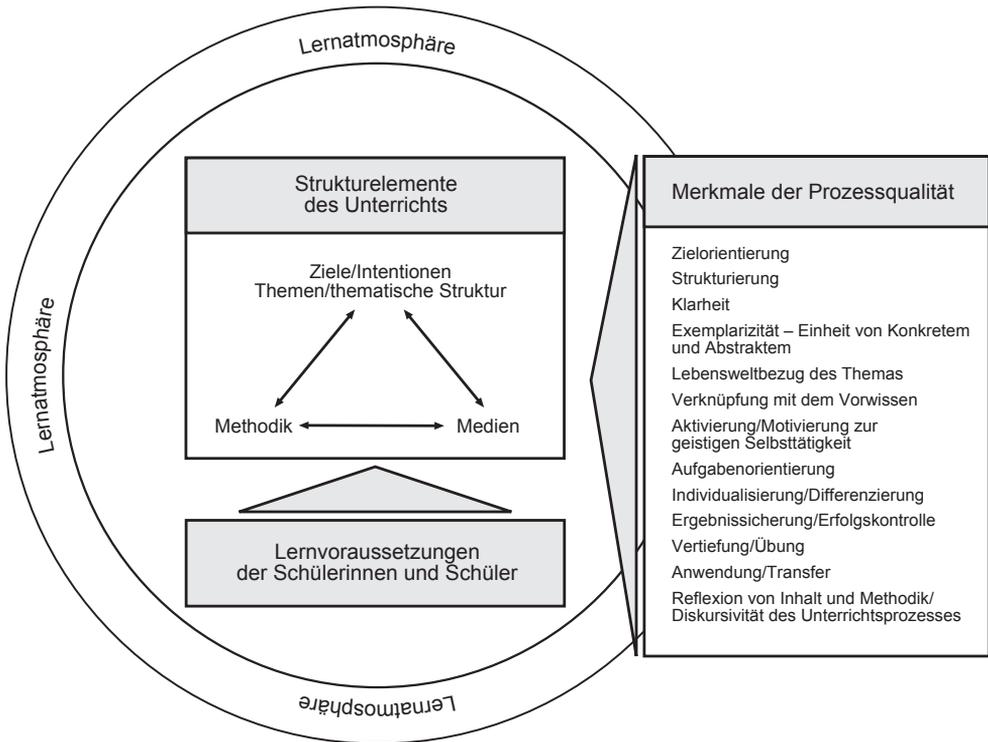


Abb. 1: Struktur- und Prozessmerkmale der Unterrichtsqualität

dieser Wissensbereich nicht umfassend abgedeckt werden. Deshalb ist die grundlegende Intention auf den *Erwerb* von *didaktisch-methodischem* Wissen, das eine wichtige *Teilmenge* des pädagogisch-psychologischen Wissens bildet, und auf die *Anwendung* dieses Wissens zur Analyse videografierten Unterrichts beschränkt.

Dieses didaktisch-methodische Wissen definieren wir über *Struktur- und Prozessmerkmale* der Unterrichtsqualität. In Abbildung 1 sind die Elemente der Strukturqualität auf der linken Seite, die Merkmale der Prozessqualität auf der rechten Seite dargestellt. Im Einzelnen ist darunter zu verstehen:

Strukturqualität

Wenn man zunächst die beiden großen didaktischen Traditionen, die bildungstheoretische und die lerntheoretische Didaktik, miteinander vergleicht, lässt sich – trotz aller Unterschiedlichkeit im Detail – ein Konsens hinsichtlich der *Struktur* von Unterricht ausmachen: Unterricht stellt sich danach als *Zusammenhang* dar (s. z. B. Heimann, 1976; Klafki, 1996), der durch mindestens diese vier Strukturmerkmale charakterisiert ist: (1) *Ziele/Intentionen/Themen/thematische Struktur*, (2) *Methodik*, (3) *Medien* und (4) *Lernvoraussetzungen der Schüler(innen)*. Sie sind im linken Teil der Abbildung an-

geführt und spiegeln das formale Gerüst wider, in dessen Rahmen sich inhaltlich die Planung und somit die *Antizipation* von Unterricht bewegt. Die konsistente Abstimmung dieser Merkmale sichert die *Strukturqualität* von Unterricht.

Prozessqualität

Die in der Planung konstruierte und antizipierte Struktur von Unterricht bildet aber zunächst nur einen *statischen* Rahmen, der für die *Realisierung* von Unterricht in einen *prozessualen* Fluss übersetzt werden muss. Während in der Planungsphase Lernwirksamkeit zunächst nur hypothetisch unterstellt werden kann, muss sie im Unterrichtsgeschehen „hergestellt“ und von der Lehrperson ständig überprüft werden. Deshalb bedarf es auch für dieses *Prozessgeschehen* der Festlegung von Qualitätsmerkmalen (rechte Seite der Abbildung). Im Einzelnen zählen dazu (vgl. im Überblick Helmke, 2009): Zielorientierung, Strukturierung, Klarheit, Exemplarizität (Einheit von Konkretem und Abstraktem), Lebensweltbezug des Themas, Verknüpfung mit dem Vorwissen, Aktivierung/Motivierung zur geistigen Selbsttätigkeit, Aufgabenorientierung, Individualisierung/Differenzierung, Ergebnissicherung/Erfolgskontrolle, Vertiefung/Übung, Anwendung/Transfer, Diskursivität als Reflexion von Inhalt und Methode.

Diese Struktur- und Prozessmerkmale stellen „Standards“ dar, die sich aus unterschiedlichen Blickwinkeln legitimieren lassen: (1) Im Sinne der KMK-Vereinbarung (2004) beziehen sich diese Merkmale auf die Kernaufgabe von Lehrer(inne)n, nämlich auf „die gezielte und nach wissenschaftlichen Erkenntnissen gestaltete Planung, Organisation und Reflexion von Lehr- und Lernprozessen sowie ihre individuelle Bewertung und systemische Evaluation“ (KMK, 2004, S. 3). (2) Sie lassen sich darüber hinaus auch den über die BilWiss-Delphi-Studie ermittelten Standardthemen für den Bereich „Unterricht“ zuordnen (genannt werden dort z. B. Transparenz von Zielen und Anforderungen, Lernvoraussetzungen, Motivation, Rückmeldungen und Beurteilung, Transfer/Aufgabenformate usw. [Kunina-Habenicht et al., 2012, S. 674]). (3) Und schließlich stellen diese Struktur- und Prozessmerkmale eine große Schnittmenge mit den Konzeptualisierungen pädagogischen Wissens im Rahmen *standardisierter* Testverfahren zur Erfassung professioneller Lehrer(innen)kompetenzen dar. In TEDS-M werden (als Dimensionen bzw. Inhaltsgebiete) u. a. Strukturierung, Methodenvielfalt, Motivation, Differenzierungsmaßnahmen und Leistungsbeurteilung als Merkmale berücksichtigt (Blömeke & König, 2010). Diese Merkmale sind auch für die von Baumert und Köller im Rahmen der TIMSS-Studie verwendete Definition von „verständnisvollem Lernen“ (2000, S. 273–274) konstitutiv. Darüber hinaus wird in dieser Studie auch den Merkmalen Vorwissen, Lebensweltbezug, multiple Repräsentation (Medieneinsatz) und Anwendung des Gelernten große Bedeutung zugemessen. In COACTIV werden zudem die „zieladäquate Orchestrierung“ (Voss & Kunter, 2011, S. 195) von Unterricht, die Beachtung individueller Lernvoraussetzungen und die Funktion der Rückmeldung genannt (S. 196 ff.).

