

HANSEN, Linn; FRIESEN, Marita & DREHER, Anika
Freiburg, Heidelberg, Freiburg

Unterrichtsqualitätsratings im Fach Mathematik: Inwiefern sind fachspezifische Beobachtungskriterien nötig?

Theoretischer Hintergrund und Fragestellung

Unterrichtsqualität ist schon seit Jahren im Fokus von bildungswissenschaftlichen und fachdidaktischen Forschungsvorhaben (Praetorius et al., 2018). Dabei gelten Unterrichtsqualitätsratings durch externe Beobachtende als „Königsweg“ (Praetorius, 2013). Es zeigt sich aber, dass Beobachterratings durch verschiedene Faktoren beeinflusst werden. Zum Beispiel können variierende Formulierungen der Beobachtungskriterien zu unterschiedlichen Qualitätsratings führen: Je nachdem, wie fachspezifisch die Kriterien formuliert sind, variiert die Berücksichtigung fachlicher Unterrichtsaspekte in der Unterrichtseinschätzung (Brunner, 2018). Außerdem interpretieren Rater:innen Beobachtungskriterien unterschiedlich (Praetorius, 2013). So ist beispielsweise zu erwarten, dass für die Beurteilung von Mathematikunterricht mit fachspezifischen Lernzielen und -prozessen fachspezifisches Wissen notwendig ist (z.B. Lindmeier & Heinze, 2020).

Um solche Zusammenhänge zu berücksichtigen, identifizierten Dreher und Leuders (2021) drei Varianzquellen für die Untersuchung der Fachspezifität von Unterrichtsqualitätsratings: Unterrichtssituationen, Rater:innen und Beobachtungskriterien. Es wird erwartet, dass sich diese Varianzquellen durch verschiedene Grade an Fachspezifität systematisch auf die Güte von Unterrichtsqualitätsratings auswirken können und auch in Wechselwirkung zueinanderstehen. In unserem Projekt wird die Fachspezifität der Unterrichtssituation hochgehalten, da ausschließlich mathespezifische Unterrichtssituationen fokussiert werden. Die Varianzquelle der Rater:innen zeigt sich durch unterschiedliches Professionswissen sowie Kompetenzen der Rater:innen. Mathematiklehrkräfte weisen beispielsweise eine hohe Fachspezifität auf, wenn es um die Beobachtung von Mathematikunterricht geht. Als dritte Varianzquelle wird die Fachspezifität der Kriterien angenommen. Einerseits sind Kriterien bezüglich der kognitiven Aktivierung beispielsweise fachspezifischer als diejenigen bezüglich Klassenmanagement (Brunner, 2018). Andererseits können auch Kriterien bezüglich kognitiver Aktivierung unterschiedlich stark fachspezifisch formuliert sein (Dreher & Leuders, 2021).

Um zu verstehen welche Rollen die Fachspezifität von Rater:innen und Kriterien für die Einschätzung fachspezifischer Unterrichtssituationen spielen, ist es hilfreich zu berücksichtigen, dass Unterrichtsbeobachtungen auf Noticingprozessen basieren: Um Unterrichtsqualität erfassen zu können, müssen

relevante Ereignisse im Unterricht wahrgenommen, interpretiert und eingeschätzt werden (Dreher & Leuders, 2021). Aus der Noticingforschung ist bekannt, dass diese Noticingprozesse professionelles Wissen, insbesondere fachliches und fachdidaktisches Wissen, erfordern (Dreher & Kuntze, 2015). Außerdem werden in der Noticingforschung unterschiedliche Beobachungskriterien in Form von Prompts eingesetzt. Dabei zeigt sich, dass fachspezifische Prompts das Noticing von zentralen fachdidaktischen Aspekten des Unterrichts begünstigen. Offen bleibt jedoch, wann generische Prompts für ein erfolgreiches Noticing ausreichen und wann fachspezifische Prompts notwendig sind (Estapa & Amador, 2023).

