

Clausen, Marten

Commentary regarding the section "The role of different perspectives on the measurement of teaching quality"

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. 173-178. - (Zeitschrift für Pädagogik, Beiheft; 66)



Quellenangabe/ Reference:

Clausen, Marten: Commentary regarding the section "The role of different perspectives on the measurement of teaching quality" - 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. 173-178 - URN: urn:nbn:de:0111-pedocs-258721 - DOI: 10.25656/01:25872

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-258721>

<https://doi.org/10.25656/01:25872>

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@dipt.de
Internet: www.pedocs.de

66. Beiheft

April 2020

ZEITSCHRIFT FÜR PÄDAGOGIK

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

BELTZ JUVENTA

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

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

Kommentare

<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

Marten Clausen

Commentary Regarding the Section "The Role of Different Perspectives on the Measurement of Teaching Quality"

Abstract: In this commentary, the paper by Fauth, Göllner, Lenske, Praetorius, and Wagner as well as the paper by Göllner, Fauth, Lenske, Praetorius, and Wagner published in this special issue of the *Zeitschrift für Pädagogik* are discussed. In the context of specificity and perspective-agreement in perceptions of teaching quality, merits and possible limitations of theoretical conceptions and empirical analyses are reviewed for both papers. In a final section, the complexity inherent to teaching research is addressed on a more general level focussing on differences between psychometric and edumetric approaches.

Keywords: Teaching Quality, Perspective Agreement, Classroom Management, Item Wording, Ratings of Teaching

1. Introduction

In this discussion, I will comment on the papers by Fauth, Göllner, Lenske, Praetorius, and Wagner (in this issue) and Göllner, Fauth, Lenske, Praetorius, and Wagner (in this issue) presented in this issue of the *Zeitschrift für Pädagogik*. From a broader perspective, both papers deal with the relation between classroom management aspects of teaching quality and their operationalization and measurement in items and scales. I will first address some more specific aspects of both papers and then will discuss more general aspects that are relevant to both papers.

2. Who Sees What? Conceptual Considerations on the Measurement of Teaching Quality from Different Perspectives

In a comparison between the different perspectives of the teacher, his students, and external observers, Clausen (2002) found evidence for a low level of agreement of teaching quality ratings. In the following years, a growing number of studies could replicate these results. Similarly to Clausen (2002), several researchers (e. g. Kunter & Baumert, 2006) have moved away from the idea of a common true score and concluded that for perceptions of instruction, perspective-specific validity should be assumed. Clausen's theoretical approach to predicting agreement as an interaction of perspectives and teaching construct characteristics (observability, demands regarding didactic understanding, evaluativeness) has had limited success in respect of empirical validation, with observability being the only dimension for which some support has been found.

Fauth et al. (in this issue) propose a similar but different approach. Instead of classifying characteristics of teaching constructs to predict agreement, they look at the item level to determine the *referent* of an item (teacher, students, both/unclear) – the perspective that determines whose behavior is focused on by an item. If a perspective is the referent of an item, Fauth et al. (in this issue) conclude this entails a higher ego involvement and more and better information. They integrate their considerations in the *perspective reference matrix* which, as a theoretical framework, allows the prediction of agreement between various combinations of item-referent and perspective. By applying this framework to reinterpret the empirical results of other studies, they find some confirmation for referent-dependent variance in the factor structures of rating instruments, and in the correlations between perspectives in agreement studies.

The paper by Fauth et al. (in this issue) offers a very valuable critical approach to the measurement of classroom management as a central aspect of the quality of teaching and to the instruments recently and currently used in empirical teaching research. Their precise focus on item level is similar to the classic facet theoretic approach to item analysis (Guttman & Greenbaum, 1998), in which the relevant item aspects are analyzed on a theoretical level to build a basic item structure (mapping sentence). The elements of the mapping sentence are then tested empirically for their effects on item inter-correlations (e.g. Alt, 2018). Lenske (2016, p. 95) provides a list of facets that may generate specific variances in ratings. While the overall approach is very convincing to me, there are two smaller points specific to the paper that caught my attention.

(1) From my view, there is no clear information advantage for the teacher in regard to the example item “Our teacher immediately notices when students start doing something else.” Teachers can only remember those times when they noticed the behavior. Both the students and external observers have at least the possibility of seeing both cases: that is, including those cases where the teacher does not notice the off-task behavior.