3.2 *Das implementierte Curriculum der Veranstaltung als spezifische Lerngelegenheit*

Das übergreifende *Ziel* der Veranstaltung liegt nicht allein in der Aneignung theoretisch-konzeptuellen Wissens (hier didaktisch-methodisches Wissen in Form von Struktur- und Prozessmerkmalen), sondern geht einen entscheidenden Schritt darüber hinaus. Damit solches Wissen kein „träges Wissen“ (Renkl, 1996) bleibt, wird es nämlich in Form *situierter Lernens* über die Analyse videografierter Unterrichtsszenen auf Praxiskontexte angewandt. Durch diese Lerngelegenheiten soll die Kluft zwischen Wissen und Handeln verringert und ein engerer Bezug zwischen Theorie und Praxis ermöglicht werden.

Jede der 13 Einheiten der Veranstaltung ist in *fünf Phasen* strukturiert: (1) *theoretische Explikation* der betreffenden Struktur- und Prozessmerkmale (didaktisch-methodisches Wissen); (2) Fokussierung auf wesentliche *Indikatoren* zur Beobachtung dieser Merkmale; (3) aktive *Anwendung* des erworbenen Wissens durch Analyse eines videografierten Unterrichtsausschnittes; (4) Sicherung und Diskussion der *Erkenntnisse*; (5) Rückbindung und Abgleich mit Erkenntnissen aus den vorangegangenen Einheiten im Sinne *kumulativen Lernens* und *immanenter Wiederholung*. Offiziell wird die Veranstaltung unter dem Label „Beobachtung und Analyse von Unterricht“ als *Vorlesung* geführt, faktisch stellt sie von der Sozialform her aber eine Mischung zwischen direkter Instruktion (Phase 1 und 2), Einzelarbeit (Phase 3) und Plenumsgespräch (Phase 4 und 5) dar.

3.3 *Zur Validität des Curriculums*

Die didaktisch-methodische Gestaltung der 13 Einheiten ist detailliert in einem gut sechzig Seiten umfassenden Papier dokumentiert (Plöger, 2012). Um die *curriculare Validität* des Konzeptes zu sichern, haben wir in Entsprechung zu dem an angloamerikanischen Universitäten genutzten Verfahren des *peer observation* (s. Spinath & Stehle, 2011) insgesamt zwölf Kolleg(inn)en um dessen Beurteilung gebeten. Neun dieser Personen haben eine Professur an einer Universität inne, eine Person lehrt als Privatdozent. Sie vertreten die Gebiete Empirische Bildungsforschung (2), Allgemeine Didaktik/Schulpädagogik (5), Fachdidaktik Physik (1), Didaktik der Musik (1), Fachdidaktik Religionslehre (1). Vier dieser zehn Kolleg(inn)en lehren an der Universität zu Köln, bei allen anderen handelt es sich um externe Expert(inn)en. Mit zwei weiteren Personen, die als Studienseminarleiter tätig sind, wurde Expertise von Vertretern der zweiten Ausbildungsphase eingeholt.

Die Beurteilung des Lehrkonzepts erfolgte über die Beantwortung von 11 Fragen auf einer sechsstufigen Likertskala (1 = trifft überhaupt nicht zu; 6 = trifft voll und ganz zu). Tabelle 1 zeigt die Ergebnisse des Expert(inn)enratings in exemplarischer Auswahl anhand von sechs Items. Der Mittelwert für zehn Fragen liegt über 5 (S_D zwischen 0.492 und 1.240), lediglich bei einer Frage wird ein M von 4.67 ($S_D = 0.651$) erreicht. Auf-

| Item | | M | S _D |
|------|--|------|----------------|
| 1 | Der <i>Umfang</i> der gebotenen Information ist hinreichend, um sich ein klares Bild von der Vorlesungskonzeption machen zu können. | 5.50 | 1.168 |
| 3 | Die Konzeption ist in sich <i>schlüssig</i> und inhaltlich <i>kohärent</i> . | 5.75 | .622 |
| 5 | Das für die Lehrerbildung konstitutive Verhältnis von <i>Theorie</i> und <i>Praxis</i> ist über die <i>theoretische</i> Einführung der Struktur- und Prozessmerkmale und deren <i>praktische</i> Einübung anhand von videografiertem Unterricht angemessen berücksichtigt. | 5.33 | .651 |
| 6 | Die Konzeption und Gestaltung der Vorlesung dürfte bei den Lehramtsstudierenden auf <i>hohe Akzeptanz</i> stoßen. | 5.33 | .778 |
| 9 | Die Vorlesung ist als <i>Vorbereitung auf Praktika</i> und das zu absolvierende <i>Praxissemester</i> gut geeignet. | 5.75 | .452 |
| 11 | Diesem Curriculum kann eine hohe <i>Lernwirksamkeit</i> hinsichtlich der Entwicklung der Analysekompetenz Studierender unterstellt werden. | 5.58 | .515 |

Tab. 1: Beurteilung des Curriculumkonzepts durch Expert(inn)en (N = 12) auf einer sechsstufigen Likertskala (1 = trifft überhaupt nicht zu; 6 = trifft voll und ganz zu)

grund dieser positiven Bewertung sehen wir die *curriculare Validität* des Vorlesungskonzepts als gegeben an.