Daher soll in diesem Beitrag der Frage nachgegangen werden, in welchen Fällen fachspezifische Rater:innen anhand generischer Items die Qualität fachspezifischer Unterrichtssituationen adäquat erfassen und in welchen Fällen fachspezifische Kriterien die Güte der Unterrichtsqualitätsratings erhöhen. Zu einem späteren Zeitpunkt des Projekts wird dann die Fachspezifität der Rater:innen variiert.

Methodik

Zur Untersuchung der Forschungsfrage wurden von uns auf Basis des Unterrichtsfeedbackbogens (UFB) (Fauth et al., 2021) sechs geskriptete Unterrichtsvideos entwickelt. Die Videos wurden zu denjenigen Qualitätsaspekten des UFBs entwickelt, die als fachspezifisch angenommen werden. Dabei wurden die Unterrichtssituationen so konzipiert, dass sie Standardsituationen des Mathematikunterrichts widerspiegeln (Seidel et al., 2021) und jeweils einen fachdidaktischen Normbruch bezüglich des fokussierten UFB-Aspekts enthalten. Zum Beispiel besteht der Normbruch für das Item "Engagement der Schülerinnen und Schüler" darin, dass die Lernenden zwar engagiert am Unterrichtsgeschehen beteiligt sind, sich dieses Engagement aber nicht auf die fachlichen Inhalte der Stunde bezieht.

Für jedes Video wurde neben dem generischen UFB-Item auch ein fachspezifisches Item bezüglich des fokussierten Unterrichtsqualitätsaspekts formuliert (vgl. Dreher & Leuders, 2021). Die Videos wurden so konzipiert, dass wir aus theoretischer Sicht erwarten, dass sich die Normverstöße in drei der Videos nur auf das fachspezifische Item auswirken. In den anderen drei Videos sollten beide Items den Normverstoß erfassen. Das Video zum Aspekt „Engagement der Schülerinnen und Schüler“ ist ein Beispiel für den ersten, inkongruenten Fall. Das UFB-Item "Die Schülerinnen und Schüler sind engagiert am Unterrichtsgeschehen beteiligt" erfasst den beschriebenen Normverstoß nicht, das fachspezifische Item "Die Schülerinnen und Schüler setzen sich engagiert mit den mathematischen Aspekten der Lerngelegenheit

auseinander" jedoch schon. Wir erwarten daher Varianz durch die Fachspezifität der Beobachtungskriterien: Fachspezifische Rater:innen sollten das generische Item, das sich auf das allgemeine Engagement der Lernenden bezieht, positiv, das fachspezifische Item jedoch negativ raten. Ein Beispiel für den zweiten, kongruenten Fall ist das Video zum Unterrichtsqualitätsaspekt „Individuelle Unterstützung“. Der Normverstoß liegt hier darin, dass die Lehrkraft nicht auf die Lösungsansätze der Schüler:innen eingeht, sondern ihnen einen Lösungsweg "überstülpt". Hier erfasst das UFB-Item: "Die Lehrkraft unterstützt die Schülerinnen und Schüler individuell in ihrem Lernprozess" bereits den Normverstoß, sodass für das fachspezifische Item keine grundsätzlich andere Einschätzung erwartet wird als für das generische Item.

Um die Fachspezifität der Beobachtungskriterien zu untersuchen, wird fachspezifischen Rater:innen in einer Online-Erhebung randomisiert eine Testversion zugeteilt (rein generische bzw. fachspezifische Items). Bevor sie das jeweilige Unterrichtsvideo sehen, sehen sie das entsprechende Item. Das Item wird auf einer 4-Punkt-Likertskala mit den Endpunkten "trifft nicht zu" und "trifft zu" geratet und die Einschätzung in einer offenen Antwort begründet. Als angemessene Einschätzung und somit hohe Güte des Unterrichtsqualitätsratings werden die Einschätzungen "trifft nicht zu" und "trifft eher nicht zu" gewertet. Die Begründungen werden danach kodiert, ob der Normverstoß erkannt wurde oder nicht (Indikator für Noticing).

