

Lindmeier, Anke; Heinze, Aiso

## Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant?

*Praetorius, Anna-Katharina [Hrsg.]; Grünkorn, Juliane [Hrsg.]; Klieme, Eckhard [Hrsg.]: Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen. 1. Auflage. Weinheim; Basel : Beltz Juventa 2020, S. 255-268. - (Zeitschrift für Pädagogik, Beiheft; 66)*



Quellenangabe/ Reference:

Lindmeier, Anke; Heinze, Aiso: Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant? - In: Praetorius, Anna-Katharina [Hrsg.]; Grünkorn, Juliane [Hrsg.]; Klieme, Eckhard [Hrsg.]: Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen. 1. Auflage. Weinheim; Basel : Beltz Juventa 2020, S. 255-268 - URN: urn:nbn:de:0111-pedocs-258785 - DOI: 10.25656/01:25878

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-258785>

<https://doi.org/10.25656/01:25878>

in Kooperation mit / in cooperation with:

# BELTZ JUVENTA

<http://www.juventa.de>

### Nutzungsbedingungen

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

### Terms of use

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

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

### Kontakt / Contact:

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

Mitglied der

  
Leibniz-Gemeinschaft

66. Beiheft

April 2020

# **ZEITSCHRIFT FÜR PÄDAGOGIK**

---

**Empirische Forschung zu Unterrichts-  
qualität. Theoretische Grundfragen und  
quantitative Modellierungen**

**BELTZ** JUVENTA

Zeitschrift für Pädagogik · 66. Beiheft



Zeitschrift für Pädagogik · 66. Beiheft

# **Empirische Forschung zu Unterrichtsqualität**

**Theoretische Grundfragen  
und quantitative Modellierungen**

Herausgegeben von  
Anna-Katharina Praetorius, Juliane Grünkorn  
und Eckhard Klieme

**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.



ISSN: 0514-2717

ISBN 978-3-7799-3534-6 Print

ISBN 978-3-7799-3535-3 E-Book (PDF)

Bestellnummer: 443534

1. Auflage 2020

© 2020 Beltz Juventa

in der Verlagsgruppe Beltz · Weinheim Basel

Werderstraße 10, 69469 Weinheim

Alle Rechte vorbehalten

Herstellung: Hannelore Molitor

Satz: text plus form, Dresden

Druck und Bindung: Beltz Grafische Betriebe, Bad Langensalza

Printed in Germany

Weitere Informationen zu unseren Autoren und Titeln finden Sie unter: [www.beltz.de](http://www.beltz.de)

# Inhaltsverzeichnis

*Anna-Katharina Praetorius/Juliane Grünkorn/Eckhard Klieme*  
Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen  
und quantitative Modellierungen. Einleitung in das Beiheft ..... 9

## **Themenblock I: Dimensionen der Unterrichtsqualität – Theoretische und empirische Grundlagen (englischsprachig)**

*Anna-Katharina Praetorius/Eckhard Klieme/Thilo Kleickmann/Esther Brunner/  
Anke Lindmeier/Sandy Taut/Charalambos Charalambous*  
Towards Developing a Theory of Generic Teaching Quality: Origin,  
Current Status, and Necessary Next Steps Regarding the Three Basic  
Dimensions Model ..... 15

*Thilo Kleickmann/Mirjam Steffensky/Anna-Katharina Praetorius*  
Quality of Teaching in Science Education: More Than Three  
Basic Dimensions? ..... 37

*Courtney A. Bell*  
Commentary Regarding the Section “Dimensions of Teaching Quality –  
Theoretical and Empirical Foundations” – Using Warrants and Alternative  
Explanations to Clarify Next Steps for the TBD Model ..... 56

## **Themenblock II: Angebots-Nutzungs-Modelle als Rahmung (deutschsprachig)**

*Svenja Vieluf/Anna-Katharina Praetorius/Katrin Rakoczy/Marc Kleinknecht/  
Marcus Pietsch*  
Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts:  
ein kritischer Vergleich verschiedener Modellvarianten ..... 63

*Sibylle Meissner/Samuel Merk/Benjamin Fauth/Marc Kleinknecht/  
Thorsten Bohl*  
Differenzielle Effekte der Unterrichtsqualität auf die aktive Lernzeit ..... 81

*Tina Seidel*

Kommentar zum Themenblock „Angebots-Nutzungs-Modelle als Rahmung“ – Quo vadis deutsche Unterrichtsforschung? Modellierung von Angebot und Nutzung im Unterricht .....	95
---	----

### **Themenblock III: Oberflächen- und Tiefenstruktur des Unterrichts (deutschsprachig)**

<i>Jasmin Decristan/Miriam Hess/Doris Holzberger/Anna-Katharina Praetorius</i> Oberflächen- und Tiefenmerkmale – eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung .....	102
--	-----

<i>Miriam Hess/Frank Lipowsky</i> Zur (Un-)Abhängigkeit von Oberflächen- und Tiefenmerkmalen im Grundschulunterricht – Fragen von Lehrpersonen im öffentlichen Unterricht und in Schülerarbeitsphasen im Vergleich .....	117
---	-----

<i>Christine Pauli</i> Kommentar zum Themenblock „Oberflächen- und Tiefenstruktur des Unterrichts“: Nutzen und Grenzen eines prominenten Begriffspaares für die Unterrichtsforschung – und das Unterrichten .....	132
--	-----

### **Themenblock IV: Zur Bedeutung unterschiedlicher Perspektiven bei der Erfassung von Unterrichtsqualität (englischsprachig)**

<i>Benjamin Fauth/Richard Göllner/Gerlinde Lenske/Anna-Katharina Praetorius/ Wolfgang Wagner</i> Who Sees What? Conceptual Considerations on the Measurement of Teaching Quality from Different Perspectives .....	138
--	-----

<i>Richard Göllner/Benjamin Fauth/Gerlinde Lenske/Anna-Katharina Praetorius/ Wolfgang Wagner</i> Do Student Ratings of Classroom Management Tell us More About Teachers or About Classroom Composition? .....	156
---	-----

<i>Marten Clausen</i> Commentary Regarding the Section “The Role of Different Perspectives on the Measurement of Teaching Quality” .....	173
--	-----

## **Themenblock V: Modellierung der Wirkungen von Unterrichtsqualität (englischsprachig)**

<i>Alexander Naumann/Susanne Kuger/Carmen Köhler/Jan Hochweber</i> Conceptual and Methodological Challenges in Detecting the Effectiveness of Learning and Teaching .....	179
<i>Carmen Köhler/Susanne Kuger/Alexander Naumann/Johannes Hartig</i> Multilevel Models for Evaluating the Effectiveness of Teaching: Conceptual and Methodological Considerations .....	197
<i>Oliver Lüdtke/Alexander Robitzsch</i> Commentary Regarding the Section “Modelling the Effectiveness of Teaching Quality” – Methodological Challenges in Assessing the Causal Effects of Teaching .....	210
 <b>Kommentare</b>	
<i>Ewald Terhart</i> Unterrichtsqualität zwischen Theorie und Empirie – Ein Kommentar zur Theoriediskussion in der empirisch-quantitativen Unterrichtsforschung .....	223
<i>Kurt Reusser</i> Unterrichtsqualität zwischen empirisch-analytischer Forschung und pädagogisch-didaktischer Theorie – Ein Kommentar .....	236
<i>Anke Lindmeier/Aiso Heinze</i> Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant? .....	255

Anke Lindmeier/Aiso Heinze

## Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant?