Über diese Beurteilung der inhaltlichen und methodischen Qualität des Curriculums durch Expert(inn)en hinaus erschien uns aber auch die Perspektive der Studierenden wichtig. Denn eine entscheidende Bedingung für das Zustandekommen entsprechender Lernzuwächse liegt in der *Akzeptanz* und *wahrgenommenen* Qualität von Lehrveranstaltungen durch die Betroffenen (Spiel, Gradinger & Lüftenegger, 2010). Um diese zu ermitteln, wurden Studierende aus einer der Experimentalgruppen (N = 82) um eine abschließende Evaluation gebeten, die auf Wunsch der Teilnehmer(innen) anonym erfolgte. Die dazu genutzten 11 Items (s. die Auswahl in Tabelle 2) sind – soweit mög-

| Item | | M | S _D |
|------|---|------|----------------|
| 2 | Der „rote Faden“ der Lehrveranstaltung ist gut erkennbar. | 5.18 | .848 |
| 4 | Die Lehrveranstaltung hilft mir beim <i>Aufbau der notwendigen Fachsprache</i> . | 4.80 | .923 |
| 5 | Das Verhältnis von <i>Theorie</i> und <i>Praxis</i> ist über die <i>theoretische</i> Einführung der Struktur- und Prozessmerkmale und deren <i>praktische</i> Einübung anhand von videografiertem Unterricht angemessen berücksichtigt. | 5.27 | .808 |
| 6 | Die Lehrveranstaltung folgt selbst dem Anspruch <i>guten Unterrichts</i> . | 4.78 | .864 |
| 8 | Die Lehrveranstaltung hat hohe Relevanz für meine Berufsperspektive. | 5.48 | .921 |
| 10 | Die Lehrveranstaltung stellt hohe <i>Anforderungen</i> an die Studierenden. | 3.74 | 1.037 |

Tab. 2: Beurteilung der Veranstaltung durch Studierende der Experimentalgruppe A4 (N = 82) auf einer sechsstufigen Likertskala (1 = trifft überhaupt nicht zu; 6 = trifft voll und ganz zu)

lich – in Entsprechung zu den Items des Expert(inn)enratings formuliert. Die sich ergebenden Abweichungen in den Itemformulierungen resultieren aus den unterschiedlichen Perspektiven (Studierende: Akzeptanz bzw. wahrgenommene Qualität – Expert[inn]en: Beurteilung der inhaltlichen und methodischen Qualität und der vermuteten Wirksamkeit der spezifischen Lerngelegenheiten).

Die Studierendenurteile bewegen sich im Mittel um den Wert 5, sodass durchgehend eine positive Bewertung des Lehrangebotes vorliegt. Lediglich der Wert für Item 10 („Die Lehrveranstaltung stellt hohe Anforderungen an die Studierenden“) weicht mit 3.74 etwas nach unten ab (zugleich höchste S_D mit 1.037). Plausible Erklärungen für dieses Urteil ließen sich aufgrund mangelnder Informationen nicht ausmachen.

4. Methode

4.1 Das quasi-experimentelle Design

Um die Wirksamkeit der den Studierenden gebotenen *spezifischen Lerngelegenheiten* zu überprüfen, wurde ein quasi-experimentelles Design in Form eines Solomon-Viergruppenplans realisiert, das zwei Experimental- und zwei Kontrollgruppen vorsieht. Im Rahmen dieses Designs stellt der Besuch der Veranstaltung „Beobachtung und Analyse von Unterricht“ das *Treatment* dar, die daran teilnehmenden Studierenden bilden also die *Experimentalgruppen*.

Bei den beiden *Kontrollgruppen* handelt es sich um Studierende, die an einer Veranstaltung zum Thema *Unterrichtsqualität* teilnahmen (Seminarform; 13 Termine á 1.5 Stunden). In dieser Veranstaltung wurden *empirische Studien* zur Frage der Unterrichtsqualität besprochen (z. B. die Scholastikstudie, die IPN-Videostudie usw.) sowie Texte erarbeitet, in denen Modelle und Merkmale von Unterrichtsqualität *theoretisch* beschrieben und begründet werden.

Im *Unterschied zur Vorlesung* (Treatment) war für die Veranstaltung der Kontrollgruppen aber nicht das spezifische Modell der Struktur- und Prozessmerkmale der Unterrichtsqualität (s. Abb. 1) leitend; es wurden zudem keine Beobachtungsindikatoren für die einzelnen Merkmale entwickelt und insbesondere *keine videografierten Unterrichtsszenen* unter dem Gesichtspunkt der Unterrichtsqualität analysiert. Aufgrund dessen hatten die Studierenden dieser Veranstaltung auch keine Gelegenheit, erworbenes Wissen in praktischen Zusammenhängen anzuwenden und einzuüben. Hierin sehen wir den zentralen Unterschied zur Veranstaltung der Experimentalgruppen. Darüber hinaus deckte die Auswahl besprochener Texte die oben beschriebenen Aspekte von Unterrichtsqualität nicht vollständig ab. So wurden z. B. nicht die Merkmale *Zielorientierung*, *Anwendung und Transfer*, *Verknüpfung mit dem Vorwissen* und *Medieneinsatz* thematisiert.

Zur Überprüfung der Wirksamkeit des Treatments wird in den Experimental- wie Kontrollgruppen am Ende des Semesters ein *Posttest* durchgeführt, mit dem die *Analysekompetenz* der Studierenden festgestellt wird. Der Test misst also das angeeignete

theoretische Wissen (Struktur- und Prozessmerkmale von Unterricht) nicht unmittelbar und „steril“, sondern handlungsnah über die Analyse videografierten Unterrichts (s. Voss, Kunter & Baumert, 2011), durch die die Teilnehmer(innen) zeigen können, ob sie ihr didaktisch-methodisches Wissen auch berufsfeld- bzw. kontextbezogen einsetzen können. Da sich die Wirksamkeit des Treatments nicht allein aus den Posttestleistungen, sondern nur im Vergleich dieser Leistungen zum anfänglich existierenden *Vorwissen* ermitteln lässt, wird dieses mittels *Pretest* gemessen.

Ein Pretest kann jedoch (Lern-)Effekte erzeugen oder dazu führen, dass Personen anders auf das Treatment reagieren, als wenn sie nicht am Pretest teilgenommen hätten (s. Bortz & Döring, 2006, S. 538–539). Um nun die Wirkung solcher Effekte kontrollieren zu können, wird der Pretest nicht in allen vier Gruppen durchgeführt, sondern jeweils nur in *einer* der beiden Experimental- und Kontrollgruppen. Dieses Vorgehen hat dann allerdings zur Konsequenz, dass man über das Vorwissen der jeweils anderen Experimental- und Kontrollgruppe, die nicht am Vortest teilnimmt, keine Aussagen machen kann.

4.2 Stichprobe

Bei den Teilnehmer(inne)n handelt es sich ausschließlich um Studierende des Lehramts Gymnasium/Gesamtschule. Die Zuteilung der Teilnehmer(innen) zu den einzelnen Gruppen erfolgte über ein universitätsinternes Einwahlverfahren, über das sich die Studierenden für eine angebotene Veranstaltung zunächst bewerben; die tatsächliche Zulassung wird dann durch das Zuteilungssystem nach vorgegebenen Kriterien geregelt. Tabelle 3 zeigt die beiden Gruppen der teilnehmenden Studierenden.

