Praetorius, Anna-Katharina; Klieme, Eckhard; Kleickmann, Thilo; Brunner, Esther; Lindmeier, Anke; Taut, Sandy; Charalambous, Charalambos

Towards developing a theory of generic teaching quality. Origin, current status, and necessary next steps regarding the Three Basic Dimensions Model

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. 15-36. - (Zeitschrift für Pädagogik, Beiheft; 66)

66. Beiheft April 2020

ZEITSCHRIFT FÜR PÄDAGOGIK

Empirische Forschung zu Unterrichtsqualität. 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

ISSN: 0514-2717
ISBN 978-3-7799-3534-6 Print
ISBN 978-3-7799-3535-3 E-Book (PDF)
Bestellnummer: 443534

1. Auflage 2020

in der Verlagsgruppe Beltz · Weinheim Basel
Werderstraße 10, 69469 Weinheim

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

Inhaltsverzeichnis

Themenblock V: Modellierung der Wirkungen von Unterrichtsqualität (englischsprachig)

Kommentare

Themenblock I: Dimensionen der Unterrichtsqualität – Theoretische und empirische Grundlagen

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

Abstract: In this paper we elaborate upon the relevance of theories of teaching (quality) in quantitative empirical research on teaching. First we introduce, the quantitative empirical research approach. Then, we present the origin and current status of research with respect to a model – the Three Basic Dimensions of teaching quality – that is especially popular in quantitative research on teaching quality in German-speaking countries. Next, we reflect on the extent to which the model fulfills criteria for a good theory, before deriving conclusions for future research that focuses on a process of successive theory building.

Keywords: Theory, Teaching Quality, Three Basic Dimensions, Instruction, Model

1. The Relevance of Theories of Teaching and Teaching Quality

Scientific endeavors generally aim to develop theory: namely, general statements that allow for understanding, explaining, predicting, or critically reflecting upon phenomena. According to modern philosophy of science, any acceptable theory should be a coherent, systematic composition of such statements, and research should constantly aim to test and revise those theories. According to Kuhn (1969) and Lakatos (1977), theories are embedded into paradigms and research programs that are grounded in shared a priori assumptions and intended applications. In the present paper we discuss quality criteria for theories of teaching (see Section 3) and apply them to a model of teaching quality, the Three Basic Dimensions (TBD) model (the origins and foundations of which are outlined in Section 2), which has been developed in quantitative research on teaching quality in German-speaking countries.

Introducing a special issue in the journal *Zeitschrift für Pädagogik* on *Unterrichtstheorie* (theory of teaching[1]), Terhart (2014) stated that educational scientists rarely dis-

1 The German concept of *Unterricht* may be translated as teaching or as instruction. In fact, the

cuss teaching from a decisively theoretical, analytical perspective free of practical considerations. For example, while traditional didactics has in some cases been claimed to be a theory of teaching (e. g., Prange, 2005), Oelkers (2000) has emphasized that the 19th and 20th century German approaches to didactics and teaching should be perceived as reflections on professional practice rather than theory. More recently, Lüders (2014) criticized didactics as eclectic, lacking empirical validation, and/or being narrowly focused on curriculum. As examples of appropriate theories of teaching, both Terhart (2014) and Lüders (2014) cite approaches that are based on qualitative methods and conceptual foundations from the social sciences (e. g., practice theory; Reckwitz, 2002).

In the same special issue, Seidel (2014) provided a review of research on teaching based on quantitative methods. She distinguished two broader paradigms. In the first paradigm, psychological theories of cognition, metacognition, or motivation are used to identify features of classroom teaching that may support these aspects of learning. The second paradigm aims to systematize aspects of teaching, studying their structure and their relations with student outcomes. Historically, this paradigm emerged as a derivative and enhancement of the process-product paradigm (Gruehn, 2000; see Section 2). TBD is mentioned as an example of this second paradigm (see Seidel, 2014).[2]

With regard to the status of theories in quantitative research on teaching, Seidel's (2014) review is quite illuminating. Although it was published in a special issue on theories of teaching, all occurrences of the term *theory* were related to psychological theories of learning, cognition, metacognition, or motivation, while the term *model* was always used in discussing aspects of teaching within the second paradigm. Overall, the paper included the term *model* 63 times and the term *theory* just 21 times. Accordingly, Terhart (2014, p. 815) characterized the research reviewed by Seidel as "a process of developing and testing models. These models function as a shared foundation of thinking within a certain context of research, and require stepwise, self-corrective verification" (translated by the authors). What Terhart described is basically the process of theory building and revision, as envisioned in the philosophy of science. But in line with Seidel, he preferred to use the notion of model instead of theory, due perhaps to the sketchy, eclectic, or data-driven nature of some of this work. Theories, we may conclude, compared to models, need to be conceptually richer, more coherent, more robust, and more general.

---

terms teaching and instruction are often used interchangeably. Cohen, Raudenbush, and Ball (2003) consider teaching to be the narrower term, since it might be thought of as "something done by teachers to learners" whereas instruction in their understanding refers to "interactions among teachers and students, around content, in environments" (p. 122). *Unterricht* is very much in line with the latter notion. Nevertheless, we are using the term teaching throughout this paper, as this term meanwhile seems to be used in a much broader sense, encompassing all kinds of classroom and professional activities (cf. e. g., Ball & Forzani, 2009).

2 Although the title of the Seidel (2014) paper refers to psychology of teaching, the second paradigm mentioned could equally be subsumed under educational science. In the present paper, we refrain from establishing borderlines between educational science, educational psychology, and sociology of education, as teaching requires an interdisciplinary approach.

This fuzzy and oftentimes rather pragmatic way of conceptualizing teaching can be found in the international literature as well. For example, the most recent *Handbook of Research on Teaching* (Gitomer & Bell, 2016) does not include any general theory of teaching. The chapter by Cappella, Aber and Kim (2016) integrates research from different areas of psychology and educational science into a complex model which in many respects is similar to the German models of opportunities and uses of instruction (see Section 4). Some prominent current work on teaching in the US, such as Deborah Ball's *high leverage teaching practices* (Ball & Forzani, 2009), is similar to the pragmatic approaches of German didactics, although the link to empirical evidence is much stronger for the former. Other prominent research on teaching evolved in the form of frameworks guiding the development of student surveys or observation protocols (e. g., Danielson, 2013). Frameworks may over time evolve into more elaborated models that also explicate relations among constructs of interest (e. g., the Dynamic Model of Educational Effectiveness; Creemers & Kyriakides, 2008). Eventually, these models might mature into theories (Leplin, 1980) fulfilling the criteria we discuss below (see Section 3). This distinction between frameworks, models, and theories is, however, often not used systematically in research on teaching quality.