**Zusammenfassung:** Dieser Beitrag diskutiert die Artikel des Beihefts „Empirische Forschung zu Unterrichtsqualität“ aus einer fachdidaktischen Perspektive. Dazu wird zunächst die eingenommene Sicht genauer spezifiziert, indem insbesondere der normative fachdidaktische Aspekt von Unterrichtsqualität betont wird. Auf dieser Basis wird dann die Rolle fachdidaktischer Merkmale bei den bisher betrachteten Indikatoren für Unterrichtsqualität untersucht. Anschließend wird diskutiert, inwieweit die in der Unterrichts-forschung betrachteten eher kurzen Ausschnitte von Unterricht eine Erfassung von genuin fachdidaktischen Qualitätsindikatoren ermöglichen. Der Beitrag schließt mit ergänzen-den, möglicherweise relevanten fachspezifischen Faktoren der Unterrichtsqualität, die in diesem Beiheft nicht betrachtet wurden.

**Schlagworte:** Unterrichtsqualität, Unterrichtsforschung, Forschungsperspektiven, Fachdidaktik, Qualität von Fachunterricht

### 1. Einleitung

Aus Sicht der Fachdidaktiken scheint die Unterrichtsqualitätsforschung einen paradoxen Charakter zu haben. So verfolgt sie einerseits das Ziel, das komplexe Konstrukt ‚Unterrichtsqualität‘ *fachunabhängig* zu konzeptualisieren und relevante Qualitätsindikatoren zu bestimmen, und hat dabei andererseits den Anspruch, die Qualität von *Fachunterricht* messen zu können. Dem Wilhelm von Ockham zugeschriebenen Forschungsprinzip der Parsimonie folgend kann dieses fachunabhängige Vorgehen zweifellos als sinnvoll angesehen werden, da ein Qualitätskonstrukt für Unterricht, das ohne fachspezifische Merkmale auskommt, äußerst sparsam wäre. Allerdings wirft so ein Vorgehen viele Fragen auf, die für die Fachdidaktik geradezu existenzielle Bedeutung haben: Wenn Unterrichtsqualität ohne Bewertung fachlicher und fachdidaktischer Merkmale des Unterrichtsgeschehens auskommt,

- warum erhalten Lehrkräfte eine fachliche und fachdidaktische Ausbildung anstelle nur einer allgemeindidaktischen Ausbildung?
- warum werden für den Fachunterricht fachdidaktische Instruktionsansätze und Lehrmaterialien für bestimmte Themengebiete entwickelt anstatt nur Inhalte entlang der Fachlogik zu lehren?
- wozu gibt es überhaupt die Fachdidaktiken?

Wenn man das Konstrukt ‚Unterrichtsqualität‘ in der Konzeptualisierung wirklich fachunabhängig versteht und entsprechend misst, dann müsste man die Qualität von Unterrichtsstunden unterschiedlicher Fächer, beispielweise von Mathematik- und Deutschunterricht, vergleichen können. Die einzige uns dazu bekannte Studie in Deutschland von Praetorius, Vieluf, Saß, Bernholt und Klieme (2016) vergleicht Deutsch- und Englischunterricht, also zwei sprachliche und sich nicht gänzlich fremde Fächer bezüglich der als fachunabhängig angesehenen Qualitätsdimensionen „Klassenführung“ und „Motivationale Unterstützung“. Es ergab sich bereits hier für die Motivationale Unterstützung ein nicht zu vernachlässigender Einfluss des Faches, und die Vermutung ist, dass dieser bei der Dimension „Kognitive Aktivierung“ noch stärker ausfallen dürfte.

Empirische Evidenz für den Einfluss fachspezifischer Merkmale auf die Effektivität von Fachunterricht liefert zusammenfassend die Metaanalyse von Seidel und Shavelson (2007). Reusser (2009) verweist auf das didaktische Dreieck, das die verschiedenen Einflussfaktoren für Unterrichtsqualität vereint und das neben Lehrkräften und Schülerinnen und Schülern auch die Rolle des Faches explizit enthält. Auch diverse Beiträge in diesem Beiheft enthalten Hinweise, dass das Fach und das Verständnis von Unterrichtsqualität nicht vollständig voneinander zu separieren sind. So enthält beispielsweise der Vorschlag des revidierten Angebots-Nutzungs-Modells in Vieluf, Praetorius, Rakoczy, Kleinknecht und Pietsch (in diesem Heft) das Fach als relevanten Kontext zur Beschreibung von Unterricht; Hess und Lipowsky (in diesem Heft) beziehen bei der Bewertung des kognitiven Frageniveaus als Tiefenmerkmal des Leseunterrichts die „inhaltliche Relevanz“ der Frage für die Leseübung explizit mit ein, und Decristan, Hess, Holzberger und Praetorius (in diesem Heft) sehen das Verhältnis von Fachlichkeit und Tiefenstrukturen als ein gegenwärtig zu klärendes Forschungsfeld an.

Die Frage, ob die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung bisher ignoriert wurde, implizit enthalten ist oder keine Relevanz hat, bleibt komplex. Die Antwort wird auch davon abhängen, ob es um das theoretische Konstrukt Unterrichtsqualität oder um dessen Messung geht. Oder kurz: Ob die Fachspezifität bereits mit der Konzeptualisierung oder erst mit der Operationalisierung beginnt (vgl. Vehmeyer, 2009, S. 169).