- 1) Die Veranstaltung „Beobachtung und Analyse von Unterricht“ wurde in vier aufeinanderfolgenden Semestern angeboten. Auf diese Weise konnte das Treatment in vier Kohorten (A1 bis A4) erprobt werden. Zusammen bilden sie mit $N = 275$ die *Experimentalgruppe A*. In allen vier Gruppen kam am Ende des Semesters ein bewährter Test (s. unten 4.3) als Posttest zum Einsatz. In A4 wurde er zusätzlich als Pretest verwendet.

| Gruppe | N | Zeitpunkt | studierte Semester | Anteil weiblich (%) | Bedingung | | |
|--------|----|------------|--------------------|---------------------|-----------|-----------|----------|
| A1 | 60 | SoSe 2012 | 8.51 | k.A. | – | Treatment | Posttest |
| A2 | 52 | WS 2012/13 | 7.69 | 69.2 | – | Treatment | Posttest |
| A3 | 81 | SoSe 2013 | 8.67 | 59.5 | – | Treatment | Posttest |
| A4 | 82 | WS 2103/14 | 8.61 | 69.0 | Pretest | Treatment | Posttest |
| B1 | 42 | SoSe 2014 | 9.19 | 57.1 | Pretest | – | Posttest |
| B2 | 42 | SoSe 2014 | 6.48 | 85.7 | – | – | Posttest |

Tab. 3: Zusammensetzung der Stichprobe: Experimentalgruppen (A1 bis A4) und Kontrollgruppen (B1 und B2)

- 2) Die Veranstaltung „Unterrichtsqualität“ wurde in einem Semester parallel mit zwei Gruppen (B1 und B2) durchgeführt; sie bilden die *Kontrollgruppe B* (N = 84 Studierende insgesamt). Sowohl in B1 als auch in B2 wurde am Ende des Semesters der Posttest, in B1 anfangs auch der Pretest durchgeführt.

In den Gruppen differiert weder die mittlere Semesteranzahl signifikant ($F[1.359] = 2.607, p > .05$) noch das Verhältnis von Männern und Frauen ($\chi^2[1] = .936, p > .05$). Somit erübrigt sich die statistische Kontrolle dieser (potenziellen) Störvariablen.

4.3 Testinstrument und -durchführung

Der als Pre- und Posttest eingesetzte Test ist im Rahmen eines vorangegangenen Projekts⁴ an einer umfangreichen Population (N = 800; je 200 Studierende, Referendarinnen und Referendare, Lehrpersonen, Fach- bzw. Seminarleiterinnen und -leiter) erprobt worden. Über die Modellierung der Analysekompetenz, ihre Messung und die Itemskala (inhaltliche Validität, Raschhomogenität, Reliabilität, Intercoderreliabilität bei der Auswertung der Itemantworten usw.) ist an anderer Stelle ausführlich berichtet worden (s. Plöger & Scholl, 2014).

Der Test umfasst insgesamt 21 Items, deren inhaltliche Formulierungen die genannten Struktur- und Prozessmerkmale und somit das didaktisch-methodische Wissen abbilden. Auf eine ausführlichere Erläuterung der Items (s. Plöger, Scholl & Seifert, 2015) muss hier verzichtet werden. Stattdessen seien zwei beispielhaft angeführt; was dabei jeweils als „richtig“ gilt, bemisst sich an einem als Referenz herangezogenen Expert(inn)enurteil.

„Zu Beginn der Unterrichtsstunde lässt der Lehrer Beispiele für Lichtbrechung nennen und führt dann selbst weitere an (mittels bildlicher Darstellung). Wie beurteilen Sie diesen Einstieg?“

Dieses Item steht für das Prozessmerkmal *Exemplarizität (Einheit von Konkretem und Abstraktem)*. Im Unterricht dargebotene Beispiele müssen die wesentlichen Elemente des zu erlernenden Inhalts repräsentieren. Ist diese Forderung eingelöst, besteht *Strukturisomorphie* zwischen konkretem Beispiel und zu erarbeitender abstrakter kognitiver Struktur. Die in der Eingangsphase der Unterrichtsstunde eingesetzten Beispiele erfüllen aber genau diese Funktion nicht.

„Verschafft sich der Lehrer einen Überblick darüber, ob seine Schülerinnen und Schüler dem Unterricht folgen können, oder tut er das nicht?“

Dieses Item bezieht sich auf das Prozessmerkmal *Ergebnissicherung/Erfolgskontrolle*, das im zu analysierenden Unterricht unzureichend eingelöst ist, weil die Lehrperson keinerlei (Kontroll- bzw. Rück-)Fragen stellt, wichtige Sachverhalte nicht (noch ein-

4 Auch dieses Vorgängerprojekt wurde von der DFG (Kennzeichen: PL 272/2-1) gefördert.

mal) von den Schülern erklären bzw. wiederholen lässt, keine Zusammenfassungen von Teilergebnissen abrufen usw.

Die Durchführung des Verfahrens erfolgt so, dass den Proband(inn)en eine videografierte Unterrichtsstunde vorgeführt und der Test in Form eines Fragebogens vorgelegt wird, dessen Beantwortung die *Analyse* der Unterrichtsstunde darstellt. Das Video zeigt eine Physikstunde (Bereich Optik), in der das Snelliussche Brechungsgesetz erarbeitet werden soll. Sie ist durch typische Schwächen von Physikunterricht geprägt, wie sie etwa durch die IPN-Videostudie (Seidel et al., 2006) charakterisiert wurden (geringe Zielorientierung, fehlende Einbettung von Experimenten in einen hypothesengeleiteten Prozess, starke Lehrerzentrierung, reproduktive Gesprächsführung).

4.4 IRT-Skalierung – Testauswertung

Der Test, welcher als Kriterium dient, wurde mithilfe der Methode der virtuellen Fälle⁵ mit der Software *Conquest* (Wu, Adams & Wilson, 1997) IRT-skaliert. Durch das Partial-Credit-Design waren für die 21 Items insgesamt 65 Parameter zu schätzen. Um für diese Schätzungen eine hinreichend große Skalierungsstichprobe zu generieren, wurden die Daten aus dem vorangegangenen Projekt (N = 800) zusammen mit den Daten aus diesem Projekt (N = 359) zur Skalierung genutzt. Für die Analyse der hier berichteten Ergebnisse wurden dann wiederum nur die in das quasi-experimentelle Design einbezogenen Fälle genutzt.

Die EAP-PV-Reliabilität ist mit .854 als gut zu bezeichnen, allerdings wäre eine höhere Varianz des Gesamtkonstrukts (.526) wünschenswert. Die gewichteten MNSQs der Items liegen in den Grenzen von .83 und 1.11. Lediglich ein Item fällt mit einem MNSQ von 1.25 etwas aus dem Rahmen. Die Trennschärfen (Discrimination Index) liegen in einem Bereich von .25 und .70 und sind somit ebenfalls akzeptabel.

Über die IRT-Skalierung werden die Testleistungen der Probanden in Form von Personen-WLE-Scores extrahiert. Sie bilden als *abhängige Variable* die Grundlage für die unten angeführten Berechnungen und Ergebnisse. Darin werden allerdings keine individuellen WLE-Scores, sondern WLE-Mittelwerte für die jeweilige Gruppe angeführt. Diese WLE-Scores wurden zwecks besserer Lesbarkeit in eine Variable mit dem Mittelwert 50 und einer Standardabweichung von 10 transformiert.