The current special issue intends to address the lack of theorizing and systematic revision of theories in quantitative research on teaching. In the present paper, we discuss the core assumption of the second paradigm mentioned by Seidel (2014): namely, the possibility of identifying a limited set of measurable dimensions that are well-founded in theories or models of teaching and that explain effects of teaching on students. As these dimensions are assumed to be the main drivers of teaching effectiveness, they are called quality dimensions. We study this assumption by choosing the TBD model[3] of teaching quality (Klieme, 2019; Kunter & Trautwein, 2013; Praetorius, Klieme, Herbert & Pinger, 2018), which has gained a lot of attention in the German-speaking research literature over the last two decades but which has also been applied internationally, for example in TIMSS (e. g., Nilsen & Gustafsson, 2016) and PISA (e. g., Kuger, Klieme, Lüdtke, Schiepe-Tiska & Reiss, 2017). Thus, studying the theoretical foundations of quantitative research on teaching, using TBD as an example, seems to be a valuable undertaking.

3 Depending on the publication, TBD is called a (theoretical) framework (e. g., Fauth, Decristan, Rieser, Klieme, & Büttner, 2014) or a (theoretical) model (e. g., Klieme, Pauli, & Reusser, 2009). At least in the abstract of an article by Klieme and Rakoczy (2008), TBD is also called a theory.

2. Three Basic Dimensions of Teaching Quality: Origins and Foundations of the Model

Carroll (1963) first introduced the notion of teaching quality to empirical educational research, defining it as the extent to which teaching enables students to learn a task as quickly and effectively as possible. In other words, teaching quality was understood as a moderator variable shaping the relation between student aptitudes, learning time, and learning outcomes. Thus it was not defined substantially, but was deemed to cover any aspect of teaching (i. e., the process) that may help to optimize the student learning outcomes (i. e., the product). In the decades to follow, the so-called process-product paradigm of quantitative empirical research aimed at identifying such aspects. During the late 1980s and the 1990s, the basic notions of the paradigm were refined to include mediator variables, enabling the coverage of cognitive and affective processes that govern student learning. At the turn of the century, a certain substantive consensus had been reached on pivotal aspects of teaching quality. A seminal meta-analysis conducted by Wang, Haertel, and Walberg (1993) showed that classroom management and the quality of student-teacher academic interactions (namely, the intensity as well as the quality of questioning and answering) had about the same mean effect size as cognitive and meta-cognitive student aptitudes, their home environment, and parental support.[4]

Building upon the research conducted in the process-(mediation)-product paradigm, large-scale research on educational trajectories and educational effectiveness was conducted in the 1990s at the Center for Educational Research at the Max Planck Institute for Human Development in Germany. In her dissertation, Gruehn (2000) combined the international research literature with national traditions in didactics (e. g., the notion of Socratic teaching as defined by Wagenschein (1992)) and research on classroom climate (Eder, 1996). Gruehn assessed students' perceptions of teaching in various subjects using a total of 21 questionnaire scales. After dropping nine scales, Gruehn was able to establish five second-order dimensions of teaching quality through Confirmatory Factor Analysis: classroom management, pacing, adaptivity, affective quality, and constructivism (Gruehn, 2000).

At the same time, the Max Planck Institute was operating as the National Center for the Third International Mathematics and Science Study (TIMSS 1995; Baumert et al., 1997). Baumert and colleagues decided to enhance the international TIMSS design by using the same 21 questionnaire scales, formerly used by Gruehn. Furthermore, Germany, the US, and Japan became part of a video study which, for the first time ever, would cover mathematics lessons in nationally representative samples of classrooms, allowing comparisons of teaching in these three countries (Stigler & Knoll, 1999). While publications from the international study were based on codes, ratings, and qualitative

4 Researchers had also set out to integrate findings conceptually (e. g., Rosenshine & Furst, 1973, with their concept of direct instruction). As in the German literature (see Section 1), however, many authors questioned whether these approaches can be called theories (see, e. g., Berliner, 2009; Biddle & Anderson, 1986; Hill & Schrum, 2002; Snow, 1973).

Factor 1 Classroom management	Factor 2 Student support	Factor 3 Cognitive activation
• Dealing effectively with disruptions • Frequency of disruptions (−) • Waste of instructional time (−) • Volatility of the teacher (−) • Clarity of rules • Clarity and structure of teaching • Monitoring • Time on task	• Social orientation • Individual frame of reference • Teachers' diagnostic competence regarding social needs • High interaction speed (−) • Achievement pressure (−)	• Socratic teaching • Challenging practicing • Repetitive practicing (−) • Teachers' ability to motivate students

Note. (−) indicates a reverse-scored scale.

Tab. 1: Scales representing the three basic dimensions in TIMSS-Video 1995

scripts derived from videos only (Stigler & Hiebert, 1999), the German team managed to combine the video recording in Grade 8 with the TIMSS assessment, plus a follow-up test and a student questionnaire implemented one year later. Thus, TIMSS and TIMSS-Video 1995 were combined into a full-size longitudinal study of teaching effectiveness in Germany.

Among others, an attempt was made to turn the 21 scales used by Gruehn (2000) into observation protocols for trained observers (Clausen, 2002). In order to systematize and structure the observational space, Klieme, Schümer, and Knoll (2001) ran an exploratory factor analysis with these high-inference rating scores, resulting in a clear three-dimensional solution (see Tab. 1; English labels provided by the present authors).[5]

Klieme et al. (2001) reported the main findings of this analysis in their non-technical paper, written for teachers interested in professional learning. They provided initial evidence that cognitive activation was positively related to gains in mathematics achievement, while student support was related to a more positive development of students' interest in mathematics. The authors suggested that there might be a non-linear relation between cognitive activation and gains in mathematics achievement (i. e., too much cognitive activation might be suboptimal). Using the same data (but without referring to TBD), Clausen (2002) additionally showed that teaching quality depends on the perspective (i. e., teacher, student, or observer perspective) from which it is evaluated, as these perspectives did not converge in their judgments of teaching quality.

All in all, the foundational work in TIMSS-Video established (a) the three-dimensional structure of teaching quality, (b) its relevance for explaining student outcomes over the course of a school year, also providing hints on (c) non-linear relations with student outcomes, and (d) the perspective-specific nature of judgments of teaching quality. Klieme et al. (2001) introduced TBD as a comprehensive model that was devel-

5 Both the scree test and the Kaiser criterion indicated that three factors, explaining a total of 73 % of the common variance of the scales, could be distinguished. Of the 21 scales, 17 could be unambiguously linked to one of the three factors (criterion: a unique loading > .65). The technical details given here were not reported in the original publication.

oped to reduce the multiplicity of measures of teaching quality available at that time to a smaller set of dimensions, applying the meta-theoretical principle of Occam's razor, which claims that scientific research should avoid redundancy. Instead of the longer lists of facets, aspects, features, factors, components, domains, or dimensions of (effective) teaching or teaching quality that had been published, a reduced set of basic dimensions was believed to be more easily interpreted and analyzed with respect to effects on student learning. The authors stated that the model should apply to all school subjects and grade levels, and potentially even to different countries.