Dass das Fach den Fachunterricht prägt, wird kaum bestritten werden. Im Englischunterricht wird englische Grammatik eine Rolle spielen, und der Chemieunterricht wird nicht ohne die Behandlung chemischer Elemente und Reaktionen stattfinden. Empirisch lässt sich zeigen, dass die *fachlichen* Inhalte im Sinne eines implementierten Curriculums Einfluss auf die Schülerleistung haben (z. B. auf Basis von Schülerberichten: Kuger, Klieme, Lüdtke, Schiepe-Tiska & Reiss, 2017; auf Basis der Schulbuchinhalte: Sievert, van den Ham, Niedermeyer & Heinze, 2019). Ob genuin *fachdidaktische* Aspekte neben fachlichen, allgemeindidaktischen und lehr-lern-psychologischen Aspekten des Unterrichts relevant sind, ist nicht mehr so klar. Dahinter steckt auch die Frage, was die Fachdidaktiken ausmacht und ob sie als eigene Disziplin angesehen werden können. So stellte etwa Paul Kirschner in Bezug auf die Lehrerprofessionsforschung infrage, dass das Konstrukt *fachdidaktisches Wissen* notwendig sei, da Lehrkräfte für das Unterrichten nur Fachwissen und darauf angewendetes bildungswissenschaftliches Wissen

benötigen würden (Kirschner, Verschaffel, Star & Van Dooren, 2017). Auf die Unterrichtsqualitätsforschung übertragen hieße dies, dass sie ohne eine genuin fachdidaktische Perspektive auskäme und Unterrichtsmerkmale aus einer Kombination von fachlichen sowie allgemeindidaktischen und lehr-lern-psychologischen Kriterien zu werten wären, die in Teilen aufeinander zu beziehen sind. Dass dieses ‚Aufeinanderbeziehen‘ jedoch komplexer ausfallen dürfte, als es vielleicht klingt, wird unten noch thematisiert.

Betrachtet man generell die Forschung in den Fachdidaktiken, so können hier grob zwei Bereiche unterschieden werden. Zum einen beschäftigen sich die Fachdidaktiken mit deskriptiven und explikativen Fragestellungen, d. h. wie die Realität im Fachunterricht ist und welche Mechanismen hinter dem fachlichen Lernen stecken. Diese Sichtweise korrespondiert mit der empirischen Unterrichtsforschung. Zum zweiten erarbeiten die Fachdidaktiken normativ-präskriptive Vorschläge für den Unterricht, d. h. wie die Realität im Zusammenhang mit dem Fachunterricht sein sollte. Auch dieses zweite Ziel spielt für die Unterrichtsqualitätsforschung eine zentrale Rolle. So beschreibt Berliner (2005) die Notwendigkeit von zwei Arten von normativ festgelegten Kriterien in Bezug auf Unterrichtsqualität: Kriterien zur unmittelbaren Bewertung der Qualität von Unterricht („good teaching“) und Kriterien, die sich auf den Effekt von Unterricht beziehen und Unterrichtsqualität damit indirekt über den Unterrichtseffekt bewerten („effective teaching“). Auch andere Autorinnen und Autoren betonen, dass die Unterrichtsforschung nicht frei von normativen Vorstellungen ist (z. B. Kunter & Ewald, 2016; Openshaw & Clarke, 1970). Wie bei Praetorius et al. (in diesem Heft) bereits herausgearbeitet, stehen Forschende im Bereich Unterrichtsqualität vor dem Dilemma, Wertneutralität anzustreben, während gleichzeitig das soziale Phänomen Unterricht per se durch geteilte Wertvorstellungen und Normen geprägt ist. Ebenso wie normative Vorstellungen zu Unterrichtsansätzen einfließen (z. B. Kunter & Ewald, 2016, zur positiven Konnotation des entdeckenden Lernens), ist in Bezug auf Fachunterricht davon auszugehen, dass in vielen Studien normative fachdidaktische Aspekte sowohl für die Unterrichtsmerkmale (z. B. Funktion des Experiments in der Chemiestunde) als auch – falls genutzt – für die Unterrichtseffekte (z. B. Konzeptualisierung von Kompetenz in Chemie als Outcome-Maß) relevant sind. Die jeweiligen normativen Setzungen werden aber in den verwendeten Modellen nicht immer expliziert, was zu Problemen bei der Interpretation der Messergebnisse führen kann. Diese Feststellung gilt im Übrigen nicht nur aus fachspezifischem Blick, sondern allgemeiner. Etwa zeigt die Analyse der Genese verschiedener Angebots-Nutzungs-Modelle von Vieluf et al. (in diesem Heft) deutlich unterschiedliche Grundsetzungen. Implizite Normdifferenzen können sich dabei generell auf den verschiedenen Kontext-Ebenen (z. B. Fach, Schularten, Bundesländer, siehe Vieluf et al., in diesem Heft) ergeben.

## 2. Die Rolle von normativen Vorstellungen in der Unterrichtsforschung

Die Rolle von Normen wird aus unserer Sicht in der aktuellen Literatur nur punktuell berücksichtigt, und man könnte kritisch anmerken, dass es häufig unklar ist, ob sich die Wissenschaftlerinnen und Wissenschaftler dieser normativen Setzungen bewusst sind. Im Folgenden soll an verschiedenen Beispielen die Rolle von Normen aufgezeigt werden, wobei ein Fokus auf die fachdidaktische Perspektive gelegt wird.

### 2.1 *Die Rolle von fachspezifischen Normen bei der Konzeptualisierung und Operationalisierung von Unterrichtsqualität*

In den bisherigen Studien wurde die generische Auffassung von Unterrichtsqualität bei der Operationalisierung von einzelnen Konstrukten – etwa kognitive Aktivierung – jeweils für das betrachtete Fach mehr oder weniger adaptiert. Fachdidaktische Qualitätsmerkmale wurden dabei selten explizit benannt, und das Verhältnis von Fachlichkeit und Unterrichtsqualitätsmerkmalen bleibt oft unklar (vgl. Decristan et al., in diesem Heft). Oft findet die Beurteilung hochinferent statt, wobei fachdidaktische Aspekte möglicherweise implizit bleiben. So geben Kunter und Ewald (2016, Tab. 1) einen Überblick zur Erfassung der kognitiven Aktivierung in verschiedenen Fächern und nennen als Beispiele u. a. die Indikatoren „Fragen, die zu langen und inhaltlichen Antworten anregen“, „Fragen, die zum Nachdenken anregen“ oder die „Vermittlung von Lesestrategien durch die Lehrkraft“, die im jeweiligen Fach beim Rating fachdidaktisch interpretiert werden können – aber nicht müssen. So können Fragen, die fachliche Fehlvorstellungen bestärken, ebenfalls zum Nachdenken anregen und lange inhaltliche Antworten induzieren, und nicht jede beliebige Lesestrategie wird das Lesen lernen nachhaltig unterstützen. Vereinfacht gesagt, stellt sich die Frage, inwieweit die kognitive Aktivierung zielgerichtet den fachlichen Lernprozess unterstützt und ihn nicht be- oder verhindert. Kriterien auf Basis fachdidaktischer Erkenntnisse dürften dabei eine zentrale Rolle spielen.

