



# Fachspezifität von Unterrichtsqualität – aus der Perspektive der Mathematikdidaktik

Anika Dreher  · Timo Leuders

Eingegangen: 18. Januar 2021 / Überarbeitet: 30. April 2021 / Angenommen: 11. Mai 2021 / Online publiziert: 9. Juni 2021  
© Der/die Autor(en) 2021

**Zusammenfassung** Der Beitrag gibt einen kurzen Überblick über die Diskussion (und aktuelle Forschung) zur Fachspezifität von Unterrichtsqualität aus der Perspektive der Mathematikdidaktik. Sowohl theoretische als auch empirische Argumente verdeutlichen, dass bei einer Konzeptualisierung von Unterrichtsqualität fachspezifische Aspekte nicht außer Acht gelassen werden sollten. Die Berücksichtigung fachspezifischer Aspekte bezieht sich derzeit meist auf die Spezifizierung von Kriterien bei der Erfassung Unterrichtsqualität durch Ratings. Wir argumentieren jedoch, dass Kriterien dabei nicht die einzige relevante Varianzquelle darstellen, in der sich Fachspezifität niederschlägt. Um der Fachspezifität von Unterrichtsqualität systematisch auf die Spur zu kommen, schlagen wir daher eine mehrdimensionale Perspektive auf die Fachspezifität von Unterrichtsqualität als Grundlage für empirische Forschungsstrategien vor. Demnach zeigt sich die Fachspezifität von Unterrichtsqualität in unterschiedlichen Varianzquellen (Kriterien, Situationen, Ratern), die sich durch verschiedene Grade an Fachspezifität systematisch auf die eingeschätzte Unterrichtsqualität auswirken können. Insbesondere illustrieren wir anhand eines Beispiels, wie das fach- und inhaltsbezogene Wissen von Ratern (und damit die Fachspezifität von Ratern) ihre Interpretationen von Kriterien und Unterrichtssituationen beeinflussen und sich systematisch in ihren Einschätzungen von Unterrichtsqualität niederschlagen können.

**Schlüsselwörter** Unterrichtsqualität · Fachspezifität · Varianzquellen · Rater · Noticing

---

Anika Dreher (✉) · Timo Leuders  
Pädagogische Hochschule Freiburg, Kunzenweg 21, 79117 Freiburg, Deutschland  
E-Mail: [anika.dreher@ph-freiburg.de](mailto:anika.dreher@ph-freiburg.de)

Timo Leuders  
E-Mail: [leuders@ph-freiburg.de](mailto:leuders@ph-freiburg.de)

## Subject-specificity of instructional quality—From the perspective of mathematics education

**Abstract** This article provides a brief overview of the discussion (and current research) on the subject-specificity of instructional quality from the perspective of mathematics education research. By drawing on theoretical and empirical arguments, we illustrate that subject-specific aspects should not be disregarded in a conceptualization of instructional quality. Currently, the consideration of subject-specific aspects mostly refers to the specification of criteria in assessing instructional quality by ratings. However, we argue that criteria are not the only relevant source of variance in which subject-specificity is reflected. In order to trace the subject-specificity of teaching quality systematically, we therefore propose a multidimensional perspective on the subject-specificity of teaching quality as a basis for empirical research strategies. Accordingly, the subject-specificity of instructional quality manifests itself in different sources of variance (criteria, situations, raters) that can affect the assessed instructional quality systematically by different degrees of subject-specificity. In particular, we use an example to illustrate how raters' subject- and content-related knowledge (and thus raters' subject specificity) can influence their interpretations of criteria and instructional situations and thus be reflected systematically in their assessments of instructional quality.