Although they were mainly focused on summarizing the core ideas for a non-research audience, Klieme et al. (2001) established some initial theoretical foundations by linking the three dimensions to a re-conceptualization of traditional (mainly German) didactics published by Diederich and Tenorth (1997), who argued – citing seminal writings from the history of school pedagogy – that classroom teaching requires a certain level of student attentiveness, student motivation, and student understanding. Klieme et al. (2001) interpreted TBD as comprising those aspects of teaching that help teachers achieve these three student outcomes.

Soon after the publication of this foundational, yet informal paper, the research group at the Max Planck Institute became involved in PISA. Indicators for TBD were therefore implemented in national extensions to PISA 2000 and 2003. This is why the model was first explained in research papers published in the context of PISA. Klieme and Rakoczy (2003) provided theoretical explanations as to how and why the dimensions should be linked to student learning (see also Fig. 1) by referring to constructivist theories (as a

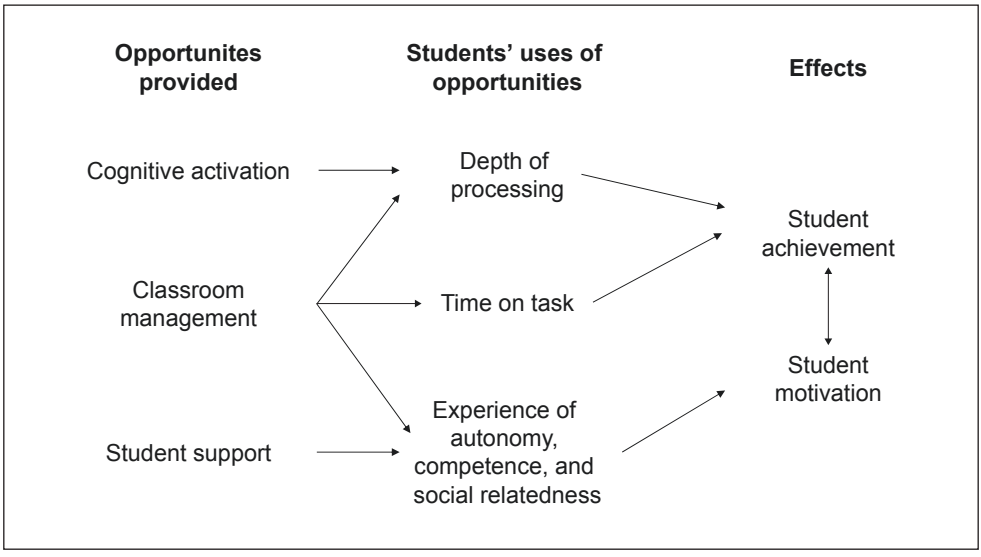

Fig. 1: The assumed relations between the three basic dimensions and student learning (adapted after Klieme et al., 2009, p. 140)

theoretical foundation for cognitive activation) and Self Determination Theory (Ryan & Deci, 2000; as a theoretical foundation for student support). Baumert, Kunter, and colleagues replicated the three dimensions for German data in PISA 2003 (published under the label COACTIV study; see Baumert et al., 2010; Kunter & Voss, 2011).

Internationally, the model of TBD was first presented by Klieme at the annual conference of the American Educational Research Association (AERA) in 2001, but the first scientific papers in English presenting TBD as a theoretical model (Klieme, Pauli & Reusser, 2009) or a theoretical conceptualization (Lipowsky et al., 2009) originated from a Swiss-German video study on teaching the Theorem of Pythagoras. This study also led to enhancements of the model. For example, Rakoczy, Klieme, Lipowsky, and Drollinger-Vetter (2010) discriminated behavioral aspects of classroom management from issues of content structure and clarity. Drollinger-Vetter (2011) and Lipowsky, Drollinger-Vetter, Klieme, Pauli, and Reusser (2018) discriminated generic aspects of cognitive activation from the quality of mathematical subject-matter presented.

Summing up, we can conclude that the TBD model, with its three-dimensional structure of teaching quality, was derived from factor analytical work. Initial empirical evidence showed its hypothesized relation to student achievement and motivation. Some effort has been spent on linking TBD to theoretical considerations, but the theoretical foundations of the TBD could still be improved, as shown in Section 4 of this paper. We envision that reflecting on possible ways to develop TBD into a theory could help to move the field forward by addressing the criticism that progress has so far largely been driven by empirical findings without a proper theoretical basis.

3. Desirable Characteristics of Theories of Teaching Quality

Quantitative empirical research on teaching, as it is addressed in the present special issue, is rooted in the understanding of scientific research that Terhart (2014) characterizes as "working on theories". This refers to a process of developing and testing models as a shared foundation of thinking in a research context that requires self-correcting examination. Despite the unclear delineation of the terms theory and model – and still more fuzzy terms such as shared foundation of thinking, or framework – there is a consensual need: quantitative empirical researchers in the field of teaching should constantly strive towards more mature frameworks/models/theories to gain more sophistication, rigorousness, and validity.

Although there are no clear, agreed-upon criteria for what qualifies as a theory in research, most researchers would probably agree that a theory is a system of concepts, connections, and explanations (e. g., Beck & Krapp, 2006; Kerlinger, 1964; Scriven, 1956/1994; Whetten, 1989). Following Rosenshine (2009), many would likely also agree that a theory on teaching "explains why and how teaching works[6], … explains

6 Others would not necessarily use the notion of "working", which seems to imply a rather technical approach.

why some kinds of teaching work better than others, … can be subjected to attempts at refutation, and … can be modified as new findings emerge" (Rosenshine, 2009, p. 1169).[7] The quantitative empirical paradigm thus clearly intends to link student learning (i. e., changes in dispositions, achievement, and behavior) to the classroom teaching in which students participate. Accordingly, it seems crucial to integrate theories of learning and motivation (e. g., social-constructivist theories of learning, Palincsar, 1998; or the self-determination theory of motivation, Ryan & Deci, 2000) with theories of teaching (e. g., Gage, 2009; Klieme et al., 2006).

In order to guide the development of TBD towards a more mature model, or even to a theory of teaching, criteria for scientific theories are needed. Some authors describe purposes and features of theories (e. g., Scriven, 1956/1994; Westermann, 2000) but do not present a set of crucial criteria on the basis of this description. Others do provide such a set (e. g., Beck & Krapp, 2006; Thagard, 1978; Whetten, 1989), but the criteria vary between authors, show considerable overlap, and are often formulated in a very general way.[8] Some authors, however, have also specified these general criteria for the field of teaching. In the following, we make pragmatic use of Kane and Marsh's (1980, p. 254) *integrated criteria for a general theory of instruction* (see Tab. 2) in order to highlight pivotal aspects regarding TBD. Although this set was proposed almost 40 years ago as a summary and integration of the assumptions of leading researchers, we are not aware of any more up-to-date set of criteria that is equally detailed and specific to theories of teaching.