Studien zur Unterrichtsqualität, die explizit die Fragen der Fachlichkeit adressieren, sind eher rar. Exemplarisch liegt eine Studie aus der Fachdidaktik Mathematik vor (Brunner, 2018), für die eine Unterrichtsstunde mit drei Erhebungsinstrumenten zur Unterrichtsqualität beurteilt wurde, darunter mit dem stark fachbezogen ausgerichteten Instrument TRU (Schoenfeld et al., 2013). Es ergaben sich je nach verwendetem Instrument für die gleiche Mathematikstunde deutlich unterschiedliche Qualitätsbewertungen, wobei insbesondere das Einbeziehen der fachlichen Korrektheit als Kriterium eine Ursache für die verschiedenen Ergebnisse war. Ein weiteres prominentes Beispiel für die Erfassung fachdidaktischer Unterrichtsmerkmale ist das Pythagoras-Projekt (Lipowsky, Drollinger-Vetter, Klieme, Pauli & Reusser, 2018). Hier wurden zunächst auf fachdidaktischer Basis theoretische ‚Verstehenselemente‘ für den Satz des Pythagoras herausgearbeitet und diese dann für die fachdidaktische Bewertung von Unterricht genutzt (Drollinger-Vetter, 2011). Anzumerken ist, dass es sich dabei um themenspezifische inhaltsbezogene Qualitätsmerkmale handelt, die in bestimmten Phasen von Unter-

richtsstunden (sog. Theoriephasen) erfasst wurden. Es zeigte sich in weiteren Analysen, dass sich ein über Verstehenselemente konzeptualisiertes Konstrukt für fachdidaktische Qualität des Unterrichts zum Satz des Pythagoras von den drei Basisdimensionen empirisch abgrenzen lässt (Lipowsky et al., 2018). Auf den ersten Blick mag es nicht verwundern, dass in beiden Studien der Einsatz anderer oder zusätzlicher Instrumente andere Bewertungsaspekte betont. Dies kann im Falle divergierender Ergebnisse der Qualitätsmessung aber zu einem Problem führen, etwa wenn fachliche Korrektheit in Studien nicht erfasst wird und die Ergebnisse normbildend rezipiert werden.

Offen bleibt an dieser Stelle die interessante Frage, ob das Modell der drei Basisdimensionen durch eine fachspezifische vierte Dimension erweitert werden sollte oder nicht. Die Ergebnisse von Lipowsky et al. (2018) deuten auf eine empirische Trennbarkeit einer fachdidaktischen Qualitätsdimension hin, beziehen sich aber auf einen sehr speziellen Unterrichtsinhalt. Schlesinger, Jentsch, Kaiser, König & Blömeke (2018) stellen ein Instrument vor, das sich global auf Mathematikunterricht bezieht und neben Skalen für die drei Basisdimensionen zwei weitere fachspezifische Skalen enthält (*subject-related quality, teaching-related quality*). Unabhängig von der noch ausstehenden empirischen Trennbarkeit möglicher fachspezifischer Basisdimensionen stellt sich vor allem die Frage der theoretischen Begründung. So können diverse Items der vorgeschlagenen fachspezifischen Skalen theoretisch auch der Basisdimension „Kognitive Aktivierung“ zugeordnet werden (z. B. *using multiple representations, dealing with mathematical errors of students, relevance of mathematics for students*). Damit wäre die Frage nicht, ob die drei Basisdimensionen durch fachspezifische Dimensionen ergänzt werden sollten, sondern eher, ob die Dimension „Kognitive Aktivierung“ nicht fachspezifisch ausdifferenziert werden muss. Letzteres würde explizit machen, was bisher als implizite fachdidaktische Operationalisierung (s. o.) bereits vorliegt. Inwieweit diese Ausdifferenzierung sinnvoll und notwendig ist, dürfte von der Forschungsfrage abhängen, die mit dem Instrument beantwortet werden soll – und damit von der in der jeweiligen Studie benötigten Auflösung des Konstrukts Unterrichtsqualität.

## 2.2 *Inwieweit ist die Operationalisierung der Qualität von Fachunterricht kulturell geprägt?*

Die zuvor angesprochene Rolle der fachspezifischen Normen bei der Untersuchung von Unterrichtsqualität wird komplexer, wenn kulturell geprägte Normen zum Fachunterricht einbezogen werden. Wie in der Debatte von Heid (2013) und Klieme (2013) herausgearbeitet wurde, muss im Kern zwischen der Deskription von Unterrichtsmerkmalen und deren Bewertung unterschieden werden. Erfasst man beispielsweise die Anzahl an offenen Lehrerfragen pro Unterrichtsstunde, so ist dies ein deskriptiver Akt. Qualifiziert man einen höheren Anteil offener Lehrerfragen als besser, so ist das der bewertende Akt. Bei genauerem Blick lassen viele der aktuell genutzten Unterrichtsqualitätsindikatoren ihren Ursprung in der Idealvorstellung eines konstruktivistisch geprägten Unterrichts erkennen, der mit einer gewissen Auffassung vom Lernen, der Rolle der

Lehrkräfte und der Rolle von Diskurs, Wissen oder Übung für individuelle Lernprozesse aufgeladen ist (vgl. auch Kunter & Ewald, 2016). Gleichzeitig weiß man aus der kulturvergleichenden Sozialpsychologie, dass solche Werthaltungen und Normen zwischen verschiedenen Kulturen stark differieren können (z. B. Hofstede, 2001). So können beispielsweise im Gegensatz zu individualistischen Gesellschaften (z. B. in Europa und Nordamerika, den sog. westlichen Ländern) in kollektivistischen Gesellschaften (z. B. den asiatischen Ländern der sog. *confucian heritage culture*) Indikatoren wie ein hoher Anteil individueller Schüleräußerungen oder das Eingehen einer Lehrkraft auf individuelle Verständnisschwierigkeiten negativ bewertet werden (z. B. Clarke, 2013a zur Rolle von Schüleräußerungen im Mathematikunterricht). Diese Normunterschiede machen die Grenzen kulturübergreifender Forschung im Bereich Unterrichtsqualität deutlich. In den internationalen Vergleichsuntersuchungen werden deswegen Konsensverfahren genutzt und strittige Indikatoren eliminiert, womit allerdings Validitätsprobleme auftreten können (*validity-comparability compromise*, Clarke, 2013b).

Der Kontrast zwischen stark unterschiedlichen Kulturen ist besonders geeignet, um das potenzielle Problem unterschiedlicher Normvorstellungen deutlich zu machen. Dieses Phänomen kann aber auch bei größerer Nähe der Kulturen auftreten. Praetorius und Charalambous (2018) stellen beispielsweise beim Vergleich von 12 generischen und mathematikspezifischen Instrumenten für Unterrichtsqualität aus dem westlichen Kontext fest, dass es auffällige Unterschiede gibt: Während die Indikatoren meist konstruktivistische Auffassungen von Lehren und Lernen spiegeln (wie z. B. kognitive Aktivierung), werden in manchen Instrumenten auch behaviorale Aspekte (wie z. B. Üben) abgebildet. Dabei ist unstrittig, dass die Wahl der Indikatoren und ihrer qualitätskonstituierenden Ausprägungen keine vollständig objektivierbaren Prozesse sein können. Es liegen entsprechend auch viele unterschiedliche Operationalisierungen vor, die allerdings die Vergleichbarkeit der Ergebnisse beeinträchtigen.