(2) I am not convinced that for all student behavior items there is a higher ego-involvement for students compared to the ego-involvement of the teacher. Even if a construct or item focuses only on the behavior of the student, a quality of teaching approach would regard the teacher as implicitly responsible or at least to some degree in control. There is a difference between “The class does not behave in an orderly manner.” and “In mathematics lessons of Mr. Smith, the class does not behave in an orderly manner”. Furthermore, I assume that ego-involvement and the resulting self-serving bias have to be regarded in the context of the data collection. If a teacher and his students fill out ratings questionnaires for TIMSS or PISA, there is little chance that the answers will have direct consequences for themselves. If these assessments take place with feedback in the context of teacher evaluation or professional development, there may be a lot more weight on the answers because they are communicated and discussed; so that disparities between the ratings of the students and the view of their teacher may become obvious.

I do agree with the authors that the deviations between perspectives should be investigated. On the other hand, I am very much convinced that a common true score does not represent the social reality of classroom interaction. Idiosyncratic perspectives may

result from different logics of action. They are real, in the sense that they influence attitudes and behavior, as noted in the Thomas Theorem (Thomas & Thomas, 1928). External observers such as school inspectors transport, communicate and establish outside norms and standards into the school – they are not supposed to take the teacher's view, nor should they try to adopt a student perspective. Students should be free to state their perceptions, which may be mainly focused on being treated respectfully, being treated fairly, being entertained, and being motivated. Teachers have to develop a realistic self, and should use feedback from other perspectives as a means of professional development. To some degree they should learn to look at their teaching from a student perspective (Wettstein, Ramseier, Scherzinger & Gasser, 2016). However, this does not mean they should make the student perspective their own. To me, there is still potential to further explore social perspectives in teaching research by applying general concepts of social perspectivity (Strack, 2004).

3. Do Student Ratings of Classroom Management Tell us More About Teachers or About Classroom Composition?

The paper by Göllner et al. (in this issue) focuses on whether the difference in referent of an aspect of classroom management does generate variation that depends more on the composition of the class than on the actual teaching. In an elaborate statistical model using student ratings from PISA 2003, the authors test the assumption that aspects of classroom composition influence ratings of those aspects of classroom management that mainly refer to the students' behavior. Results indicate that teachers of classes with more low-achieving students and more male students are rated lower on aspects of classroom management that refer to the students, in contrast to measures referring to the teacher.

The paper by Göllner et al. (in this issue) has a sound theoretical background closely related to the theoretical approach of Fauth et al. (in this issue). The authors apply a methodologically sophisticated multi-level model to work out the importance of the specific referent-dependent variance of the "student referent items" for student ratings and their relations to academic achievement and classroom composition. From my point of view, the referent-dependent variance results from characteristics of general class behavior and its corresponding psychological dimension, classroom climate; both of which influence the specific behavior towards the teacher who is being rated. The classroom climate should stabilize this behavior at the class level. These results are of special importance when ratings are used in the context of teacher evaluation, because they might impair the fairness of the evaluation (Campbell & Ronfeldt, 2018).

4. General Comments

Aspects of teaching and classroom interaction are still being treated like psychological constructs, with unidimensionality and normal distribution being common assumptions. Looking at some possible “states” of classroom management along a very basic abstract continuum of “order vs. chaos”, it is obvious however that the continuum is not normally distributed. (1) A class that is very difficult – perhaps because of its composition – will make teaching hard for almost any teacher. (2) A teacher with great authority will control even the most difficult class. (3) With a class of committed and intrinsically motivated students, there is almost no need for classroom management. (4) A stable balanced state in the middle of the continuum has a relatively low probability. Disturbances and disciplinary problems that are not dealt with effectively have a tendency to escalate and spread by contagion. All of the different examples described here have different referents or combinations of referents.

As Fauth et al. (in this issue) and Göllner et al. (in this issue) point out, for typical aspects of classroom management the item referent varies on the item level, and the item referent influences both the internal structure and the correlations to external criteria. On the other hand, this is also true on a construct level. In their analyses of expert sortings of teaching constructs, Clausen, Schnabel & Schröder (2002, p. 254) generated a differentiation regarding the “subject” and “object” of a construct (in brackets):

- 1) students → (students): e. g., student-student relations
- 2) students → (teacher): e. g., student engagement
- 3) teacher → (students): e. g., monitoring and
- 4) teacher → (teacher): e. g., clarity and structure.