**Keywords** Instructional quality · Subject-specificity · Sources of variance · Raters · Noticing

### 1 Die Diskussion der Fachspezifität von Unterrichtsqualität in der Mathematikdidaktik

Die empirische Unterrichtsforschung nahm den Mathematikunterricht bereits häufig in den Blick und konzeptualisierte seine Qualität insbesondere im deutschsprachigen Raum mit Hilfe der weitgehend fachunspezifischen drei Basisdimensionen (z. B. Praetorius et al. 2018). Es wird jedoch vielfach darauf hingewiesen, dass die Qualität von Mathematikunterricht mit generischen Dimensionen nicht hinreichend gut erfasst werden kann (Praetorius und Charalambous 2018). Illustriert wird dieses Phänomen häufig dadurch, dass einer aus mathematikdidaktischer Sicht unzureichenden – beispielsweise fachlich nicht korrekten – Unterrichtsstunde auf der Basis generischer Instrumente eine gute Unterrichtsqualität beigemessen wird (Brunner 2018; Learning Mathematics for Teaching Project 2011). Mathematikdidaktiker\*innen betonen die fachliche Korrektheit daher oft als wichtiges ergänzendes Kriterium (z. B. Brunner 2018; Schlesinger et al. 2018). Darüber hinaus wird die Relevanz von fachdidaktischen Kriterien für Instrumente zur Erfassung von Unterrichtsqualität in den Blick genommen. Ein bekanntes Beispiel dafür ist das Pythagoras-Projekt, in dem die Forschungsgruppe ein Konstrukt „fachdidaktische Unterrichtsqualität“ über die drei Subdimensionen *Vorkommen sogenannter Verstehenselemente*, *die Qualität von Repräsentationsformen* und *strukturelle Klarheit* konzeptualisierte. Dieses Konstrukt konnte empirisch von den drei Basisdimensionen getrennt werden (Lipowsky et al.

2018) und erwies sich als relevant für mathematischen Lernerfolg (Drollinger-Vetter 2011).

Aufgrund der Kritik an ausschließlich generischen Unterrichtsqualitätsmerkmalen entwickelten Mathematikdidaktiker\*innen stärker fachspezifische Instrumente, indem sie teilweise zusätzliche Dimensionen ergänzten (z. B. Schlesinger et al. 2018), teilweise aber auch fachspezifische Ausdifferenzierungen der generischen Dimensionen vornahm. Dies gilt insbesondere für die Basisdimension der kognitiven Aktivierung, die für verschiedene Studien auf unterschiedliche Weisen konzeptualisiert wurde (Leuders und Holzäpfel 2011; Praetorius et al. 2018), aber auch für weitere generische Kriterien, wie Feedback geben. Die Forschungsgruppe um Hill und Ball differenzierte beispielsweise „responding to students appropriately“ im Sinne von mathematikdidaktisch angemessenen Reaktionen z. B. auf Lernendenfehler fachspezifisch aus (Hill et al. 2008). Viel beachtete fachspezifische Instrumente wurden dabei vor allem in den USA entwickelt (z. B. MQI, vgl. Hill et al. 2008; TRU, vgl. Schoenfeld 2018). Da diese und weitere Instrumente aus der internationalen mathematikdidaktischen Community Grundlage für das Syntheseframework von Praetorius und Charalambous (2018) waren, sind die in diesem Framework enthaltenen Qualitätskriterien durchaus anschlussfähig an mathematikdidaktische Perspektiven. Auch die in Praetorius et al. (2020) vorgeschlagenen Ergänzungen aus Sicht anderer Fachdidaktiken finden sich teilweise im mathematikdidaktischen Diskurs wieder. So unterrichten auch Mathematiklehrkräfte Lerninhalte im Sinne des Spiralcurriculums, was eine *adäquate didaktische Reduktion unter Berücksichtigung zukünftiger fachlicher Lernschritte* erfordert (z. B. Dreher et al. 2018). Andere Ansätze betonen die *Diskursqualität* als Qualitätskriterium für Mathematikunterricht (z. B. Nowinska 2016).

Mathematikdidaktiker\*innen diskutieren Fachspezifität von Unterrichtsqualität jedoch nicht nur in Bezug auf Instrumente und Kriterien, sondern auch hinsichtlich der Frage, wie viel fachspezifisches Wissen notwendig ist, um die Qualität von Unterricht adäquat beurteilen zu können. So wird insbesondere angezweifelt, dass Unterrichtsqualität von fachfremden Ratern adäquat beurteilt werden kann (z. B. Brunner 2018; Lindmeier und Heinze 2020). Hierzu gibt es allerdings unseres Wissens noch keine spezifische Forschung. Einen Vorschlag, wie man diesbezüglich in Theorie und Empirie systematisch vorgehen kann, diskutieren wir im folgenden Abschnitt.