Vielleicht lassen sich einige ‚blind spots‘ dadurch erklären, dass ohne kontrastierende Zugänge die häufig impliziten eigenen Normen und Wertvorstellungen kaum der Reflektion zugänglich sind. Deswegen wird beispielsweise im „lexicon project“ (Mesiti & Clarke, 2017) daran gearbeitet, Begriffe zur Beschreibung von (Mathematik-)Unterricht in verschiedenen Sprachen zu sammeln und zu sortieren. Ein Ziel ist, für zentrale Terme eine standardisierte Übersetzung zur Erleichterung komparativer Forschung zu gewinnen. Ein zweites Ziel ist, die genuinen Entwicklungen der verschiedenen Kulturen wechselseitig zugänglich zu machen. Dabei folgt das Projekt der Erkenntnis, dass ein Unterrichtsmerkmal, für das es in einer Sprache keinen Begriff gibt, in der zugehörigen Kultur auch nicht leicht zugänglich ist und entsprechend in Operationalisierungen kaum berücksichtigt werden kann (Clarke, 2013b). Dieses Grundprinzip trifft auch für Unterschiede in der Unterrichtskultur zwischen den Fächern oder zwischen einem allgemeinen und fachspezifischen Blick auf Unterricht zu, beispielsweise wenn generische Messinstrumente nicht zwischen verschiedenen Erklärqualitäten unterscheiden. Für die Schülerfrage: „Warum hat 4:0 kein Ergebnis?“ könnten etwa die zwei Erklärungen „Weil das eine Division durch 0 ist und das ist nicht definiert“ und „Weil geteilt durch 0 ja bedeuten würde, man verteilt 4 Dinge auf 0 Personen und das geht nicht“ aus

generischer Sicht ähnlich bewertet werden, während die Mathematikdidaktik ihnen sehr unterschiedliche Qualitäten in Bezug auf das für den Mathematikunterricht wichtige Kriterium der Verständnisorientierung zuweisen würde.

Vergleiche von verschiedenen Unterrichtsqualitätsmessinstrumenten (Praetorius & Charalambous, 2018), die aus unterschiedlichen Kontexten mit unterschiedlichen Zielsetzungen vor unterschiedlichen theoretischen Rahmungen entstanden sind, stellen einen ersten Ansatz dar, um die Vielfältigkeit der Operationalisierungen sichtbar zu machen. Im nächsten Schritt wäre es notwendig zu untersuchen, woher die Unterschiede stammen. Das kann zur Begriffsschärfung beitragen, da beispielsweise verschiedene Operationalisierungen desselben Konstrukts (etwa für verschiedene Fächer) von Unterschieden im Verständnis der Konstrukte (etwa zwischen Fächern oder Kulturen) differenziert werden könnten.

### 2.3 Fachspezifische Effektivität als Kriterium für Unterrichtsqualität

Wie in der Einleitung erwähnt, kann Unterrichtsqualität nicht nur unmittelbar bewertet werden, sondern auch mittelbar über den Unterrichtseffekt. Erkenntnisse zur Qualität von Unterricht nach diesem zweiten Kriterium hängen in hohem Grad von der Wahl der Forschungsfragen ab (siehe auch Seidel, in diesem Heft). So zeigen sich bei Studien zu kognitiven Outcomes andere Qualitätsmerkmale als prädiktiv als für nicht-kognitive Outcomes. Aus fachdidaktischer Sicht besteht die Schwierigkeit, dass die Wahl des Messinstruments für bzw. von Daten zu Unterrichtserfolg häufig von der Verfügbarkeit abhängt. Dies führt einerseits dazu, dass die Effekte von Unterrichtsqualität bevorzugt für Fächer untersucht werden, die als ‚leichter‘ messbar angesehen werden (wie z. B. Mathematik im Falle von kognitiven Outcomes). Andererseits stellt sich die Frage, ob das jeweilig eingesetzte Instrument für den intendierten Zweck als valide anzusehen ist. Wird beispielsweise die PISA-Mathematikskala als Outcome-Maß eingesetzt, so ist zu beachten, dass die Aufgaben aus fachdidaktischer Sicht eine klare normative Schwerpunktsetzung aufweisen (*mathematical literacy* ohne Anspruch auf curriculare Validität). Inwieweit die für internationale Vergleiche von Bildungssystemen entwickelten PISA-Aufgaben inhaltlich geeignet sind, um zur Unterrichtsqualitätsforschung im Fach Mathematik beizutragen, wäre also zu diskutieren (vgl. etwa die unterschiedliche Veränderungssensitivität der PISA-Skala und der Bildungsstandards-Skala in Lehner et al., 2017). Die Wahl des Outcome-Maßes als externes Kriterium für Unterrichtsqualität kann deutliche Auswirkungen auf die Ergebnisse haben. Systematische Prüfungen, inwiefern die Ergebnisse der Unterrichtsqualitätsforschung von der Wahl des Faches oder des Outcome-Messinstruments abhängen, fehlen bisher.

#### 2.4 *Wer erkennt was und warum? Zur Rolle von Raterinnen und Ratern bei der Messung von Unterrichtsqualität*

Während sich die zuvor diskutierten Problempunkte hauptsächlich auf die Frage beziehen, *was* beobachtet wird, tritt bei hochinferenten Ratings während der Ausführung eine weitere Herausforderung hinzu, die aus fachspezifischer Sicht Aufmerksamkeit erregt. Dabei handelt es sich um die Frage, *wer* die Unterrichtsqualitätsmerkmale im Ratingprozess beurteilt. In der Lehrerkompetenzforschung wurde in den letzten Jahren betont, dass die professionelle Wahrnehmung, dazu gehört das Erkennen und Interpretieren von relevanten Unterrichtsmerkmalen, eine Funktion von professionellem Wissen ist. Im sog. advokatorischen Ansatz (Oser, Heinzer & Salzmann, 2010) werden beispielsweise Noticing-Fähigkeiten von Lehrkräften als Indikator für deren Expertise verwendet. In den Beiträgen von Vieluf et al. (in diesem Heft) und Fauth, Göllner, Lenske, Praetorius und Wagner (in diesem Heft) wird die Rolle von Vorerfahrungen und Wissen sowie eigenen Prädispositionen für die Wahrnehmung des Unterrichtsangebots durch Lernende, Lehrende und durch Forschende (an-)diskutiert. Es stellt sich die Frage, welche Kompetenzen die Beurteilenden aufweisen müssen, um hochinferente Ratings mit (explizitem oder implizitem) fachlichem Bezug durchführen zu können, beispielsweise bei der Einschätzung, ob eine Frage im Fachunterricht kognitiv aktivierend ist. Beurteilende können die Lernenden selbst, die betroffenen Lehrkräfte oder aber externe Ratende sein, wobei das Phänomen nicht-korrespondierender Urteile zwischen verschiedenen ‚Beurteilergruppen‘ als „perspektivenspezifische Unterrichtsqualität“ bezeichnet wird. Fauth et al. (in diesem Heft) reflektieren dafür verschiedene Ursachen, von den gewählten Methoden und Indikatoren bis hin zur Formulierungsebene, und arbeiten so die damit verbundenen Herausforderungen auf. Wir möchten hier darüber hinaus einen bisher eher wenig betrachteten Aspekt innerhalb einer Perspektive adressieren.