As sections IA and IIB clearly show, Kane and Marsh (1980) subscribed to the view of analytical philosophy of science, defining theory as a consistent, hierarchical set of statements with empirical support and predictive value. Section IIA, testability, refers to a main aspect of critical rationalism. However, the set of criteria transcends these philosophical views, as limitations and restrictions of intended use are considered (IB), untested hypotheses are taken into account (IIB(3)), and prescriptions for teaching practice are requested (III). An aspect not included in the set of Kane and Marsh (1980) but emphasized in other places (e. g., Kuhn, 1970; Seidel, Prenzel & Krapp, 2014; Westermann, 2000) is a focus on the relation of a newly developed theory to existing approaches: Researchers need to make clear the extent to which a new theory builds upon, extends, refines, or overrules existing theories and observations. Only in doing so can we help in "developing the field by building cumulatively on existing knowledge and theory, rather than constantly attempting to reinvent the wheel" (Muijs et al., 2014, p. 251).

---

7 This view is in line with analytical philosophy of science and critical rationalism (Popper, 2005). Modern philosophy of science (for an overview, see Chalmers, 2007) also offers different perspectives, such as the non-statement view of theory, and scientific anti-realism. Nevertheless, analytical and critical-rationalist views dominate in the thinking of researchers conducting quantitative empirical research.

8 At the same time, such sets also show that in building theories, striking a balance between different criteria is necessary, as some of them are not independent, and some even contradict each other (e. g., comprehensiveness and parsimony in the set by Kane and Marsh, 1980).

I.	**Theoretical Characteristics**
A.	Characteristics and organization of the components. A theory of instruction should consist of a set of:
(1)	logically, and
(2)	theoretically related
(3)	internally consistent statements (axioms, corollaries, postulates), arranged in a
(4)	hierarchical or systematic order, so that
(5)	the higher level constructs integrate the constructs below.
(6)	These statements should be as few as possible to cover all of the theories and findings relevant to the area specified and should be
(7)	clearly defined.
(8)	If possible, these statements should be quantitatively related, as well as
(9)	qualitatively related.
B.	Boundaries
	The boundaries or limitations of concern of the theory should be stated, including such limitations as theories of learning and development subscribed to, philosophies adhered to, characteristics of the students and organizations deemed suitable. The most general theory will have as few such limitations as possible.
II.	**Empirical Characteristics**
	The statements included (except for axiomatic statements and those noted in IIB(3)) should relate to existing empirical evidence in the following manner:
A.	Testability
	The statements should be:
(1)	capable of being easily and clearly restated in the form of hypotheses about which
(2)	evidence can be collected to either verify or refute them.
B.	Support
	The statements should have
(1)	demonstrable empirical support and
(2)	predictive value in similar situations.
(3)	However, at the present time it may be necessary to include as yet untested hypotheses to meet the completeness criteria noted in section IA above.
III.	**Prescriptive Characteristics**
	To be of practical use, a theory of instruction should contain or clearly imply a series of prescriptive statements, specifying how best to obtain given ends, if they are desired. Areas to be covered include strategies, sequencing, materials, reinforcements, motivation.

Tab. 2: Set of integrated criteria for a general theory of instruction (Kane & Marsh, 1980, p. 254)

4. Reflecting on the Theoretical Foundation of TBD

In the following, we apply the core quality criteria mentioned above in an illustrative manner to the current understanding of TBD (see also Section 2). The literature included is based on the TBD review by Praetorius et al. (2018), some reviews available in German (Klieme, 2019; Kunter & Ewald, 2016; Lipowsky & Bleck, 2019) and on the authors' knowledge of the literature.

Our aim in applying the theory quality criteria to TBD is to help researchers in the field of teaching understand how their work can better contribute to theory building by using a model that holds considerable potential to evolve into a theory of teaching. To minimize the risk of subjective interpretations regarding the theoretical and empirical evidence for TBD, we ensured that our author group was diverse by (a) including the main developer of TBD, while at the same time making sure that we included researchers (b) from different disciplines, (c) working in more applied or research-oriented settings, (d) from both German-speaking and other countries, and (e) both working with TBD themselves, and not working with TBD. To ensure a broad and consensual perspective in application to each criterion, the initial drafts of the sections to follow were developed further until a consensus was reached by this interdisciplinary group.

4.1 Are the Statements Logically and Theoretically Related as well as Internally Consistent?

TBD can basically be seen as the combination of a structural component (quality dimensions), and a process component (effectiveness), including different psychological mediators (see Fig. 1).

The structural part of the model, distinguishing the three dimensions, could be elaborated further by including assumed relations between the dimensions (see also Sections 4.2 and 4.6). Conceptually, some overlap exists across dimensions with regard to the specific sub-dimensions used to conceptualize the dimensions (for details, see Praetorius et al., 2018). For example, the sub-dimension of support of competence experience (e. g., constructive approach to errors, differentiation and adaptive support, as well as constructive feedback) may contribute not only to the dimension of student support but also to cognitive activation.

The process part of the model takes up the idea of the model of opportunities and uses of instruction (e. g., Fend, 1998; see also Vieluf, Praetorius, Rakoczy, Kleinknecht & Pietsch, this issue): On the basis of different theories of learning and motivation, assumed mediating processes (i. e., the use of opportunities by students) between the teaching quality dimensions (i. e., the opportunities provided by the teacher) and student outcomes are included. The assumed effects of the three dimensions on student outcomes could, however, be described more specifically in terms of how these effects are assumed to work, at both individual and group levels (see Vieluf et al., this issue).

Additionally, some relations may be missing as, for example, cognitive activation might also have an effect on experience of competence.

4.2 Are the Statements Arranged in a Hierarchical or Systematic Order?

No hierarchy or order has been suggested explicitly for the TBD. In some publications, however, classroom management is described as the prerequisite for other aspects of teaching quality (e. g., Brunner, 2018; Klieme, Schümer & Knoll, 2001). This also fits well with Openshaw and Clarke's (1970) suggestion to distinguish three levels of teaching acts: those that set the stage for learning (i. e., classroom management and student support), those that are at the core of learning (i. e., cognitive activation), and those that appraise the process and the product (this last level is largely missing from the TBD conception).

4.3 How Explicitly is TBD Related to Existing Attempts to Capture a Theory of Teaching?

Klieme et al. (2001) related TBD to the three requirements of classroom teaching distinguished by Diederich and Tenorth (1997): student attentiveness, student motivation, and student understanding. This approach therefore can be seen as an initial theoretical foundation for TBD. In further developments of TBD, motivational and learning theories were added to this understanding (see Section 2).

