

SIMBÖCK, Laura; LINDL, Alfred & KRAUSS, Stefan
Regensburg

Kriterienbasierte Einschätzung von (fachspezifischer) Unterrichtsqualität mittels AMADEUS

1 Forschungshintergrund und Fragestellung

Zur Untersuchung der Merkmale von Unterrichtsqualität, die als wesentliche Einflussgrößen für Lernerfolge von Schüler*innen gelten, liegt mittlerweile eine Vielfalt an theoretischen und methodischen Zugängen vor (Helmke, 2022). Um diese Ansätze miteinander in Verbindung zu bringen und für weitere Forschung fruchtbar zu machen, erarbeiteten Praetorius et al. (2020) aus der vergleichenden Zusammenschau von zwölf für die Unterrichtsbeobachtung konzipierten Schemata das sogenannte (erweiterte) Syntheseframework.

Dieses Framework wurde in der interdisziplinären Forschungsgruppe FALKO-PV (**F**achspezifische **L**ehrkräfte**k**ompetenzen – **P**rädiktive **V**alidierung), welche primär die Zusammenhänge zwischen professioneller Kompetenz von Lehrkräften, der Qualität und Zielkriterien von Unterricht untersucht, als konzeptueller Rahmen zur Beschreibung von Unterrichtsqualität herangezogen sowie erstmals ein darauf fußendes Beurteilungsinstrument aus der Perspektive von Schüler*innen entwickelt und eingesetzt. In diesem Beitrag werden erste Befunde dazu berichtet, beispielsweise inwiefern damit Merkmale von Mathematikunterricht reliabel, differenzsensibel und valide erfasst werden.

2 Methode

Stichprobe. Die erste Erhebungsrunde in FALKO-PV fand im Schuljahr 2023/24 an bayerischen Gymnasien statt. Daran beteiligten sich in Mathematik sechs Lehrkräfte (eine weiblich, fünf männlich) mit acht Klassen (ein Lehrer unterrichtete drei Klassen) jeweils der achten Jahrgangsstufe. Diese umfassten insgesamt 170 Schüler*innen (58,4 % weiblich). Die Teilnahme erfolgte freiwillig, unentgeltlich und unter Berücksichtigung forschungsethischer und datenschutzrechtlicher Bestimmungen.

Instrumente. Zur kriterienbasierten Einschätzung von Unterrichtsmerkmalen aus Schüler*innensicht wurde ein neu konzipierter Fragebogen mit sieben generischen Dimensionen und einigen fachspezifischen Aspekten genutzt (vgl. Tabelle 1). Die Konzeptualisierung der generischen Dimensionen orientierte sich an dem erweiterten Syntheseframework von Praetorius et al. (2020), fachspezifische Ergänzungen wurden auf Basis eines interdisziplinären Austauschs der beteiligten sechs Fachdidaktiken (z. B. interkulturelle

In: L. Schick, M. Platz & A. Lambert (Hrsg.),
Beiträge zum Mathematikunterricht 2025.

Bezüge / Fremdheitserfahrung) und einer systematischen Literaturrecherche (z. B. gendersensibler Unterricht) vorgenommen. Die zu allen Aspekten formulierten Items wurden einem verbal probing (Frei, 2019) unterzogen, mehrfach pilotiert und gegebenenfalls adjustiert (Rader et al., 2022). Itembeispiele für ausgewählte Merkmale sind „*Unsere Lehrkraft hat die Klasse im Griff*“ (Klassenführung) oder „*Im Unterricht üben wir immer wieder, wie man Probleme des Alltags in die Welt der Mathematik überführen kann*“ (mathematisches Modellieren). Alle 154 Items, die im finalen Fragebogen enthalten sind, werden auf einer fünfstufigen Likertskala beantwortet (0: trifft nicht zu, ..., 4: trifft zu).

Weiterhin wurden bei Schüler*innen mittels eines Fragebogens demographische Angaben, aber auch wichtige Kontextvariablen wie Lernfreude, fachliches Selbstkonzept oder Ängstlichkeit erfasst.