## 2 Fachspezifität von Unterrichtsqualität: Spezifität von Kriterien, Situationen und Ratern

Die theoretischen und empirischen Argumente des vorigen Abschnitts verdeutlichen, dass bei einer Konzeptualisierung von Unterrichtsqualität fachspezifische Aspekte nicht außer Acht gelassen werden können. Die Berücksichtigung fachspezifischer Aspekte von Unterrichtsqualität bezieht sich zwar meist auf die Spezifität von Kriterien, doch wir vermuten, dass Kriterien nicht die einzige relevante Varianzquelle bei der Erfassung von Unterrichtsqualität durch Ratings darstellen, in der sich Fachspezifität niederschlägt. Um der Fachspezifität von Unterrichtsqualität systematisch auf

die Spur zu kommen, nehmen wir daher im Folgenden eine mehrdimensionale Perspektive auf die Fachspezifität von Unterrichtsqualität als Grundlage für empirische Forschungsstrategien ein.

Bestehende Forschung zur Erfassung von Unterrichtsqualität durch Ratings zielt im Wesentlichen darauf ab, das Verhältnis von Kriterien, Unterrichtssituationen und Ratern zu klären, beispielsweise durch Fragen nach der Varianz oder Stabilität der erfassten Unterrichtsqualität in Abhängigkeit von variierenden Unterrichtssituationen. Dabei zeigte sich, dass sich die Variation der betrachteten Situationen zwischen verschiedenen Unterrichtsstunden – und auch zwischen verschiedenen Phasen innerhalb von Unterrichtsstunden – insbesondere auf die durch bestimmte fachspezifische Kriterien erfasste Einschätzung von Unterrichtsqualität auswirkt (Jentsch et al. 2020; Praetorius et al. 2014). Jentsch et al. (2020) argumentieren, dass dies nicht schlicht als Messfehler anzusehen ist, sondern dass Varianz in den Unterrichtssituationen liegt, die insbesondere durch die jeweiligen Unterrichtsinhalte und -ziele bedingt ist. So können Unterrichtssituationen beispielsweise aufgrund ihrer Ziele unterschiedlich stark fach- bzw. inhaltspezifisch sein, abhängig davon, ob eher fachunabhängige Kompetenzen, übergreifende fachspezifische Kompetenzen (z. B. mathematisches Beweisen) oder die Beherrschung von inhaltspezifischen Konzepten und Prozeduren im Vordergrund stehen. Letzteres kann zum Beispiel bedeuten, dass in einer Situation zentral ist, Lernende dabei zu unterstützen, anschlussfähige Grundvorstellungen für ein mathematisches Konzept, wie den Unbestimmtenaspekt einer Variablen (d. h. eine Variable steht für eine beliebige Zahl) auszubilden. Dies erfordert die Verwendung von in der Fachdidaktik empirisch fundierten inhaltspezifischen Lehrkonzepten (z. B. Nutzung bestimmter Repräsentationen oder Erklärungen), was sich dann wiederum in fach- und inhaltspezifischen Kriterien für Unterrichtsqualität niederschlägt.

Umgekehrt untersuchen Mathematikdidaktiker\*innen auch (wenn auch meist nicht systematisch), wie sich die Variation von Kriterien (mehr oder weniger fachspezifisch) auf die erfasste Unterrichtsqualität bestimmter Situationen auswirkt (z. B. Brunner 2018), um dadurch der Rolle von Fachspezifität bei der Erfassung von Unterrichtsqualität durch Ratings auf die Spur zu kommen.