5 Bei der Methode der virtuellen Fälle werden die getesteten Personen unabhängig von ihrer Teilnahme an einem der beiden oder an beiden Messzeitpunkten einbezogen, wobei echte Längsschnittprobanden auch zwei Fälle in der Skalierungsdatei darstellen. Dieses Verfahren zur Skalierung bietet den Vorteil, dass durch die simultane Nutzung aller Fälle dem Problem der Einschränkung der Varianz, wie sie bei der Skalierung zu einem Messzeitpunkt vorliegen kann, begegnet wird. Zudem lassen sich die Daten der beiden Messzeitpunkte im Sinne einer einheitlichen Skala aufeinander beziehen (vgl. Hartig & Kühnbach, 2005).

4.5 Hypothesen

Für die Überprüfung der Wirksamkeit der gebotenen spezifischen Lerngelegenheiten sind fünf Hypothesen leitend:

H1: Insbesondere durch das aktive Anwenden theoretischer Inhalte auf videografierte Unterrichtssituationen während der gesamten Lehrveranstaltung nehmen wir an, dass sich das Ausmaß der Analysekompetenz der Studierenden vom Beginn bis zum Ende der Veranstaltung erhöht: *In der Treatmentgruppe A4 ergibt sich vom Pretest zum Posttest ein signifikanter Leistungszuwachs.*

H2: Im Vergleich von Treatment- und Kontrollgruppe erwarten wir, dass der *Anstieg* der Analysekompetenz über die Zeit in der Kontrollgruppe wesentlich geringer ist als in der Treatmentgruppe. *Die Leistungsentwicklung in der Treatmentgruppe A4 ist signifikant größer als in der Kontrollgruppe B1.*

H3: Hypothese 3 und 4 dienen der Kontrolle von Testeffekten. Hierfür wird zunächst postuliert, dass aufgrund der unterschiedlichen Leistungsentwicklung, die schon in H1 und H2 thematisiert wurde, die Leistungen im Posttest in *allen* Treatmentgruppen höher sind als in den Kontrollgruppen. Da keine *Leistungsentwicklung* getestet wird, können auch Gruppen zur Analyse genutzt werden, für die nur ein Posttest-Ergebnis vorliegt: *Die Treatmentgruppen A1 bis A4 unterscheiden sich im Posttest signifikant von den Kontrollgruppen B1 und B2.*

H4: Um den Effekt des Pretests zu kontrollieren, wird nachfolgend ein zweiter Faktor (Pretest ja/nein) in die Analyse eingeführt. Hierbei sollte weder der Haupteffekt dieses zweiten Faktors signifikant werden (alleiniger Pretest-Effekt) noch der Effekt der Interaktion aus erstem und zweitem Faktor (also ein zusammengesetzter Effekt aus Pretest und Treatment). Der Haupteffekt des ersten Faktors (Kontrollgruppe/Experimentalgruppe) sollte hingegen bestehen bleiben. H4 lautet demnach: *Der Unterschied in den Gruppen im Posttest ist nicht durch den Pretest verursacht.*

H5: Schließlich wird die Annahme überprüft, inwieweit das Treatment in den vier verschiedenen Kohorten ein vergleichbares Ergebnis im Posttest zeitigt, also das Treatment eine gewisse Stabilität aufweist: *In den Gruppen A1 bis A4 existieren keine signifikanten Mittelwertsunterschiede.*

5. Ergebnisse

H1: In der Treatmentgruppe A4 ergibt sich vom Pretest zum Posttest ein signifikanter Leistungszuwachs.

Um über die verschiedenen Analysen eine bessere Vergleichbarkeit herzustellen, wird auch an dieser Stelle schon eine Varianzanalyse mit Messwiederholung statt eines T-Tests für verbundene Stichproben berichtet. Der Mittelwert des Pretests, $M = 45.59$, $S_D = 7.33$, ist signifikant geringer als der des Posttests, $M = 49.21$, $S_D = 10.09$, $F(1.86) = 11.822$, $p < .001$, $\eta^2 = .121$. Somit kann Hypothese 1 als bestätigt angesehen werden. Zudem liegt ein praktisch bedeutsamer Effekt vor; nach Cohen (1988) steht ein $\eta^2 > .01$ für einen kleinen, ein $\eta^2 > .06$ für einen mittleren und ein $\eta^2 > .14$ für einen hohen Effekt. Das Treatment zeigt hier also einen mittleren bis hohen Effekt. Erwähnenswert ist die Erhöhung der Standardabweichung vom ersten zum zweiten Zeitpunkt. Dies könnte daher rühren, dass das Treatment einige Personen besonders gut anspricht, wohingegen andere „auf der Stelle treten“. Diese Erklärung deckt sich mit der allenfalls als moderat zu bezeichnenden Korrelation ($r = .397$, $p < .001$) zwischen den WLE-Scores des Pre- und Posttests.

H2: Die Leistungsentwicklung in der Treatmentgruppe A4 ist signifikant größer als in der Kontrollgruppe B1.

Tabelle 4 zeigt die Mittelwerte und Standardabweichungen in den beiden Gruppen A4 und B1 für den Pre- und Posttest.

Der Haupteffekt des Faktors „Zeitpunkt“ ist signifikant, $F(1.127) = 5.68$, $p < .05$, $\eta^2 = .043$, der Haupteffekt des Faktors „Gruppe“ dagegen nicht, $F(1.127) = 3.46$, $p > .05$, $\eta^2 = .027$. Die Interaktion zwischen den beiden Faktoren fällt wiederum signifikant aus, $F(1.127) = 5.68$, $p < .05$, $\eta^2 = .043$, was für die Wirkung des Treatments spricht, die in Abbildung 2 veranschaulicht wird. Somit kann auch Hypothese 2 als bestätigt angesehen werden.

| | Pretest | | Posttest | |
|--------|---------|----------------|----------|----------------|
| | M | S _D | M | S _D |
| A4 | 45.59 | 7.33 | 49.21 | 10.09 |
| B1 | 44.79 | 7.75 | 44.79 | 7.75 |
| Gesamt | 45.33 | 7.45 | 47.77 | 9.59 |

Tab. 4: Mittelwerte und Standardabweichungen der WLE-Scores in den beiden Gruppen A4 und B1 für den Pre- und Posttest