Merkmaldimension	Items	α	$M (SD)$	$[M_{\min}; M_{\max}]$	ICC
Auswahl und Thematisierung von Inhalten und Fachmethoden	29	,90	2,74 (0,55)	[2,48; 3,03]	,08
Kognitive Aktivierung	21	,88	2,66 (0,59)	[2,32; 2,94]	,12
Unterstützung des Übens	10	,87	3,15 (0,66)	[2,72; 3,46]	,09
Formatives Assessment	16	,92	2,61 (0,81)	[2,10; 2,90]	,09
Unterstützung des Lernens aller Schüler*innen	15	,87	2,51(0,65)	[2,28; 2,76]	,05
Sozio-emotionale Unterstützung	14	,82	2,96 (0,59)	[2,61; 3,28]	,06
Klassenführung	18	,84	2,63 (0,57)	[2,31; 2,94]	,09
Interkulturelle Bezüge / Fremdheitserfahrung	4	,75	1,74 (0,89)	[1,49; 2,08]	,00
Gendersensibler Unterricht	6	,72	3,24 (0,76)	[2,86; 3,52]	,04
Modellieren	9	,88	2,38 (0,82)	[1,80; 2,72]	,10
Historische Kontextualisierung	3	,62	1,80 (0,93)	[1,21; 2,26]	,07
Verständnisfördernde Technologienutzung	7	,91	2,24 (1,05)	[1,76; 3,28]	,26

Tabelle 1: Interne Konsistenz (Cronbachs α), arithmetische Mittelwerte (M), Standardabweichungen (SD), minimale bzw. maximale Klassenmittelwerte, Intraklassenkorrelation (ICC) pro Merkmaldimension ($N = 170$ Schüler*innen)

Implementation. Die Items zu Merkmalen von Unterrichtsqualität wurden in die projektseitig neu entwickelte Webapp AMADEUS (anonym nutzbare

Mobile App zur digitalen Evaluation des Unterrichts durch Schüler*innen) implementiert, was eine technisch einfache, zeitökonomische Durchführung ermöglicht. Der Einsatz der Items folgt dabei in einem Rotationsdesign mit drei Erhebungszeitpunkten zu je zwei Durchgängen. Dies bedeutet, jede*r Schüler*in einer bestimmten Klasse beantwortet pro Durchgang ca. 1/3 der Items, was ca. zehn Minuten erfordert, sodass je Durchgang und Klasse alle Items von ca. 1/3 der Schüler:innen bearbeitet werden und nach drei Durchgängen jede*r Schüler*in alle Items einmal beurteilt hat.

Grundlage für die nachstehenden Analysen ist das vollständige Datenset nach den drei ersten Durchgängen aus dem Schuljahr 2023/24. Dieses wird mit Angaben aus dem Hintergrundfragebogen, der zu Beginn des Schuljahres ebenfalls (mittels LimeSurvey) digital administriert wurde, in Verbindung gebracht.

3 Ausgewählte Ergebnisse

Nach Ausschluss zweier Items aus psychometrischen Gründen (negative Trennschärfen) aus den Skalen *Auswahl und Thematisierung von Inhalten und Fachmethoden* sowie *Unterstützung des Übens* ist, wie aus Tabelle 1 hervorgeht, die interne Konsistenz aller Merkmaldimensionen zufriedenstellend bis gut ($,62 < \alpha < ,92$). *Interkulturelle Bezüge / Fremdheitserfahrung* und *historische Kontextualisierungen* nehmen die Schüler*innen im Mittel tendenziell nicht als Merkmal ihres Mathematikunterrichts wahr, *sozio-emotionale Unterstützung* und *Unterstützung des Übens* kommen ihrer Einschätzung nach im Unterricht vor. Dabei liegen mit Blick auf die Variabilität innerhalb und zwischen den Klassen auch Bewertungsunterschiede vor, welche eine gewisse Sensitivität der Skalen indizieren. Bei *kognitiver Aktivierung* und *verständnisfördernder Technologienutzung* erweisen sich die Unterschiede zwischen den Klassenmittelwerten (vgl. die ICC in Tabelle 1) in einfaktoriellen Varianzanalysen als signifikant ($F(7, 54,33) = 3,79, p < ,05$, bzw. $F(7, 52,48) = 15,24, p < ,001$ adjustiert nach Holm).