TBD has also been connected to other frameworks or models that conceptualize teaching quality. This is particularly true for the Classroom Assessment Scoring System (CLASS; see Pianta & Hamre, 2009) with its three teaching dimensions: classroom organization (e. g., behavior management), emotional support (e. g., positive climate), and instructional support (e. g., content understanding). Without explicitly describing similarities and differences between TBD and CLASS, some publications seem to largely equate the dimensions included in both (e. g., Decristan et al., 2015; Fauth, Decristan, Rieser, Klieme & Büttner, 2014; Praetorius et al., 2017; Taut & Rakoczy, 2016). If both are indeed capturing exactly the same dimensions of teaching quality, one would need to critically challenge whether we need both. As the aspects covered in each of the dimensions are not structured and named in the same way, direct comparison of the two approaches is difficult. Through the synthesis of several teaching dimensions that resulted from considering different frameworks and models, including TBD and CLASS, Praetorius and Charalambous (2018), showed differences between the two frameworks/models: whereas TBD covers 10 of the 20 sub-dimensions of teaching quality in this synthesis, CLASS covers 15, and groups some of them differently. We therefore see value in better understanding the degree of overlap between different frameworks or models, as well in comparing the empirical support for the different assumptions (see also Sections 4.8 and 4.9).

4.4 Are the Statements Parsimonious and Comprehensive?

Whereas TBD was described as parsimonious and comprehensive in the initial publication by Klieme et al. (2001), subsequently it has been emphasized that it is parsimonious but likely not comprehensive. Lipowsky and Bleck (2019), for example, have claimed that a fourth dimension, subject matter quality (in their case, mathematics), needs to be added. Nilsen and Gustafsson (2016) enhanced the TBD model with a dimension called clarity, while Kleickmann, Steffensky, and Praetorius (this issue) established a dimension called cognitive support, which they distinguish from motivational support. Such claims receive further support from the synthesis by Praetorius and Charalambous (2018) across TBD and 11 other commonly used observational frameworks or models for describing teaching quality. Of the seven dimensions distinguished in the synthesis, one is not covered at all (i. e., support of practicing); of the 20 sub-dimensions across the seven dimensions, ten are missing in TBD (two generic, e. g., "teacher regularly checks for students' understanding"; three content-specific, e. g., "selecting worthwhile and developmentally appropriate content"; and five correspond to the interaction of generic and content-specific aspects, e. g., "presenting the content in a structured way"[9]). One could argue that at least the generic aspects should be included in TBD, to enable a comprehensive generic view on teaching quality.

4.5 Are the Statements Clearly Defined and Can They Be Tested?

For being able to test statements, explicit and clear hypotheses as well as clearly defined statements are necessary.

For TBD, such explicit and clear hypotheses have been formulated (e. g., the assumed effects of these dimensions on student achievement and student motivation; see Section 2, see also Fig. 1).

Clearly defined statements involve proper and agreed upon definitions and operationalizations of the constructs in which we are interested. However, in the case of TBD, the respective literature does not convey a consistent definition of its main elements and we find great differences in the operationalizations across studies (see Praetorius et al., 2018). This is even the case for classroom management, which is often considered a dimension in which the research community has developed a common understanding over the last decades. Some studies use rather narrow operationalizations (focusing exclusively on disruptions or effective use of time, e. g., Fauth et al., 2014, whereas others are broader (focusing also on aspects such as monitoring or clear rules; Lenske et al., 2016). This variability exists to an even larger extent for cognitive activation and student support. Here, narrow foci on challenging tasks and questions as well

9 In the initial publication on TBD by Klieme et al. (2001), structuredness was included as part of classroom management. In later publications (e. g., Lipowsky et al., 2009), classroom management was focused on time and behavior management.

as on exploration of the students' way of thinking exist for cognitive activation (e. g., Fauth et al., 2014) or on teacher-student relationships and differentiation for student support, respectively (e. g., Praetorius, Vieluf, Saß, Bernholt & Klieme, 2016). Broader foci also include discursive and co-constructive learning or the support of metacognition for cognitive activation (e. g., Korneck et al., 2017) or encompass constructive feedback and choice options for student support, respectively (e. g., Rakoczy & Pauli, 2006). Another question that needs to be addressed more clearly in this context is which perspectives (e. g., observer, student, and teacher ratings) are best suited to capturing each of the (sub)dimensions (see also Fauth, Göllner, Lenske, Praetorius & Wagner, this issue).

4.6 Are the Statements Quantitatively or Qualitatively Related?

The degree to which the statements of a theory are quantitatively and/or qualitatively related is an important prerequisite for the testability of theories (see Section 4.5). For example, although the three basic dimensions are partly positively correlated in empirical studies (e. g., Fauth et al., 2014; Kleickmann et al., this issue; Kunter et al., 2013), publications usually do not include explicit statements on expected relations. Such statements would, however, enhance the possibilities for validating the TBD model.

The non-deterministic relations of specific teaching dimensions and student outcomes are highlighted in the TBD by including students' uses of opportunities (e. g., time on task, depth of processing; see Fig. 1). What could be investigated in more detail is the nature of these relations – for instance, whether they are linear or non-linear (for an initial investigation, see Klieme et al., 2001) or whether an optimum or the maximum level of a teaching characteristic is most conducive for student learning (e. g., Brunner, 2018; Marzano & Marzano, 2003; Puntambekar & Hübscher, 2005).

4.7 Are the Boundaries Stated?

Although TBD is assumed to be broadly applicable to all age levels, school subjects, school forms, and possibly even countries (e. g., Klieme et al., 2001), boundaries can be identified. For example, despite the suggestion that research be value neutral (e. g., Seidel et al., 2014; Reiss & Sprenger, 2017), social research focuses on social phenomena that are per se intertwined with values and socio-cultural aspects in complex ways. International comparative research, for instance, has shown teaching and its effects on student outcomes to be influenced by culture (e. g., Bellens, van Damme, van den Noortgate, Wendt & Nilsen, 2019; Clarke, 2013; Stigler & Hiebert, 1999). It is evident that a theory of teaching cannot evade the complexities arising from cultural-societal contexts. Hammersley (2000) discusses that the only way of dealing with the dilemma is that researchers pay close attention to the values and socio-cultural assumptions implicated in their work and make them explicit. Thus, one could enhance the

explicitness of the TBD with respect to its boundaries by elaborating on the underlying socio-cultural assumptions more explicitly (for an example, see Fischer, Praetorius & Klieme, 2019).

The boundaries of TBD could also be made more explicit with respect to its capability to model content-specific teaching aspects. In general, TBD aims to cover generic aspects of teaching quality. Content-specific aspects therefore are explicitly excluded (e. g., accuracy of the content taught; see for example Charalambous & Litke, 2018). However, due to the fact that most studies on TBD have been tied to mathematics (see Section 4.9), as well as the challenge that cognitive activation is closely intertwined with the content taught, the focus on only generic aspects of teaching quality is not as clear-cut as one might think.