Erste Ergebnisse und Ausblick

Pilotierungsdaten deuten darauf hin, dass sich die Ratings in den kongruenten Fällen tatsächlich nicht signifikant voneinander unterscheiden. In den inkongruenten Fällen wurden die fachspezifischen Items jedoch signifikant negativer geratet. Die Kodierung der offenen Antworten bestätigt zusätzlich den Zusammenhang der negativen Ratings der fachspezifischen Items mit dem Erkennen der Normverstöße. Zu beachten ist allerdings, dass die Daten für die Pilotierung in einem Within-Subjects-Design erhoben und die Formulierungen der fachspezifischen Items auf Grundlage dieser Daten zwischenzeitlich weiterentwickelt wurden. Insofern bleiben die Ergebnisse der Haupterhebung abzuwarten, die bis zum Vortrag zumindest teilweise vorhanden sein sollten.

Es wird für alle sechs Videos erwartet, dass die fachdidaktischen Normverstöße für generische Rater:innen kaum sichtbar sind, so dass auch die Varianz durch die Fachspezifität der Rater:innen untersucht werden kann. Daher sollen zusätzlich Daten von fachfremden Rater:innen erhoben werden. So können die Ratings und Noticingprozesse von fachfremden und fachspezifischen Rater:innen verglichen werden. Im Verlauf des Projekts sollen weitere

Aspekte der Unterrichtsqualitätsratings in den Blick genommen werden: Es stellt sich beispielsweise die Frage, ob sich die Güte der Unterrichtsqualitätsratings fachfremder Rater:innen durch eine Intervention erhöhen lässt.

Insgesamt soll durch die Untersuchung der im Forschungsprojekt aufgeworfenen Fragestellungen ein Beitrag zur Diskussion über die Fachspezifität von Unterrichtsqualitätsratings geleistet werden. Dabei ist zu bedenken, dass Unterrichtsqualitätsratings nicht nur in Forschungskontexten stattfinden, sondern auch in der Praxis. Tiefere Erkenntnisse über die Fachspezifität von Unterrichtsqualitätsratings sind somit auch für die professionellen Entwicklung in der Aus- und Weiterbildung von Lehrkräften relevant.

Literatur

- Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik*, 39(2), 257–284. <https://doi.org/10.1007/s13138-017-0122-z>
- Dreher, A. & Kuntze, S. (2015). Teachers' professional knowledge and noticing: The case of multiple representations in the mathematics classroom. *Educational Evaluation and Policy Analysis*, 88(1), 89–114. <https://doi.org/10.1007/s10649-014-9577-8>
- Dreher, A., & Leuders, T. (2021). Fachspezifität von Unterrichtsqualität – aus der Perspektive der Mathematikdidaktik. *Unterrichtswissenschaft*, 49(2), 285–292. <https://doi.org/10.1007/s42010-021-00116-9>
- Estapa, A., & Amador, J.-M. (2023). A qualitative metasynthesis of video-based prompts and noticing in mathematics education. *Mathematics Education Research Journal* 35, 105–131. <https://doi.org/10.1007/s13394-021-00378-7>
- Fauth, B., Ruth-Herbein, E., & Maier, J. L. (2021). Beobachtungsmanual zum Unterrichtsfeedbackbogen Tiefenstrukturen - Überarbeiteter Entwurf. Institut für Bildungsanalysen Baden-Württemberg. <https://doi.org/10.1007/s13394-021-00378-7>
- Lindmeier, A., & Heinze, A. (2020). Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant? *Zeitschrift für Pädagogik*, 66, 255–267. <https://doi.org/10.25656/01:25878>
- Praetorius, A.-K. (2013). *Messung von Unterrichtsqualität durch Ratings*. Waxmann Verlag.
- 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. <https://doi.org/10.1007/s11858-018-0918-4>
- Seidel, T., Renkl, A., & Werner, R. (2021). Basisdimensionen für Unterrichtsqualität im Fachkontext konkretisieren: Die Rolle von Unterrichtsartefakten und Bestimmung von Standardsituationen. *Unterrichtswissenschaft*, 49(2), 293–301. <https://doi.org/10.1007/s42010-021-00108-9>