Zwischen allen betrachteten Merkmaldimensionen liegen mittlere bis hohe signifikante bivariate Korrelationen vor ($,31 \leq \text{Pearsons } r \leq ,81$). Diese scheinen also differente, wenn auch erwartungsgemäß nicht disjunkte Aspekte von Unterricht zu erfassen. Diskussionswürdig ist insbesondere die hohe Korrelation zwischen *kognitiver Aktivierung* und *formativem Assessment*.

In weiterführenden Korrelationsanalysen zu allen erfassten Unterrichtsaspekten und verschiedenen Hintergrundmerkmalen von Schüler*innen werden drei Tendenzen deutlich: Erstens finden sich keine signifikanten Zusammenhänge mit sozio-demographischen Merkmalen wie Geschlecht, Familiensprache, familialem Bildungs- oder ökonomischem Hintergrund. Zweitens

treten plausible Korrelationen mit affektiven Bildungszielen wie Lernfreude, Ängstlichkeit und fachlichem Selbstkonzept auf. Drittens ist die Vorjahresnote in Mathematik, aber nicht diejenige in Deutsch signifikant mit den Einschätzungen der Schüler*innen zu Unterrichtsmerkmalen assoziiert.

4 Fazit und Ausblick

Die hier vorgestellten ersten Befunde deuten darauf hin, dass eine reliable und differenzsensible Erfassung verschiedener generischer und fachspezifischer Merkmale von Mathematikunterricht aus der Sicht von Schüler*innen mittels der neu konstruierten Webapp AMADEUS möglich ist. Korrelationsanalysen indizieren einerseits faktorielle, andererseits auch inhaltliche und kriteriale Validität, indem keine signifikanten Zusammenhänge mit Hintergrundvariablen, wohl aber mit affektiven und kognitiven Kriterien des Unterrichtserfolgs im Fach Mathematik bestehen.

Diese Ergebnisse sind auf Basis der umfassenderen Stichprobe der zweiten Runde der Haupterhebung im Schuljahr 2024/25 zu evaluieren und durch zusätzliche Analysen abzusichern (z. B. konfirmatorische Faktorenanalysen). Zudem soll die zeitliche Stabilität der Merkmalseinschätzungen über alle drei über das Schuljahr verteilten Erhebungszeitpunkte hinweg überprüft werden. Schließlich wird auch die Lehrkraftebene, insbesondere das domänenspezifische Professionswissen, sowie die Kompetenzentwicklung der Schüler*innen in die Betrachtung einbezogen, um die unterrichtliche Wirkungskette von der Lehrkraftkompetenz zur Performanz von Schüler*innen zu untersuchen.

Literatur

- Frei, M. (2019). Probing als unterstützende Methode bei der Entwicklung eines Fragebogens. Ein Anwendungsbericht aus dem Forschungsprojekt FALKE-Musik. In G. Puffer, A. Becker, F. Körndle & K. Sprau (Hrsg.), *Musik—Pädagogik—Professionalität. Festschrift für Bernhard Hofmann zum 60. Geburtstag* (S. 121–136). Helbling.
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung. Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*. Klett.
- Praetorius, A.-K., Rogh, W. & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, 48(3), 303–318. <https://doi.org/10.1007/s42010-020-00072-w>
- Rader, M., Simböck, L., Krämer, A., Gutsmedl, M., Gürtner, M., Ehrich, P., Böhringer, S. & Lindl, A. (2022). 3. Statusseminar der BMBF-Nachwuchsforschungsgruppen im Rahmenprogramm Empirische Bildungsforschung, Institut zur Qualitätsentwicklung im Bildungswesen (IQB) Berlin: Posterpräsentation zu „Was macht guten Unterricht im Fach aus? Messung unterrichtlicher Qualitätsmerkmale in sechs Fächern“. <https://doi.org/10.13140/RG.2.2.10780.81281>