Effekte unterschiedlicher externer Ratender<sup>1</sup> werden in der Regel mit Blick auf die Objektivität und Reliabilität der Urteile untersucht. Die in diesem Beiheft vorliegenden empirischen Beiträge berichten entsprechende Qualitätssicherungsmaßnahmen, womit sichergestellt wird, dass geschulte Personen ein ausgearbeitetes Ratingmanual möglichst intersubjektiv reliabel anwenden können. Unklar bleibt dabei, welchen Einfluss Wissen, Normen, Werthaltungen oder gar die Expertise der Beurteilenden selbst haben. Können Raterinnen und Rater mit einem bestimmten Hintergrund, beispielsweise in Erziehungswissenschaften oder Psychologie, fachliche Fehlvorstellungen von Schülerinnen und Schülern im (Chemie-, Deutsch-, Geschichts-, Mathematik-, Physik-)Unterricht erkennen<sup>2</sup> und valide einschätzen, ob Lehrkräfte damit fachdidaktisch adäquat um-

1 Bei Beurteilungen durch Lehrende selbst ist die Frage nach der Intersubjektivität hinfällig (Ausnahme: team-teaching). Bei der Beurteilung durch Lernende kann die Kohärenz der Urteile in einer Klasse als Indikator für die Beurteilungsgüte gesehen werden (z. B. Göllner, Fauth, Lenske, Praetorius & Wagner, in diesem Heft), in anderen Studien wird aber Varianz innerhalb der Klasse als Hinweis auf differenziell wirkenden Unterricht interpretiert.

2 Noch anspruchsvoller und gravierender sind fachliche Fehler der Lehrkräfte im Unterricht.

gehen oder die Fehlvorstellungen ggf. noch verstärken? Konkret geht es um die Frage, bis zu welchem Grad in Schulungen neben intersubjektiv übereinstimmenden Ratings auch kriterial suffiziente Beurteilungen der jeweiligen Unterrichtsqualitätsmerkmale erreicht werden können. Ausgehend von den oben erwähnten Ergebnissen zu Noticing-Fähigkeiten von Lehrkräften ist dies ein fachdidaktischer Aspekt der Messung von Unterrichtsqualität, der nicht zu verlässigen ist, bislang empirisch aber nicht in den Blick genommen wird.

### **3. Thin Slices, Unterrichtseinheit oder Unterrichtssequenz: Wann sind fachdidaktische Qualitätskriterien beobachtbar?**

Beobachtungsstudien zur Erfassung von Unterrichtsqualität beschränken sich aufgrund des hohen Aufwands notgedrungen auf wenige Unterrichtsstunden (oft sogar nur eine Stunde pro Klasse). Aufgrund der hohen Kosten sind die Forschenden bemüht, möglichst ökonomische Methoden zur Qualitätsmessung von Unterricht zu entwickeln, um bei gleichem Aufwand mehr Unterrichtsstunden untersuchen zu können. Ein interessanter Ansatz dieser Optimierungsversuche stellt das Thin-Slices-Verfahren bei der video-basierten Erfassung von Unterrichtsqualität dar. Beispielsweise wurden in der Studie von Begrich, Fauth, Kunter und Klieme (2017) von jeder der 37 Sachunterrichtsstunden drei zufällig ausgewählte 10-sekündige Ausschnitte von ungeschulten Beurteilenden in Bezug auf (generische) Qualitätsmerkmale bewertet. Es ergab sich eine hohe Inter-rater-Übereinstimmung, und die gemessene Unterrichtsqualität wies eine angemessene prognostische Validität im Hinblick auf den Lernerfolg auf. Auch wenn die Betrachtung von fachdidaktischen Merkmalen der Unterrichtsqualität im Thin-Slices-Verfahren noch aussteht, so kann vor dem Hintergrund der Diskussion im vorherigen Abschnitt mit aller Vorsicht vermutet werden, dass eine fachdidaktische Qualitätseinschätzung auf Basis von drei 10-sekündigen Ausschnitten durch fachdidaktisch ungeschulte Ratende herausfordernd sein dürfte. Diese Annahme beruht auf dem Hintergrund, dass eine Einschätzung fachdidaktischer Aspekte von Lehr-Lern-Prozessen in kurzen und damit situationalen Ausschnitten von Unterricht nur schwer möglich ist.

Die meisten Unterrichtsstudien betrachten eine Unterrichtsstunde. Damit liegt eine zeitliche Beobachtungseinheit vor, die der Planungseinheit der Lehrkraft für den Fachunterricht entspricht und somit eine weitergehende fachdidaktische Bewertung erlaubt, die Fischer, Reyer, Wirz, Bos und Höllrich (2002, S. 132) als den Versuch, „den ‚latenten Plan‘ des Lehrers für den Unterricht im Verlauf zu interpretieren“ beschrieben haben. Dabei geht es um die fachspezifische Tiefenstruktur des Fachunterrichts, die Fischer und Kollegen für den Physikunterricht operationalisiert haben, indem sie die Basismodelle von Oser und Patry (1990) für das unterrichtliche Physiklernen adaptierten (s. a. QuIP-Projekt, Fischer, Labudde, Neumann & Viiri, 2014; vgl. auch Decristan et al., in diesem Heft).

Aber auch die Betrachtung einer Unterrichtsstunde ermöglicht noch keine Bewertung von Angebotsstrukturen, die sich über mehrere Unterrichtsstunden entwickeln.