Die graduierte Spezifität von Kriterien lässt sich beispielsweise wie folgt beschreiben:

- *Generisches Kriterium*, z. B. „Die Lehrkraft verwendet anschauliche Visualisierungen, um Vorstellungen zu relevanten Konzepten aufzubauen“.
- *Fachspezifisches Kriterium*, z. B. „Die Lehrkraft verwendet anschauliche Visualisierungen, um tragfähige Vorstellungen zu mathematischen Konzepten aufzubauen“.
- *Inhaltspezifisches Kriterium*, z. B. „Die Lehrkraft verwendet anschauliche Visualisierungen, um tragfähige Vorstellungen zum Variablenbegriff (Variable als Unbestimmte, Unbekannte und Veränderliche) aufzubauen.“

Diese unterschiedlichen Kriterien sorgen insbesondere dann für Varianz in der erfassten Unterrichtsqualität, wenn eine Unterrichtssituation vorliegt, die zwar das generische Kriterium erfüllt, das inhaltspezifische jedoch nicht. Ein illustrierendes Beispiel bezieht sich auf das Phänomen der sogenannten „Fruchtsalat-Algebra“:

Dabei werden beispielsweise erlaubte Termumformungsregeln, wie das Distributivgesetz, anschaulich mit Hilfe von Äpfeln und Birnen begründet. Mit Hilfe einer Visualisierung, welche die gleiche Menge aus Äpfeln und Birnen in zwei Sortierungen zeigt, wird begründet, dass die Terme, die diese Menge in der jeweiligen Sortierung beschreiben, gleichwertig sind. (Konkret:  $6a + 4b = 2 \cdot (3a + 2b)$ , denn 6 Äpfel und 4 Birnen kann man auch darstellen als 2 Gruppen aus jeweils 3 Äpfeln und 2 Birnen). Eine solche Erklärung nutzt eine anschauliche Visualisierung, um inhaltliche Vorstellungen aufzubauen, die diese Termumformung für Lernende als Regel durchaus plausibel macht. Das oben genannte generische Kriterium ist damit erfüllt. Aus der fachdidaktischen Forschung ist jedoch bekannt, dass diese Fruchtsalat-Algebra die weit verbreitete Fehlvorstellung von Variablen als Bezeichner realer Objekte fördert, statt eine tragfähige Grundvorstellung zu Variablen zu unterstützen (z. B. MacGregor und Stacey 1997). Die genutzte Visualisierung ist bei näherer Betrachtung also eher eine oberflächlich analoge „Strukturmetapher“, die den inhaltlichen Kern von allgemeiner Termgleichheit bei unbestimmten Anzahlen verdeckt und die Verstehensgrundlagen für das Weiterlernen gefährdet. Insofern sind hier die entsprechenden fach- und inhaltspezifischen Kriterien nicht erfüllt.

Dieses Beispiel macht deutlich, dass generische Kriterien fachspezifisch (insbesondere fachdidaktisch) präzisiert werden müssen, um die Qualität von Mathematikunterricht in solchen Situationen erfassen zu können. Es zeigt aber auch, dass für ein entsprechendes Urteil anhand eines fachspezifischen Kriteriums (zumindest implizit) eine inhaltspezifische Ausdifferenzierung nötig ist (z. B. Was sind tragfähige Vorstellungen zum Variablenbegriff? Welche Visualisierungen können diese unterstützen?). Eine vergleichbare inhaltspezifische Ausdifferenzierung wurde beispielsweise in dem oben genannten Pythagoras-Projekt bezüglich der fachdidaktischen Subdimensionen (z. B. Qualität der Repräsentationsformen) im Zusammenhang mit dem Satz des Pythagoras vorgenommen (Lipowsky et al. 2018).

Dies bedeutet jedoch, dass es eigentlich geboten ist, Kriterien für Unterrichtsqualität für jeden Unterrichtsgegenstand inhaltspezifisch ausdifferenzieren. Dies ist in inhaltlich eng fokussierten Forschungsprojekten möglich, würde allerdings für breit anzuwendende Instrumente zur Erfassung von Unterrichtsqualität über viele Unterrichtsstunden hinweg den Entwicklungsaufwand stark aufblähen. Es stellt sich daher die Frage, ob man der Fachspezifität von Unterrichtsqualität überhaupt durch Ausdifferenzierung von Kriterien im Rahmen von Instrumenten gerecht werden muss, oder ob die fach- und inhaltspezifischen Ebenen durch entsprechende Kompetenzen von Ratern berücksichtigt werden können.