This differentiation is not a question of item formulation, but a question of the construct. At the core of teaching research, the constructs of quality of teaching represent a dynamic social interaction of teachers and students which is also stressed in the German “Angebot-Nutzungs-Modell” by Helmke (2003; roughly translated as “offer-use model” or “utilization-of-learning-opportunities” model). From this point of view, the most relevant constructs are the interactive dimensions that focus the students’ behavior towards the teacher (category 2) and the teacher’s behavior towards the students (category 3). For these constructs the reference perspective matrix does not make clear predictions. Reducing constructs to a “referent differentiated” level would avoid the referent-dependent variance problem, but would also reduce the behavioral universe to the less-relevant pure teacher variables and pure student variables. To put all referents into every single item, would make the resulting items more complex and more difficult to agree to. Balancing the different referents relevant to a construct across different items of a scale would not solve the referent problem – and yet in a way, this is how the instruments discussed by Fauth et al. (in this issue) work.

Edumetrics is not psychometrics – some of the complexity is inherent to the domain of teaching research, and thus cannot be solved easily. Classroom management is too

broad and too complex a field for one to be able to expect it to be homogeneous in terms of internal structure, as well as in terms of correlations to other constructs. Both papers focus mainly on behavior and classroom management, while the more cognitive, motivational, and emotional basic dimensions of teaching play marginal roles. If the authors were able to extend the theoretical scope of the reference perspective matrix, it would gain more relevance.

Updating the theoretical and empirical approaches to teaching research and perspective agreement, and connecting them closer to the Anglo-American research, is a major merit of both papers. The new approaches still need to be further elaborated and validated. The authors of both papers are well aware of this, and many of the points discussed in this comment paper are addressed at least to some degree in these papers. The quantitative dimensional approach to describing the quality of teaching has its limitations, and the more elaborate the methodology and the design, the more obvious are these limitations. Many of the items and scales discussed in Fauth et al. (in this issue) and Göllner et al. (in this issue) have their roots in instruments from the German classroom climate research of the 1980s and 1990s (e.g. Saldern & Littig, 1987). They were created with different theoretical and empirical frameworks in mind. This new generation of researchers has developed promising ideas and they are capable of testing them in elaborate designs with state-of-the-art methodology. I am very curious to see them apply these ideas to more recent data with modern multi-perspective instruments.

References

- Alt, D. (2018). Students' perceived constructivist learning environment. *European Journal of Psychological Assessment*, 34(6), 432–443.
- Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for? *American Educational Research Journal*, 55(6), 1233–1267.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Schnabel, K., & Schröder, S. (2002). Konstrukte der Unterrichtsqualität im Expertenurteil. *Unterrichtswissenschaft*, 30(3), 246–260.
- Guttman, R., & Greenbaum, C. W. (1998). Facet theory. *European Psychologist*, 3(1), 13–36.
- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Schulisches Qualitätsmanagement. Seelze: Kallmeyer.
- Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9(3), 231–251.
- Lenke, G. (2016). *Schülerfeedback in der Grundschule: Untersuchung zur Validität*. Münster: Waxmann.
- Saldern, M. v., & Littig, K.-E. (1987). *Landauer Skalen zum Sozialklima: 4.–13. Klassen. LASSO 4–13*. Weinheim: Beltz.
- Strack, M. (2004). *Sozialperspektivität – Theoretische Bezüge, Forschungsmethodik und wirtschaftliche Praktikabilität eines beziehungsdiagnostischen Konstrukts*. Göttingen: Universitätsverlag.
- Thomas, W. I., & Thomas, D. S. (1928). *The child in America*. New York: Knopf.
- Wettstein, A., Ramseier, E., Scherzinger, M., & Gasser, L. (2016). Unterrichtsstörungen aus Lehrer- und Schülersicht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48(4), 171–183.

Zusammenfassung: In diesem Kommentar werden die in diesem Themenheft der Zeitschrift für Pädagogik erschienenen Beiträge von Fauth, Göllner, Lenske, Praetorius und Wagner sowie Göllner, Fauth, Lenske, Praetorius und Wagner diskutiert. Vor dem Hintergrund der Frage von Spezifität und Perspektivenübereinstimmung von Unterrichtswahrnehmungen werden Verdienste und mögliche Einschränkungen der theoretischen Ansätze und empirischen Analysen beider Beiträge reflektiert. Im abschließenden Abschnitt wird die inherente Komplexität der Unterrichtsforschung auf allgemeinerer Ebene thematisiert mit Blick auf Unterschiede zwischen psychometrischen und edumetrischen Ansätzen.

Schlagworte: Unterrichtsqualität, Perspektivenübereinstimmung, Klassenführung, Itemformulierung, Unterrichtsbeurteilungen

Contact

Prof. Dr. Marten Clausen, Universität Duisburg-Essen,
AE Unterrichtsforschung,
Universitätsstr. 2, 45117 Essen, Deutschland
E-Mail: marten.clausen@uni-due.de