Praetorius, Pauli, Reusser, Rakoczy und Klieme (2014) konnten zeigen, dass insbesondere die Basisdimension der kognitiven Aktivierung, die am stärksten durch fachdidaktische Aspekte beeinflusst ist (s. Abschnitt 2.1), ein instabiler Merkmalsbereich ist und für verlässliche Aussagen eine Beobachtung von mindestens neun Unterrichtsstunden benötigt wird. Geht man davon aus, dass die fachdidaktische Planung von Unterricht idealtypisch nicht in Form von isolierten Unterrichtsstunden erfolgt, sondern als Unterrichtssequenz mit aufeinander aufbauenden Stunden unterschiedlicher inhaltlicher Zielsetzung, so sollte eine Betrachtung von themenbezogenen Unterrichtssequenzen das Potenzial für die Erfassung fachdidaktischer Qualitätsmerkmale erhöhen. Würden längerfristige Angebotsstrukturen betrachtet, so könnte man die Stabilität von Konstrukten wie der kognitiven Aktivierung (vgl. auch Seidel, in diesem Heft) und gleichzeitig die fachdidaktische Angebotsstruktur eines Themenbereichs untersuchen. Fachdidaktische Konzepte wie etwa die instruktionale Kohärenz eines inhaltlichen Unterrichtsangebots, die als Kohärenz der Lernziele, Kohärenz innerhalb von (mehrstündigen) Unterrichtseinheiten und Kohärenz über mehrere, zeitlich voneinander getrennte Unterrichtseinheiten konzeptualisiert wird (Shwartz, Weizman, Fortus, Krajcik & Reiser, 2008), bieten hier Ansätze zur Entwicklung von fachdidaktischen Qualitätsskalen. Da die Einschätzung der fachdidaktischen Qualität der Angebotsstruktur eine entsprechende Expertise voraussetzt, ist eine Erfassung auf Basis von Schülereinschätzungen nur schwer vorstellbar. Um aufwändige Unterrichtsbeobachtungen oder Videoaufzeichnungen zu vermeiden, wären beispielsweise Selbstberichte von Lehrkräften über ihre Unterrichtsplanung, die verwendeten Materialien, Aufgaben oder Arbeitsblätter eine ökonomische Variante zur Erhebung von Rohdaten.

#### 4. Zusammenfassung und Ausblick

Zusammenfassend ergibt sich durch die Diskussion der Unterrichtsforschung aus fachdidaktischer Perspektive ein durchaus komplexes Bild. Normen beeinflussen die Unterrichtsqualitätsforschung bei der Fokussierung des Forschungsinteresses, der Operationalisierung und dem eigentlichen Messvorgang. Normen unterscheiden sich zwischen Kulturen, aber auch zwischen anderen sozialen Kontextsystemen, beispielsweise Schulformen und Fächern, da unterschiedliche Zielsetzungen in unterschiedlichen pädagogischen Traditionen verfolgt werden. Bisher findet eine explizite Reflektion dieser implizit wirkenden Normen (ob generisch oder fachspezifisch) nur in geringem Ausmaß statt, sodass die Reichweite bzw. Generalisierbarkeit der vorhandenen Ergebnisse nicht eingeschätzt werden kann (vgl. auch Praetorius et al., in diesem Heft). Vergleichende oder kontrastierende Studien, die explizit den Einfluss von Normen adressieren, wären hier wünschenswert (z. B. analog zu Dreher, Lindmeier, Wang & Hsieh, 2018). Zur Erfassung und Bewertung von fachlichen und fachdidaktischen Merkmalen der Unterrichtsqualität (ob implizit oder explizit) sollten Personen mit entsprechender fachlicher Expertise einbezogen werden, da schon der Austausch zwischen den Disziplinen erhellend für die eigene Kultur wirken und so die Validität der Vorgehensweisen erhöhen kann.

Darüber hinaus ist anzudenken, fachdidaktische Aspekte der Unterrichtsqualität über die Betrachtung längerfristiger Angebotsstrukturen zu erfassen und ihre Prädikktivität für den Lernerfolg zu untersuchen.

Den bisherigen Ergebnissen der Unterrichtsforschung folgend und aufgrund der Integration generischer und fachspezifischer Aspekte in der Praxis des Unterrichts gehen wir nicht davon aus, dass es genügt, generische und fachspezifische Merkmale gesondert zu betrachten. Hingegen bietet sich die Hierarchisierung von Qualitätsmerkmalen des Fachunterrichts an. So wurde in Praetorius et al. (in diesem Heft) mit Bezug auf Openshaw und Clarke (1970) bereits erwähnt, dass die Klassenführung Voraussetzung für die konstruktive Lernerunterstützung und die kognitive Aktivierung sei. Brunner (2018) schlug in ihrem Beitrag ein erweitertes hierarchisches Modell von Qualitätsmerkmalen vor, bei dem neben der Klassenführung auch die fachliche Fundierung des Unterrichts (inkl. der fachlichen Korrektheit) die Voraussetzung für weitere generische und fachdidaktische Merkmale eines effektiven Fachunterrichts darstellt. Vor dem Hintergrund der in der Einleitung erwähnten Erkenntnisse, dass das implementierte Curriculum einen Einfluss auf die Effektivität von Unterricht hat, ist dieser Vorschlag plausibel.

Abschließend sei noch auf einen weiteren Ansatz für mögliche Studien hingewiesen: Die bisher nur geringe Rolle von fachdidaktischen Merkmalen in der Literatur zur Unterrichtsforschung könnte auch darauf zurückzuführen sein, dass die fachdidaktische Qualität im realen Unterricht – wenn sie untersucht wird – gering ist. Damit würden Skalen zu fachdidaktischen Qualitätsmerkmalen wenig Varianz aufweisen bzw. nicht ‚funktionieren‘. Um dies als mögliche Erklärung zu untersuchen, könnten Studien interessant sein, in denen idealtypische, d. h. fachdidaktisch optimierte Stundensequenzen umgesetzt und dann mit den entsprechenden Ratingverfahren bewertet werden. Eine ähnliche Idee wurde in der IGEL-Studie (Fauth, Decristan, Rieser, Klieme & Büttner, 2014) mit generischen Unterrichtselementen verfolgt (z. B. *peer assisted learning*, *formative assessment*). Auch in der Fachdidaktik gibt es Modelle für das fachliche Lernen, auf deren Basis idealtypische *best practice*-Stunden umgesetzt werden könnten (vgl. den Verweis von Seidel, in diesem Heft auf die *best practice*-Forschung in den USA).

Bleibt zum Schluss die Antwort auf die Frage im Titel dieses Diskussionsbeitrags: Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant? Einerseits lässt sich feststellen, dass die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung bisher weder vollständig ignoriert wurde noch dass sie als irrelevant angesehen wird. Andererseits muss aber auch konstatiert werden, dass die fachdidaktische Perspektive selten explizit einbezogen wurde, sondern eher ein implizites Dasein im Rahmen fachdidaktischer Interpretationen von Qualitätsindikatoren fristete. Dies mag mit einer bisher stark psychologischen und erziehungswissenschaftlichen Prägung der Unterrichtsforschung zu tun haben, die mit einem speziellen Blick auf Unterricht und damit verbundenen Fragestellungen und Forschungszugängen einhergeht. Eine Erweiterung dieses Blicks um fachdidaktische Fragestellungen wäre aus unserer Sicht ein naheliegender und wichtiger nächster Schritt.