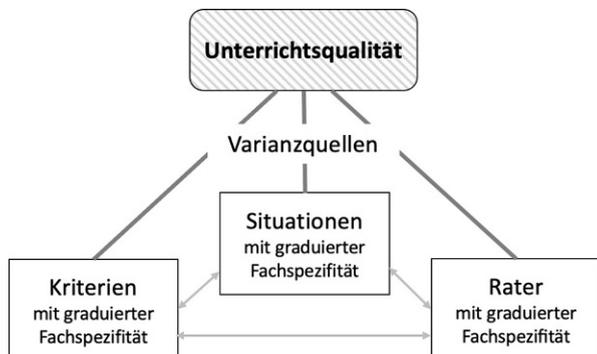
Bislang wird Varianz in der erfassten Unterrichtsqualität, die durch unterschiedliche geschulte Rater entsteht, meist als unerwünschter Messfehler im Sinne von Verzerrungen (Rater-Bias) und unzureichender Übereinstimmung (Interrater-Reliabilität) betrachtet. Doch Rater müssen bei der Anwendung von (fachspezifischen) Kriterien auf konkrete Unterrichtssituationen das komplexe Unterrichtsgeschehen immer mit Blick auf inhaltsgebundene Zielsetzungen und Interaktionen erfassen und dabei auch spontan inhaltliche Ausdifferenzierungen der Kriterien vornehmen. Angesichts der Erkenntnisse der Noticing-Forschung bezüglich der Bedeutung von fach- und inhaltspezifischem Wissen für die Wahrnehmung und Interpretation von Unterrichtssituationen (z. B. Dreher und Kuntze 2015; Stahnke et al. 2016), lässt

sich also Folgendes vermuten: Bei Ratings von Unterrichtsqualität geht häufig fach- und inhaltsbezogene Expertise ein, die über das hinausgeht, was Raterschulungen leisten können. Es liegt folglich nahe, dass auch hier eine systematische Varianz in der erfassten Unterrichtsqualität vorliegt, die sich auf Fachspezifität zurückführen lässt: Ähnlich wie man die Relevanz fachspezifischer Kriterien dadurch zeigt, dass sich die Variation von Kriterien in ihrer Fachspezifität systematisch auf die erfasste Unterrichtsqualität bestimmter Situationen auswirkt, kann man also in Bezug auf Rater untersuchen, inwiefern sich eine Variation in der „Fachspezifität von Ratern“ systematisch auf die erfasste Unterrichtsqualität dieser Situationen auswirkt, um die Bedeutung von Ratern mit fach- und inhaltsbezogener Expertise zu untermauern. So würde man beispielsweise für eine Unterrichtssituation nach dem oben skizzierten Fruchtsalat-Algebra Zugang erwarten, dass generische Expertinnen und Experten für Unterrichtsqualität das gleiche, oben genannte fachspezifische Kriterium systematisch anders bewerten würden als Personen mit mathematikdidaktischer Expertise. Generische Rater würden in diesem Fall vermutlich einen entscheidenden Aspekt für die Qualität der Unterrichtssituation trotz Anwendung eines fachspezifischen Kriteriums nicht erkennen.

Vor dem Hintergrund dieser Überlegungen schlagen wir vor, die Untersuchung der Fachspezifität von Unterrichtsqualität und Berücksichtigung dieser in der Messung von Unterrichtsqualität nicht auf die Betrachtung von mehr oder weniger fachspezifischen Kriterien und Situationen zu beschränken, sondern wie in Abb. 1 dargestellt, auch die Fachspezifität von Ratern einzubeziehen.

Demnach ist zu erwarten, dass sich die Fachspezifität von Unterrichtsqualität bei ihrer Erfassung durch Ratings in den drei Varianzquellen „Unterrichtssituationen“, „Kriterien“ sowie „Rater“ und ihren Wechselwirkungen zeigt (ganz im Sinne der Generalisierbarkeitstheorie, Cronbach et al. 1972). Diese Varianzquellen weisen in Bezug auf Fachspezifität jeweils unterschiedliche Abstufungen auf, die systematisch berücksichtigt werden sollten, um der Fachspezifität von Unterrichtsqualität auf die Spur zu kommen. Insbesondere nehmen wir an, dass das fach- und inhaltsbezogene Wissen von Ratern ihre Interpretationen von Kriterien sowie Situationen beeinflussen, was sich dann systematisch in ihren Einschätzungen der Unterrichtsqualität niederschlägt.