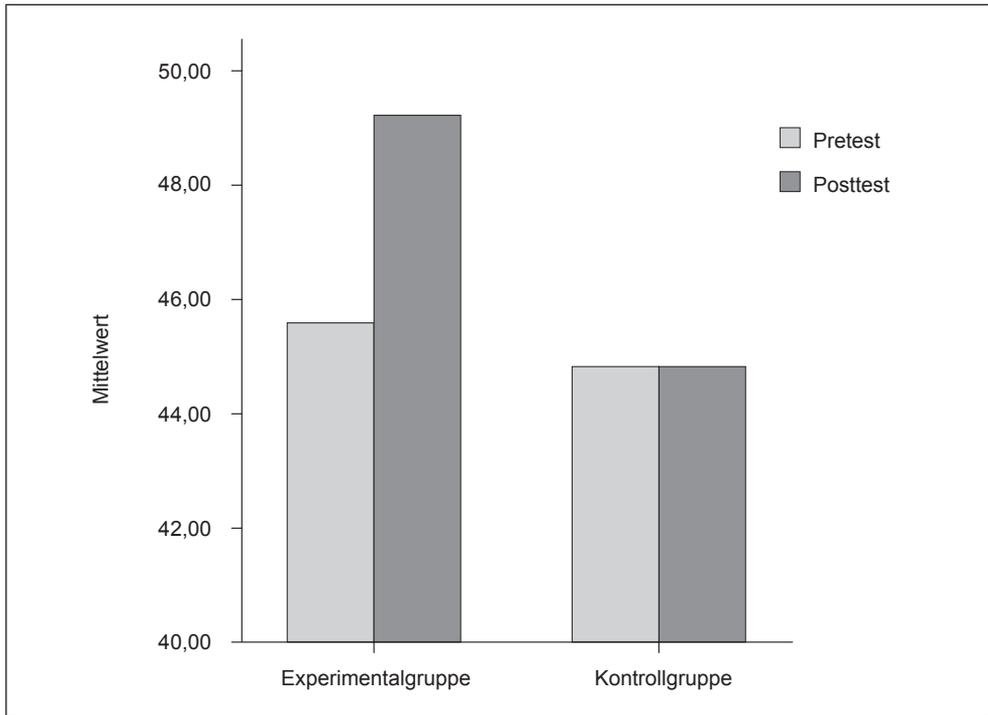


Abb. 2: Vergleich der Experimental- und Kontrollgruppe im Pre- und Posttest anhand der Mittelwerte der WLE-Scores

H3: Die Treatmentgruppen A1 bis A4 unterscheiden sich im Posttest signifikant von den Kontrollgruppen B1 und B2.

Tabelle 5 zeigt die Mittelwerte und Standardabweichungen in den Experimental- und Kontrollgruppen für den Posttest.

Eine Varianzanalyse bestätigt den Unterschied der mittleren Leistung zwischen den Experimental- und Kontrollgruppen im Posttest, $F(1.362) = 6.21$, $p < .05$, $\eta^2 = .017$.

H4: Der Unterschied in den Gruppen im Posttest ist nicht durch den Pretest verursacht.

Diese Fragestellung wird überprüft, indem in die vorangegangene Varianzanalyse ein zweiter Faktor „Pretest ja/nein“ eingefügt wird. Der Haupteffekt dieses zweiten Faktors ist nicht signifikant, $F(1.360) = 2.425$, $p > .05$, $\eta^2 = .007$. Ebenso ist die Interaktion zwischen „Gruppe“ und „Pretest“ nicht signifikant, $F(1.360) = 2.18$, $p > .05$, $\eta^2 = .006$. Der Haupteffekt des Faktors „Gruppe“ bleibt bestehen, $F(1.360) = 5.93$, $p < .05$, $\eta^2 = .016$. Es ist somit festzuhalten, dass die im Posttest gemessenen Lernzuwächse allein auf das Treatment und nicht auf (mögliche) Testeffekte zurückzuführen sind.

| | Posttest | |
|-----------|----------|----------------|
| | M | S _D |
| A1 bis A4 | 49.27 | 9.26 |
| B1 und B2 | 44.79 | 7.75 |
| Gesamt | 48.69 | 9.19 |

Tab. 5: Mittelwerte und Standardabweichungen der WLE-Scores in den Experimental- und Kontrollgruppen für den Posttest

| | Posttest | |
|--------|----------|----------------|
| | M | S _D |
| A1 | 50.60 | 7.92 |
| A2 | 49.22 | 9.26 |
| A3 | 48.39 | 9.29 |
| A4 | 49.21 | 10.09 |
| Gesamt | 49.27 | 9.26 |

Tab. 6: Mittelwerte und Standardabweichungen der WLE-Scores in A1 bis A4

H5: In den Gruppen A1 bis A4 existieren keine signifikanten Mittelwertsunterschiede. Tabelle 6 zeigt die Mittelwerte und Standardabweichungen in den vier Subgruppen A1 bis A4 für den Posttest.

Der Effekt des Faktors „Untergruppe“ in dieser Varianzanalyse wird nicht signifikant, $F(3.276) = .656$, $p > .05$, $\eta^2 = .007$. Dieses Ergebnis spricht für die Stabilität des Treatments, da es in verschiedenen Treatmentgruppen zu ähnlich hohen Leistungen führt.

6. Zusammenfassung und Diskussion

Ziel der hier referierten Evaluationsstudie war es, die *Wirksamkeit spezifischer Lerngelegenheiten* im Rahmen des erziehungswissenschaftlichen Studiums empirisch zu überprüfen. Die Realisierung dieser Intention setzte ein an „Standards“ orientiertes Curriculum voraus, das Studierenden die Möglichkeit bietet, didaktisch-methodisches Wissen zu erwerben und in der Analyse von Unterricht berufsfeldbezogen einzuüben.

Die Validität des Curriculums sehen wir über ein Expert(inn)enrating (Tab. 1) als gegeben. In Ergänzung dazu kann die Evaluation durch die Studierenden (Gruppe A4, N = 82; s. Tab. 2) als positives Indiz für die Akzeptanz und wahrgenommene Qualität des tatsächlichen Veranstaltungsverlaufs gedeutet werden.

Die erzielten Testleistungen der Probanden stehen für signifikante Unterschiede zwischen Experimental- und Kontrollgruppen. Die mittlere bis hohe Effektstärke belegt zudem die praktische Bedeutsamkeit des Lernzuwachses in A4. Hinsichtlich der erzielten Lernerfolge in den vier Teilgruppen A1 bis A4 ist hervorzuheben, dass zwischen ihnen keine signifikanten Leistungsunterschiede vorliegen. Das Treatment führt also in verschiedenen Gruppen zu ähnlich hohen positiven Wirkungen.

Hinsichtlich dieser Wirkungen muss allerdings einschränkend hinzugefügt werden, dass wir keine Nachhaltigkeitsmessung zu einem späteren Zeitpunkt durchgeführt haben. Insofern lässt sich nichts über längerfristige Effekte des Treatments sagen, die zu-

sammen mit den kurzfristigen Effekten erst eine Gesamtbeurteilung ermöglichen. Wir haben aus pragmatischen Gründen auf eine solche Messung verzichtet, weil diese im universitären Kontext äußerst schwierig zu realisieren ist.