4.8 Are the Statements Supported Empirically?

Models of teaching quality are often highly complex, considering a large number of aspects that are interrelated. Hence it is not possible to generate empirical data in a single study that is suited to support such models as a whole. Empirical support therefore typically pertains to sub-parts, and researchers aim to accumulate evidence across several studies to eventually come close to a complete picture. The structural assumption that three dimensions can be distinguished has been supported by several studies using factor analysis (for an overview, see Praetorius et al., 2018). The fact that these studies differed in regard to the instruments and operationalizations used, could be seen as additional support for the model, as it is robust in this way. However, in most studies the decision-making with respect to selecting or developing the specific scales used for measuring TBD, is not explained. Therefore, one could question whether such decisions were conceptually-driven prior to data collection, or whether they resulted from a data-informed a posteriori decision process – a problem that, of course applies not only to research on TBD but to empirical studies in general without an open science approach (see e. g., Open Science Collaboration, 2017).

An important assumption in the initial publication on TBD by Klieme et al. (2001) – one that is, however, not explicitly stated in later publications – was that the three dimensions are comprehensive for capturing teaching quality. One way of testing this assumption is by testing the increment in predictive value of student outcomes when taking into account additional teaching aspects. Existing studies indicate the need to include aspects of teaching quality that go beyond TBD (see Section 4.4).

Furthermore, the assumption that all three dimensions have effects on student outcomes could only be confirmed for approximately half of the findings, according to a review of the current empirical literature that included only multi-level longitudinal designs, as these allow methodologically proper testing of such effects (see Praetorius et al., 2018). These inconsistent findings should be discussed, and should inform a revision of TBD in future publications. Such revisions have been suggested by Klieme in several conference presentations, although these suggestions have not as yet been

published. In 2012 for example, Klieme proposed a six-factor model, splitting each of the basic dimensions into a behavioral/interactive factor and a content-related factor (Klieme, 2012).

4.9 Are There Yet Untested Main Statements?

One of the assumptions of TBD is that the three dimensions are relevant for all subjects, grade levels, school forms and, potentially, different countries. Of the 21 empirical studies provided in the overview by Praetorius et al. (2018), the majority were conducted in mathematics (12 studies), with the remaining studies being distributed between German (5 studies), science education/physics (4 studies), English as a Foreign Language (2 studies), accounting (1 study), or across all subjects (3 studies).[10] This points to the need to test TBD more frequently in subjects other than mathematics and science education (e. g., language, social sciences, or musical/aesthetic subjects).

These studies mainly took place in lower secondary grades (9 studies), some additionally covering the lower part of upper secondary school (6 studies), whereas studies in primary school (4 studies) and those spread across all grades (2 studies) were rather the exception. This variation shows that evidence is needed on TBD in upper secondary grades and primary school.

TBD has been implemented in international studies such as PISA and TIMSS (see Section 2), and its dimensional structure, as well as the cross-sectional relations with outcomes, provides initial support for its applicability on an international level. It is important to note, however, that these studies also partly indicate that the dimensions might not be interpretable in the same way across countries (e. g., Fischer et al., 2019; see also Section 4.7).

Finally, the more complex assumptions on mediating processes between TBD and student outcomes – such as students' time on task for classroom management, students' experiences of autonomy, competence, and social relatedness for student support, as well as students' depth of processing for cognitive activation – have been stated theoretically (Klieme, Lipowsky, Rakoczy & Ratzka, 2006), but so far have rarely been investigated empirically (exceptions are, for example, Helm, 2016, as well as Rakoczy et al., 2019). The reasons for this, among others, include the lack of clarity with respect to how to conceptualize the mediating processes (see Vieluf et al., this issue), and how to develop reliable and valid measures capturing student thinking and attentiveness (e. g., using experience sampling methods).

10 The sum of the named studies exceeds 21, as some studies covered several subject matters.

4.10 Are There Yet Untested Main Statements?

According to Kane and Marsh (1980), theories of teaching quality should not only fulfill certain criteria for research purposes, but they should also contain, or clearly imply, statements that are useful for guiding and improving teaching practice. The purpose of TBD has been to serve both research and professional practice by providing an analytic tool on how to conceptualize teaching quality and its effects on student outcomes. Although the model was not developed with the goal of offering concrete directions for practice, it was meant to stimulate professional learning: TBD provides a general heuristic that may guide feedback and reflection on teaching. In fact, even the initial publication on TBD by Klieme et al. (2001) was accompanied by a CD-ROM providing material for professional training. Further research projects using TBD have produced professional development materials on cognitive activation for teachers (see Krammer et al., 2006). Furthermore, in educational policy (e. g., school evaluation and inspection systems in Germany), TBD has been used as an input in the development of standards describing quality teaching, and for the validation of classroom observation protocols (Taut & Rakoczy, 2016).

5. Conclusion

"[P]leas for better theory fall on receptive ears but recalcitrant hands. Everyone agrees that our theories should be stronger, as long as it does not require us to do anything differently" (Sutton & Staw, 1995, p. 378). Although developing stronger theories represents an arduous task, we believe it is time to take action so as to bring the field of teaching quality forward. The answers to the questions above provide some initial ideas on how to foster theory development for TBD specifically; yet the analytic process outlined above can also be transferred to other teaching models and frameworks. A summary of the most important needs for action might read as follows: (a) defining and summarizing the main statements of a model or framework in a testable, parsimonious but comprehensive way; (b) providing an overview of the extent to which the statements are empirically supported, and revising them in cases where they are not; (c) relating these single statements logically, theoretically, and internally consistently to each other, as well as arranging them hierarchically or systematically; (d) clearly stating the boundaries of the model or framework; (e) being more explicit on how the model or framework of interest is related to other existing approaches of conceptualizing teaching quality; and eventually (f) ensuring its relevance and potential for improving teaching practice.

It is hoped that in the years to come scholars will undertake this work more consistently and collaboratively, in producing stronger theories of teaching quality that, in turn, can help us better understand how teaching contributes to student learning.

References

Ball, D. L., & Forzani, F. M. (2009). The work on teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47,* 133–180.

Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O., & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde.* Wiesbaden: Springer.

Beck, K., & Krapp, A. (2006). Wissenschaftstheoretische Grundfragen der Pädagogischen Psychologie. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 33–73). Weinheim: Beltz.

Bellens, K., van Damme, J., van den Noortgate, W., Wendt, H., & Nilsen, T. (2019). Instructional quality: Catalyst or pitfall in educational systems' aim for high achievement and equity? An answer based on multilevel SEM analyses of TIMSS 2015 data in Flanders (Belgium), Germany, and Norway. *Large-Scale Assessments in Education, 7,* 1–27.

Berliner, D. C. (2009). Review regarding the book. In N. L. Gage (Eds.), *A conception of teaching* (see back of the book). New York: Springer Science and + Business Media.