**Abb. 1** Die Fachspezifität von Unterrichtsqualität zeigt sich in unterschiedlichen Varianzquellen: In allen drei (miteinander zusammenhängenden) Dimensionen können sich verschiedene Grade an Fachspezifität systematisch auf die eingeschätzte Unterrichtsqualität auswirken



Es gilt folglich zu untersuchen, wie sich die systematische Variation der einzelnen Elemente in ihrer Fachspezifität auf die erfasste Unterrichtsqualität auswirkt. Methodisch kann man hier beispielsweise mit qualitativen Analysen verbaler Raterurteile den Auswirkungen verschiedener Grade an Fachspezifität auf die Spur kommen und vermutete systematische Zusammenhänge quantitativ, zum Beispiel mit Hilfe der Generalisierbarkeitstheorie untersuchen (Praetorius et al. 2014; Wiedmann et al. 2019; Jentsch et al. 2020). Ein solcher Ansatz liefert aus unserer Sicht vielversprechende Fragestellungen für Forschung, in denen generische und fachspezifische Aspekte von Unterrichts- und Noticing-Forschung zusammenkommen.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik*, 39(2), 257–284.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: Wiley.
- Dreher, A., & Kuntze, S. (2015). Teachers' professional knowledge and noticing: the case of multiple representations in the mathematics classroom. *Educational Studies in Mathematics*, 88(1), 89–114.
- Dreher, A., Lindmeier, A., Heinze, A., & Niemand, C. (2018). What kind of content knowledge do secondary mathematics teachers need? A conceptualization taking into account academic and school mathematics. *Journal für Mathematik-Didaktik*, 39(2), 319–341.
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit: Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Münster: Waxmann.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study. *Cognition and Instruction*, 26(4), 430–511.
- Jentsch, A., Casale, G., Schlesinger, L., Kaiser, G., König, J., & Blömeke, S. (2020). Variabilität und Generalisierbarkeit von Ratings zur Qualität von Mathematikunterricht zwischen und innerhalb von Unterrichtsstunden. *Unterrichtswissenschaft*, 48(2), 179–197.
- Learning Mathematics for Teaching Project (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education*, 14, 25–47.
- Leuders, T., & Holzäpfel, L. (2011). Kognitive Aktivierung im Mathematikunterricht. *Unterrichtswissenschaft*, 39(3), 213–230.
- Lindmeier, A., & Heinze, A. (2020). Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant? *Zeitschrift für Pädagogik*, 66, 255–267.

- 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.
- MacGregor, M., & Stacey, K. (1997). Students' understanding of algebraic notation: 11–15. *Educational Studies in Mathematics*, 33, 1–19.
- Nowinska, E. (2016). *Leitfragen zur Analyse und Beurteilung metakognitiv-diskursiver Unterrichtsqualität*. Osnabrück: Forschungsinstitut für Mathematikdidaktik.
- Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: looking back and looking forward. *ZDM*, 50(3), 535–553.
- Praetorius, A. K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B., & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, 48(3), 409–446.
- Praetorius, A. K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of three basic dimensions. *ZDM*, 50(3), 407–426.
- Praetorius, A. K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Schlesinger, L., Jentsch, A., Kaiser, G., König, J., & Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM*, 50(3), 475–490.
- Schoenfeld, A. H. (2018). Video analyses for research and professional development: the teaching for robust understanding (TRU) framework. *ZDM Mathematics Education*, 50(3), 491–506.
- Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: a systematic review of empirical mathematics education research. *ZDM*, 48(1–2), 1–27.
- Wiedmann, M., Kaendler, C., Leuders, T., Spada, H., & Rummel, N. (2019). Measuring teachers' competence to monitor student interaction in collaborative learning settings. *Unterrichtswissenschaft*, 47(2), 177–199.