## Literatur

- Begrich, L., Fauth, B., Kunter, M., & Klieme, E. (2017). Wie informativ ist der erste Eindruck? Das Thin-Slices-Verfahren zur videobasierten Erfassung des Unterrichts. *Zeitschrift für Erziehungswissenschaft*, 20 (Suppl. 1), 23–47.
- Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of teacher education*, 56(3), 205–213.
- Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik*, 39(2), 257–284.
- Clarke, D. J. (2013a). Contingent conceptions of accomplished practice: The cultural specificity of discourse in and about the mathematics classroom. *ZDM*, 45(1), 21–33.
- Clarke, D. J. (2013b). The validity-comparability compromise in crosscultural studies in mathematics education. In *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education* (S. 1855–1864).
- Dreher, A., Lindmeier, A., Wang, T.-Y., & Hsieh, F.-J. (2018). Teacher Noticing in Taiwan und Deutschland – Wie stark prägen kulturelle Normen das Verständnis von Unterrichtsqualitätsmerkmalen? In *Beiträge zum Mathematikunterricht 2018* (S. 461–464). Münster: WTM.
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit: fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Münster: Waxmann.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.
- Fischer, H. E., Reyer, T., Wirz, C., Bos, W., & Höllrich, N. (2002). Unterrichtsgestaltung und Lernerfolg im Physikunterricht. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 124–138). Weinheim: Beltz.
- Fischer, H. E., Labudde, P., Neumann, K., & Viiri, J. (Hrsg.) (2014). *Quality of instruction in physics: Comparing Finland, Switzerland and Germany*. Münster: Waxmann.
- Heid, H. (2013). Logik, Struktur und Prozess der Qualitätsbeurteilung von Schule und Unterricht. *Zeitschrift für Erziehungswissenschaft*, 16(2), 405–431.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks: Sage publications.
- Kirschner, P. A., Verschaffel, L., Star, J., & Van Dooren, W. (2017). There is more variation within than across domains: An interview with Paul A. Kirschner about applying cognitive psychology-based instructional design principles in mathematics teaching and learning. *ZDM*, 49(4), 637–643.
- Klieme, E. (2013). Qualitätsbeurteilung von Schule und Unterricht: Möglichkeiten und Grenzen einer begriffsanalytischen Reflexion – ein Kommentar zu Helmut Heid. *Zeitschrift für Erziehungswissenschaft*, 16(2), 433–441.
- Kuger, S., Klieme, E., Lüdtke, O., Schiepe-Tiska, A., & Reiss, K. (2017). Mathematikunterricht und Schülerleistung in der Sekundarstufe: Zur Validität von Schülerbefragungen in Schulleistungsstudien. *Zeitschrift für Erziehungswissenschaft*, 33, 61–98.
- Kunter, M., & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 9–31). Münster: Waxmann.
- Lehner, M. C., Heine, J.-H., Sälzer, C., Reiss, K., Haag, N., & Heinze, A. (2017). Veränderung der mathematischen Kompetenz von der neunten zur zehnten Klassenstufe. *Zeitschrift für Erziehungswissenschaft*, 33, 7–36.

- Lipowsky, F., Drollinger-Vetter, B., Klieme, E., Pauli, C., & Reusser, K. (2018). Generische und fachdidaktische Dimensionen von Unterrichtsqualität – Zwei Seiten einer Medaille? In M. Martens, K. Rabenstein, K. Bräu, M. Fetzer, H. Gresch, I. Hardy & C. Schelle (Hrsg.), *Konstruktionen von Fachlichkeit: Ansätze, Erträge und Diskussionen in der empirischen Unterrichtsforschung* (S. 183–202). Bad Heilbrunn: Klinkhardt.
- Mesiti, C., & Clarke, D. (2017). The international lexicon project: Giving a name to what we do. In R. Seah, M. Horne, J. Ocean, & C. Orellana (Hrsg.), *Proceedings of the Mathematical Association of Victoria annual conference* (S. 31–38). Brunswick, Vic.: Mathematical Association of Victoria.
- Openshaw, K., & Clarke, S. C. T. (1970). General teaching theory. *Journal of Teacher Education*, 21(3), 403–416.
- Oser, F., & Patry, J.-L. (1990). Choreographien unterrichtlichen Lernens: Basismodelle des Unterrichts. *Berichte zur Erziehungswissenschaft* (Nr. 89). Freiburg (CH): Pädagogisches Institut der Universität Freiburg.
- Oser F., Heinzer S., & Salzmann P. (2010), Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. Chancen und Grenzen des advokatorischen Ansatzes. *Unterrichtswissenschaft*, 38(1)1, 5–28.
- Praetorius, A. K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM*, 50(3), 535–553.
- Praetorius, A. K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Praetorius, A. K., Vieluf, S., Saß, S., Bernholt, A., & Klieme, E. (2016). The same in German as in English? Investigating the subject-specificity of teaching quality. *Zeitschrift für Erziehungswissenschaft*, 19(1), 191–209.
- Reusser, K. (2009). Empirisch fundierte Didaktik – didaktisch fundierte Unterrichtsforschung. In *Perspektiven der Didaktik* (S. 219–237). VS Verlag für Sozialwissenschaften.
- Schlesinger, L., Jentsch, A., Kaiser, G., König, J., & Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM*, 50(3), 475–490.
- Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM*, 45(4), 607–621.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of educational research*, 77(4), 454–499.
- Shwartz, Y., Weizman, A., Fortus, D., Krajcik, J. S., & Reiser, B. J. (2008). The IQWST experience: Using coherence as a design principle for a middle school science curriculum. *Elementary School Journal*, 109(2), 199–219.
- Sievert, H., van den Ham, A.-K., Niedermeyer, I., & Heinze, A. (2019). Effects of mathematics textbooks on the development of primary school children's adaptive expertise in arithmetic. *Learning and Individual Differences*, 74, 1–13.
- Vehmeyer, J. K. (2009). *Kognitiv anregende Verhaltensweisen von Lehrkräften im naturwissenschaftlichen Sachunterricht: Konzeptualisierung und Erfassung*. Dissertation Universität Münster. <https://miami.uni-muenster.de/Record/74b36e17-38c8-4130-8e9c-f14834595217> [08. 10. 2019].

**Abstract:** The paper discusses the articles on teaching quality in this special issue from the perspective of subject-specific educational research. First, we specify the view adopted by focusing on normative aspects of teaching quality. Against this background, we then analyse the role of subject-specific aspects for indicators of teaching quality considered so far. Subsequently, we discuss to what extent the common practice of analysing short time segments of classroom teaching can be used to measure teaching quality with criteria derived from subject-specific educational research. Finally, we address aspects of teaching quality that are not considered in the articles of this special issue, but may be relevant for a subject-specific perspective.

**Keywords:** Teaching Quality, Instructional Research, Research Perspectives, Subject-Specific Educational Research, Quality of Subject Teaching

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

Prof. Dr. Anke Lindmeier, IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik Kiel,  
Olshausenstraße 62, 24118 Kiel, Deutschland  
E-Mail: lindmeier@leibniz-ipn.de

Prof. Dr. Aiso Heinze, IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik Kiel,  
Olshausenstraße 62, 24118 Kiel, Deutschland  
E-Mail: heinze@leibniz-ipn.de