Biddle, B. J., & Anderson, D. S. (1986). Theory, methods, knowledge, and research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 230–252). New York: MacMillan.

Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik, 39*(2), 257–284.

Cappella, E., Aber, J. L., & Kim, H. K. (2016). Teaching beyond achievement tests: Perspectives from developmental and education science. In D. H. Gitomer, & C. A. Bell (Eds.), *Handbook of research on teaching* (pp. 249–348). Washington, DC: American Education Research Association.

Carroll, J. B. (1963). A primer of programmed instruction in foreign language teaching. *International Review of Applied Linguistics in Language Teaching, 1*(2), 115–141.

Chalmers, A. F. (2007). *Wege der Wissenschaft: Einführung in die Wissenschaftstheorie* (6. Band). Berlin: Springer.

Charalambous, C. Y., & Litke, E. (2018). Studying instructional quality by using a content-specific lens: The case of the Mathematical Quality of Instruction framework. *ZDM Mathematics Education, 50,* 445–460.

Clarke, D. J. (2013). International comparative research into educational interaction: Constructing and concealing difference. In K. Tirri & E. Kuusisto (Eds.), *Interaction in educational settings* (pp. 5–22). Rotterdam: Sense Publishers.

Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität.* Münster: Waxmann.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119–142.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools.* London/New York: Routledge.

Danielson, C. (2013) *The framework for teaching evaluation instrument.* www.danielsongroup.org [24. 09. 2019].

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., Hondrich, A. L., Rieser, S., Hertel, S., & Hardy, I. (2015). Embedded formative assessment and classroom process quality: How do they interact in promoting students' science understanding? *American Educational Research Journal, 52*(6), 1133–1159.

Diederich, J., & Tenorth, H.-E. (1997). *Theorie der Schule: Ein Studienbuch zu Geschichte, Funktionen und Gestaltung.* Berlin: Cornelsen Scriptor.

Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit.* Münster: Waxmann.

Eder, F. (1996). *Schul- und Klassenklima. Ausprägung, Determinanten und Wirkungen des Klimas an höheren Schulen.* Innsbruck: Studien Verlag.

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.

Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistungen.* Weinheim/München: Juventa.

Fischer, J., Praetorius, A.-K., & Klieme, E. (2019). *The impact of linguistic similarity on cross-cultural comparability of students' perceptions of teaching quality. Educational Assessment, Evaluation and Accountability.* https://www.springerprofessional.de/the-impact-of-linguistic-similarity-on-cross-cultural-comparabil/16658412 [24. 09. 2019].

Gage, N. L. (2009). *A conception of teaching.* New York: Springer Science and + Business Media.

Gitomer, D. H., & Bell, C. A. (2016) (Eds.). *Handbook of research on teaching* (5th ed.). Washington: American Educational Research Association.

Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung.* Münster: Waxmann.

Hammersley, M. (2000). Varieties of social research. *International Journal of Social Research Methodology, 3*(3), 221–229.

Helm, C. (2016). Zentrale Qualitätsdimensionen von Unterricht und ihre Effekte auf Schüleroutcomes im Fach Rechnungswesen. *Zeitschrift für Bildungsforschung, 6*(2), 101–119.

Hill, J. R., & Schrum, L. (2002). Theories on teaching: Why are they so hard to find?!? *TechTrends, 46*(5), 22–26.

Kane, R., & Marsh, C. J. (1980). Progress toward a general theory of instruction? *Educational Leadership,* 253–255.

Kerlinger, F. N. (1964). *Foundations of behavioral research: Educational and psychological inquiry.* New York: Holt, Rinehart and Winston.

Klieme, E. (2012, August). Qualities and effects of teaching. Integrating findings across subjects and cultures. *Keynote lecture at the EARLI SIG 18 conference,* Zurich.

Klieme, E. (2019). Unterrichtsqualität. In M. Gläser-Zikuda, M. Harring & C. Rohlfs (2018). *Handbuch Schulpädagogik* (S. 393–408). Münster: Waxmann.

Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht: Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts "Pythagoras". In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.

Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiss (Hrsg.), *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 333–359). Wiesbaden: VS Verlag für Sozialwissenschaften.

Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik: Outcomeorientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik, 54*(2), 222–227.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS-Impulse für Schule und Unterricht. Forschunsbefunde, Refominitiativen, Prasixberichte und Video-Dokumente* (S. 43–57). Bonn: Bundesministerium für Bildung und Forschung.

Korneck, F., Krüger, M., & Szogs, M. (2017). Professionswissen, Lehrerüberzeugungen und Unterrichtsqualität angehender Physiklehrkräfte unterschiedlicher Schulformen. In E. Sumfleth & H. Fischler (Hrsg.), *Professionelle Kompetenzen von Lehrkräften der Chemie und Physik. Studien zum Physik- und Chemielernen Bd. 200* (S. 113–133). Berlin: Logos.

Krammer, K., Ratzka, N., Klieme, E., Lipowsky, F., Pauli, C., & Reusser, K. (2006). Learning with classroom videos: Conception and first results of an online teacher-training program. *ZDM Mathematics Education, 38*(5), 422–432.

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, 20*(2), 61–98.

Kuhn, T.S. (1969/1997). *Die Struktur wissenschaftlicher Revolutionen* (14. Aufl.). Frankfurt a.M.: Suhrkamp.

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.

Kunter, M., Klusmann, U., Baumert, S., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. https://doi.org/10.1037/a0032583.

Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts.* Paderborn: Ferdinand Schöningh.

Kunter, M., & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85–113). Münster: Waxmann.

Lakatos, I. (1977). *The methodology of scientific research programmes: Philosophical papers Volume 1.* Cambridge: Cambridge University Press.

Lenske, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., Liepertz, S., & Leutner, S. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht. *Zeitschrift für Erziehungswissenschaft, 19,* 211–233.

Leplin, J. (1980). The role of models in theory construction. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 267–283). Dordrecht: Springer.

Lipowsky, F., Rakoczy, K., Drollinger-Vetter, B., Klieme, E., Reusser, K., & Pauli, C. (2009). Quality of geometry instruction and its short-term impact on students' understanding of Pythagorean Theorem. *Learning and Instruction, 19*(6), 527–537.

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.

Lipowsky, F., & Bleck, V. (2019). Was wissen wir über guten Unterricht? – Ein Update. In U. Steffens & R. Messner (Hrsg.), *Unterrichtsqualität. Konzepte und Bilanzen gelingenden Lehrens und Lernens. Grundlagen der Qualität von Schule* (Bd. 3, S. 219–249). Münster: Waxmann.

Lüders, M. (2014). Erziehungswissenschaftliche Unterrichtstheorien. *Zeitschrift für Pädagogik, 60*(6), 832–849.

Marzano, R. J., & Marzano, J. S. (2003). The key to classroom management. *Educational Leadership, 61*(1), 6–13.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256.