Die gleichmäßige Wirkung des Treatments könnte man als Indiz für die Reproduzierbarkeit der Interventionsmaßnahme und für eine gewisse *Generalisierbarkeit* der erzielten Ergebnisse sehen. Dieser Aspekt der Generalisierbarkeit (externe Validität) spielt gerade bei der Beurteilung von Evaluationsstudien eine zentrale Rolle, in denen die interne Validität begrenzt ist, da die potenziell wirksamen Variablen nur bedingt kontrolliert werden können. Das gilt selbstverständlich auch für ein komplexes Setting einer Lehrveranstaltung, in der eine ganze Reihe von institutionellen, kontextuellen und individuellen Determinanten (Motivation, thematisches Interesse, Arbeitsbelastung usw.) den Lernerfolg beeinflussen können (s. Helmke, 1996). Für den Bereich von Evaluationsstudien hat Cronbach (1982) deshalb das prinzipielle Spannungsverhältnis zwischen interner und externer Validität stärker zugunsten der letzteren ausgelegt und damit die Frage der Generalisierbarkeit von Ergebnissen über die der Kontrolle von Störbedingungen gestellt. Diese Forderung Cronbachs sehen wir in unserer Studie über die gleichmäßige Wirkung des Treatments in vier voneinander unabhängigen Gruppen (A1 bis A4) ansatzweise realisiert. Freilich bedürfte dieser Sachverhalt der zusätzlichen Überprüfung an anderen Populationen (z. B. Lehramtsstudierende für die Primarstufe und Sekundarstufe I).

Prinzipielle Grenzen von Evaluationsstudien zeigen sich nicht nur im Verhältnis von interner und externer Validität, sondern auch in der Verfügbarkeit „einer wirklich gut geeigneten Vergleichsgruppe“ (Moosbrugger & Schweizer, 2002, S. 29), die gerade auch für Lehrevaluationen schwierig zu finden ist. Je stärker sich das thematische Angebot in der Kontrollgruppe von dem in der Experimentalgruppe unterscheidet, desto wahrscheinlicher sind Treatmenteffekte nachweisbar. In unserem Fall hätte z. B. ein Angebot zu historischen, soziologischen oder bildungstheoretischen Themen die Experimentalgruppe vermutlich von vornherein zum „Gewinner“ gemacht. Wir haben dieses Problem durch einen Balanceakt zu lösen versucht, indem wir in der Kontrollgruppe ein thematisches Angebot („Unterrichtsqualität“) gewählt haben, das noch dem Aufgabenfeld des Unterrichtens zuzurechnen ist, aber nicht auf die Aneignung didaktisch-methodischen Wissens in Form von *Struktur- und Prozessmerkmalen* des Unterrichts und insbesondere nicht auf die praktische Anwendung dieses Wissens ausgerichtet war. Solche Entscheidungen muss man bei der Beurteilung der an sich positiven Ergebnisse im Blick haben.

Schließlich muss der Stellenwert der hier berichteten positiven Wirkungen spezifischer Lerngelegenheiten vor dem Hintergrund gesehen werden, dass es sich um eine einzelne Veranstaltung handelt, die (noch) nicht in ein universitäres Gesamtkonzept bildungswissenschaftlicher Studien eingebunden war. Vor dem Hintergrund der BilWiss-Studie wäre das aber dringend angeraten, denn deren Ergebnisse „können dahingehend interpretiert werden, dass es derzeit den Universitäten in NRW nur unzureichend gelingt, bildungswissenschaftliches Wissen systematisch und kumulativ aufzubauen“ (Kunina-Habenicht et al., 2013, S. 18). Freilich ist die Qualität von Gesamtkonzepten

nicht durch eine formal und inhaltlich eher allgemein gehaltene Dokumentation von Modulhandbüchern gesichert; entscheidend ist vielmehr, „die Studienwege und Veranstaltungsangebote inhaltlich und methodisch so zu gestalten, dass die erforderlichen Kompetenzen schrittweise entstehen können“ (Terhart, 2007, S. 11).

Der dadurch angestrebte systematische und kumulative Aufbau bildungswissenschaftlicher Kompetenz ist zweifelsohne ein hoch gestecktes Ziel, das vermutlich nur in kleinen Schritten realisiert werden kann. In dieser Hinsicht scheinen uns *spezifische Lerngelegenheiten* und die *Evaluation ihrer Wirkungen* geeignete Mittel zu sein. Wir sehen unser Modell dabei nicht als *best practice*, sondern als mögliches Beispiel, zu dem weitere empirisch fundierte Modelle spezifischer Lerngelegenheiten in Vergleich treten sollten. Solche konkurrierenden Modelle könnten möglicherweise wieder *Dynamik* in die Lehrerbildung bringen. Sollte das gelingen, hätte die schon lange andauernde Kritik – wie sie eingangs skizziert wurde – dann doch ihre Funktion als Motor des Fortschritts erfüllt, sodass man künftig über die Lehrerbildung das sagen könnte, was ein großer Naturforscher seinen Kritikern entgegeng gehalten haben soll: „Und sie bewegt sich doch!“

Literatur

- Bauer, J., Diercks, U., Rösler, L., Möller, J., & Prenzel, M. (2012). Lehramtsausbildung in Deutschland: Wie groß ist die strukturelle Vielfalt? *Unterrichtswissenschaft*, 40(2), 101–120.
- Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, R. Bromme & R. Lehmann (Hrsg.), *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 2: Mathematische und physikalische Kompetenzen in der Oberstufe* (S. 271–315). Opladen: Leske + Budrich.
- Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 197–238). Münster/New York/München/Berlin: Waxmann.
- Blömeke, S., & König, J. (2010). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 239–263). Münster/New York/München/Berlin: Waxmann.
- Blömeke, S., & König, J. (2011). Zum Zusammenhang von Ausbildungsformen, -inhalten und -methoden mit dem erworbenen pädagogischen Professionswissen von Grundschullehrkräften. *Zeitschrift für Grundschulforschung*, 4, 33–46.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. überarb. Aufl.). Berlin: Springer.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd edition. Hillsdale: Lawrence Erlbaum.