Nilsen, T., & Gustafsson, J.-E. (Eds.) (2016). *Teacher quality, instructional quality and student outcomes*. Heidelberg: Springer.

Oelkers, (2000). Anmerkungen zur Reflexion von "Unterricht" in der deutschsprachigen Pädagogik des 20. Jahrhundert. In D. Benner & H. E. Tenorth (Hrsg.), *Bildungsprozesse und Erziehungsverhältnisse im 20. Jahrhundert* (42. Beiheft der Zeitschrift für Pädagogik S. 166–185). Weinheim: Beltz.

Open Science Collaboration (2017). Maximizing the reproducibility of your research. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 1–21). New York: Wiley.

Openshaw, K., & Clarke, S. C. T. (1970). General teaching theory. *Journal of Teacher Education, 21*(3), 403–416.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49,* 345–375.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38,* 109–119.

Popper, K. R. (2005). *Logik der Forschung. Gesammelte Werke* (Band 3). Tübingen: Mohr Siebeck.

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.

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM Mathematics Education, 50*(3), 407–426.

Praetorius, A.-K., Lauermann, F., Klassen, R. M., Dickhäuser, O., Janke, S., & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and Teacher Education, 65,* 241–254.

Praetorius, A.-K., & Charalambous, C. Y. (2018) Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM Mathematics Education, 50*(3), 535–553.

Prange, K. A. (2005). Didaktik. In J. Kade, W. Helsper, C. Lüders, B. Egloff, F.-O. Radtke & W. Thole (Hrsg.), *Pädagogisches Wissen* (S. 183–188). Stuttgart: Kohlhammer.

Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex environment: What have we gained and what have we missed? *Educational Psychologist, 40,* 1–12.

Rakoczy, K., Klieme, E., Lipowsky, F., & Drollinger-Vetter, B. (2010). Strukturierung, kognitive Aktivität und Leistungsentwicklung im Mathematikunterricht. *Unterrichtswissenschaft, 38*(3), 229–246.

Rakoczy, K., & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis". Teil 3: Videoanalysen. Materialien zur Bildungsforschung* (S. 206–233). Frankfurt a. M.: GFPF.

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction, 60,* 154–165.

Reckwitz, A. (2002) Toward a theory of social practices. A development in culturalist theorizing. *European Journal of Social Theory, 52,* 245–265.

Reiss, J., & Sprenger, J. (2017). *Scientific objectivity. The Stanford encyclopedia of philosophy* (Winter 2017 Edition). https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/ [24.09.2019].

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Eds.), *Second handbook of research on teaching* (pp. 122–183). Chicago: Rand McNally.

Rosenshine, B. (2009). A conception of teaching. Review of N.L. Gage. Springer, Norwell, MA (2009). *Teaching and Teacher Education, 25,* 1169–1171.

Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and wellbeing. *American Psychologist, 55*(1), 68–78.

Scriven, M. (1994). A possible distinction between traditional scientific disciplines and the study of human behavior. In M. Martin & L.C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 71–78). Cambridge: The MIT Press.

Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie: Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik, 60*(6), 850–866.

Seidel, T., Prenzel, M., & Krapp, A. (2014). Grundlagen der Pädagogischen Psychologie. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (S. 21–36). Weinheim: Beltz.

Snow, R.E. (1973). Theory construction for research on teaching. In R.M.W. Travers (Ed.), *Second handbook of research on teaching* (pp. 77–112). Chicago, IL: Rand McNally.

Stigler, J.W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom.* New York: Free Press.

Stigler, J.W., & Knoll, S. (1999). The TIMSS Videotape Classroom Study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States. *Education Statistics Quarterly, 1,* 109–112.

Sutton, R.I., & Staw, B.M. (1995). What theory is not. *Administrative Science Quarterly, 40*(3), 371–384.

Taut, S., & Rakoczy, K. (2016). Observing instructional quality in the context of school evaluation. *Learning and Instruction, 46,* 45–60.

Terhart, E. (2014). Unterrichtstheorie: Einführung in den Thementeil. *Zeitschrift für Pädagogik, 60*(6), 813–816.

Thagard, P.R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy, 75*(2), 76–92.

Wagenschein, M. (1992). *Verstehen lehren. Genetisch – Sokratisch – Exemplarisch* (9. ed.). Weinheim: Beltz.

Wang, M.C., Haertel, G.D., & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249–294.

Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik. Ein Lehrbuch zur Psychologischen Methodenlehre.* Göttingen: Hogrefe.

Whetten, D.A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490–495.

Zusammenfassung: Der vorliegende Beitrag fokussiert auf die Bedeutung von Theorien zu Unterricht und dessen Qualität für die quantitative empirische Unterrichtsforschung. Dabei wird zunächst der quantitative empirische Forschungsansatz vorgestellt. Anschließend stellen wir die Herkunft und den aktuellen Status eines populären Modells in der deutschsprachigen quantitativen Unterrichtsforschung dar, dem Modell der drei Basisdimensionen von Unterrichtsqualität. Es folgt eine Reflexion in welchem Ausmaß dieses Modell Kriterien guter Theorien erfüllt. Abschließend werden Schlussfolgerungen für zukünftige Forschung mit einem Fokus auf die Entwicklung von Theorien gezogen.

Schlagworte: Theorie, Unterrichtsqualität, drei Basisdimensionen, Unterricht, Modell

Contact

Prof. Dr. Anna-Katharina Praetorius, Universität Zürich,
Lehrstuhl für pädagogisch-psychologische Lehr-Lernforschung und Didaktik,
Institut für Erziehungswissenschaft,
Freiestrasse 36, 8032 Zürich, Schweiz
E-Mail: anna.praetorius@ife.uzh.ch

Prof. Dr. Eckhard Klieme, DIPF | Leibniz-Institut für Bildungsforschung
und Bildungsinformation,
Rostocker Straße 6, 60323 Frankfurt a. M., Deutschland
E-Mail: klieme@dipf.de

Prof. Dr. Thilo Kleickmann, Christian-Albrechts-Universität zu Kiel,
Institut für Pädagogik, Abteilung Schulpädagogik,
Olshausenstr. 75, 24118 Kiel, Deutschland
E-Mail: kleickmann@paedagogik.uni-kiel.de

Prof. Dr. Esther Brunner, Pädagogische Hochschule Thurgau
Unterer Schulweg 3, 8280 Kreuzlingen, Schweiz
E-Mail: esther.brunner@phtg.ch

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

Prof. Dr. Sandy Taut, Bayerisches Landesamt für Schule,
Qualitätsagentur,
Stuttgarter Straße 1, 91710 Gunzenhausen, Deutschland
E-Mail: sandy.taut@isb.bayern.de

Assistant Prof. Dr. Charalambos Charalambous, University of Cyprus,
Department of Education,
Theophanides Building,11 – 13 Dramas Str., 1077 Nicosia, Cyprus
E-Mail: cycharal@ucy.ac.cy