- Cramer, C. (2013). Beurteilung des bildungswissenschaftlichen Studiums durch Lehramtsstudierende in der ersten Ausbildungsphase im Längsschnitt. *Zeitschrift für Pädagogik*, 59(1), 66–82.
- Cronbach, L. J. (1982). *Designing Evaluation of Educational and Social Programs*. San Francisco: Jossey-Bass.
- Czerwenka, K., & Nölle, K. (2011). Forschung zu ersten Phase der Lehrerbildung. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 362–380). Münster: Waxmann.
- Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education*, 26, 247–273.
- Expertenkommission (2007). Ausbildung von Lehrerinnen und Lehrern in Nordrhein-Westfalen (Auszüge). Bericht der Expertenkommission unter Vorsitz von Jürgen Baumert. *Seminar – Lehrerbildung und Schule (Sonderheft)*, 12, 10–20.
- Hartig, J., & Kühnrich, O. (2005). Schätzung von Veränderung mit Plausible Values in mehrdimensionalen Rasch-Modellen. In A. Ittel & H. Merrens (Hrsg.), *Veränderungsmessung und Längsschnittstudien in der Erziehungswissenschaft* (S. 27–44). Wiesbaden: Springer VS.
- Heimann, P. (1976). *Didaktik als Unterrichtswissenschaft* (hrsg. u. eingel. v. K. Reich u. H. Thomas). Stuttgart: Ernst Klett.
- Helmke, A. (1996). Studentische Evaluation der Lehre – Sackgassen und Perspektiven. Anmerkung zum Beitrag von Rosemann & Schweer. *Zeitschrift für Pädagogische Psychologie*, 10, 181–186.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Klafki, W. (1996). *Neue Studien zur Bildungstheorie und Didaktik: Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* (5., unveränd. Aufl.). Weinheim/Basel: Beltz.
- Kleickmann, T., & Anders, Y. (2011). Lernen an der Universität. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 305–316). Münster: Waxmann.
- KMK (2004). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.), *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16. 12. 2004*. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf [02. 11. 2014].
- König, J., & Blömeke, S. (2012). Future teachers' general pedagogical knowledge from a comparative perspective: Does school experience matter? *ZDM – Mathematics Education*, 44, 341–354.
- König, J., & Seifert, J. (Hrsg.) (2012). *Lehramtsstudierende erwerben pädagogisches Wissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung*. Münster: Waxmann.
- 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, 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.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.

- Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55–68). Münster: Waxmann.
- Lohmann, V., Seidel, V., & Terhart, E. (2012). Der Kompetenzbereich „Unterrichten“ im intendierten Curriculum der universitären Lehrer/innen/bildung: Struktur, Substanz, Varianz. In T. Hascher & G. H. Neuweg (Hrsg.), *Forschung zur (Wirksamkeit der) Lehrer/innen/bildung* (S. 65–83). Wien: LIT.
- McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, 17, 305–322.
- Moosbrugger, H., & Schweizer, K. (2002). Evaluationsforschung in der Psychologie. *Zeitschrift für Evaluation*, 1, 19–37.
- Nolle, A. (2004). *Evaluation der universitären Lehrerinnen- und Lehrerbildung. Erhebung zur pädagogischen Kompetenz von Studierenden der Lehramtsstudiengänge*. München: Martin Meidenbauer.
- Plöger, W. (2012). „Beobachtung und Analyse von Unterricht“ – Ein Vorlesungskonzept zur Förderung der Analysekompetenz im Lehramtsstudium (unveröffentlichtes Manuskript).
- Plöger, W., & Scholl, D. (2014). Analysekompetenz von Lehrpersonen – Modellierung und Messung. *Zeitschrift für Erziehungswissenschaft*, 17, 85–112.
- Plöger, W., Scholl, D., & Seifert, A. (2015). Analysekompetenz – ein zweidimensionales Konstrukt?! *Unterrichtswissenschaft*, 43, 166–184.
- Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird. *Psychologische Rundschau*, 47, 78–92.
- Seidel, T., Prenzel, M., Rimmerle, R., Schwindt, K., Kobarg, M., Herweg, C., & Dalehefte, I. M. (2006). Unterrichtsmuster und ihre Wirkungen. Eine Videostudie im Physikunterricht. In M. Prenzel & L. A. Näcke (Hrsg.), *Untersuchungen zur Qualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 99–123). Münster: Waxmann.
- Spiel, C., Gradinger, P., & Lüftenegger, M. (2010). Grundlagen der Evaluationsforschung. In H. Holling & B. Schmitz (Hrsg.), *Handbuch Statistik, Methoden und Evaluation* (S. 223–232). Göttingen: Hogrefe.
- Spinath, B., & Stehle, S. (2011). Evaluation von Hochschullehre. In N. Birbaumer, D. Frey, J. Kuhl, W. Schneider & R. Schwarzer (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich B, Serie II, Bd. 1: Grundfragen und Anwendungsfelder psychologischer Diagnostik* (S. 617–667). Göttingen: Hogrefe.
- Terhart, E. (2005). Standards für die Lehrerbildung – ein Kommentar. *Zeitschrift für Pädagogik*, 51, 275–279.
- Terhart, E. (2007). Standards in der Lehrerbildung – eine Einführung. *Unterrichtswissenschaft*, 36, 2–14.
- Terhart, E. (2009). Erste Phase. Lehrerbildung an der Universität. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nikolaus & R. Mulder (Hrsg.), *Lehrerprofessionalität* (S. 425–437). Weinheim/Basel: Beltz.
- Terhart, E. (2012). „Bildungswissenschaften“: Verlegenheitslösung, Sammeldisziplin, Kampfbegriff? *Zeitschrift für Pädagogik*, 58(1), 22–39.
- Voss, T., & Kunter, M. (2011). Pädagogisch-psychologisches Wissen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 193–214). Münster: Waxmann.
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical and psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103, 952–969.

Wissenschaftsrat (2001). *Empfehlungen zur künftigen Struktur der Lehrerbildung*. Köln.
Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-Aspect Test Software* (Computer Program). Camberwell: Australian Council for Educational Research.

Abstract: At present it seems that the attempt to systematically build up sufficient educational-scientific knowledge during the first phase of teacher training is not successful. The key to an improvement of that situation lies not so much in the quantity of the respective offers, but rather in their *quality*. This is decisively shaped by the content-related and methodological organization of *specific learning opportunities*. The present contribution reports how such learning opportunities can be given a curricular structure and how they may be effectively realized through an adequate teaching event. The evaluation is carried out via a Solomon four-group design. The significant and strong effects shown by the experimental group with regard to performance development indicate the importance of *specific learning opportunities* in the acquisition of educational-scientific knowledge and substantiate the need for the development and empirical evaluation of further models of adequate learning offers.

Keywords: Teacher Education, Learning Opportunities, Development of Educational Competencies, Quality of Teaching, Evaluation

Anschrift der Autoren

Prof. Dr. Wilfried Plöger, Universität zu Köln, Humanwissenschaftliche Fakultät,
Institut für Allgemeine Didaktik und Schulforschung,
Aachener Straße 201, 50931 Köln, Deutschland
E-Mail: wilfried.ploeger@uni-koeln.de

Dr. Daniel Scholl, Universität zu Köln, Humanwissenschaftliche Fakultät,
Institut für Allgemeine Didaktik und Schulforschung,
Aachener Straße 201, 50931 Köln, Deutschland
E-Mail: daniel.scholl@uni-koeln.de

Dr. Andreas Seifert, Universität Paderborn,
Zentrum für Bildungsforschung und Lehrerbildung (PLAZ)
Mersinweg 3, 33098 Paderborn, Deutschland
E-Mail: seifert@plaz.upb.de