

## Reflexionsfähigkeit: Validierungsstudien zu einem Test mit Feedback

Die Reflexion von Unterricht gilt als probates Mittel zur Förderung der Professionalisierung angehender Lehrkräfte und demgemäß als Kernaufgabe von Lehrkräften (z. B. Carlson et al., 2019; von Aufschnaiter et al., 2019). Da Reflexionen (angehender) Lehrkräfte allerdings häufig auf niedrigen bzw. deskriptiven Niveaus verbleiben (z. B. Wyss, 2013; Hatton & Smith, 1995), wird ihr Potential für die Entwicklung von Unterricht und der eigenen Fähigkeiten nicht ausgeschöpft. Um Physik-Lehramtsstudierende bei der Entwicklung ihrer Reflexionsfähigkeit zu unterstützen, wurde im Rahmen des Projekts *ProfiLe-P-Transfer* ein Testinstrument mit Feedback entwickelt, das die fundierte und valide Einschätzung dieser Fähigkeit ermöglichen soll. Das Instrument steht unter [www.unterrichtsreflexion.de](http://www.unterrichtsreflexion.de) frei zur Verfügung.

### Die Materialien: Testinstrument und Feedback

Die Erfassung der Reflexionsfähigkeit angehender Physiklehrkräfte erfolgt mit einem online-basierten Testinstrument. Darin werden sieben (geskriptete) inhaltlich aufeinander folgende Videovignetten einer Physik-Doppelstunde reflektiert. Der Einsatz von Videovignetten ermöglicht dabei die systematische Betrachtung authentischer Unterrichtssituationen in komplexitätsreduzierter Form (Lohse-Bossenz et al., 2023; Oser et al., 2010). Die Reflexion wird in eine fiktive kollegiale Beratungssituation eingebettet: Robert, ein fiktiver Mitpraktikant, bittet um Feedback, das die angehenden Lehrkräfte zu 16 Aspekten jeweils im Rahmen von zwei Multiple-Choice (MC)-Aufgaben geben (je eine Aufgabe zur Bewertung und eine zum Vorschlagen von Handlungsalternativen). Betrachtet wird dabei u. a. der stereotypische Umgang mit Schüler\*innen. Das anschließende Feedback soll den Studierenden eine fundierte Einschätzung der eigenen Fähigkeit ermöglichen. Um das Feedback möglichst effektiv zu gestalten (Hattie & Timperley, 2007), enthält es Informationen zur Konzeptualisierung der Unterrichtsreflexion, dessen Umsetzung im Testinstrument, die individuellen Ergebnisse mit einer Einordnung in eine Vergleichsgruppe sowie den Hinweis auf ein Selbstlernmaterial zur unterstützten Reflexion dreier weiterer Unterrichtsausschnitte.

### Validierungsstudien zu Testinstrument und Feedback

Testinstrument und Feedback werden evaluiert, um die Interpretation der Testwerte als Maß für die Reflexionsfähigkeit der Studierenden und damit als Ausgangspunkt für weitere Professionalisierung zu validieren. Im Sinne des Argument-based-Approach (Kane, 2013) werden Argumente für die Validität gesammelt. Dazu werden die Übersetzungsschritte evaluiert, die notwendig sind, um den Studierenden ausgehend vom Konstrukt der Unterrichtsreflexion die Interpretation ihres Feedbacks zu ermöglichen (s. Abb. 1; vgl. Dickmann, 2016). Dabei werden Aspekte der inhaltlichen, kognitiven, externen und konsequentiellen Validität berücksichtigt (Messick, 1995).

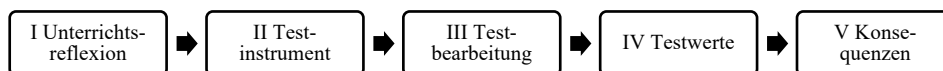


Abb. 1: Übersetzungsschritte vom betrachteten Konstrukt Unterrichtsreflexion bis hin zu Konsequenzen aus dem Feedback

### **Zusammenfassung der Ergebnisse der Validierungsstudien**

Das entwickelte Testinstrument basiert auf dem von Kempin et al. (2018) entwickelten Reflexionsperformanztest, der Reflexionen zu Videovignetten verbal und in einem offenen Antwortformat einfordert. Für die *inhaltliche Validität* des Testinstruments (Übersetzungsschritt I zu II) ist sicherzustellen, dass die Auswahl der Szenen und Aufgaben zur Abbildung des Konstrukts der Unterrichtsreflexion angemessen und authentisch ist. Die Auswahl der Testinhalte erfolgte mit einem Fokus auf fachdidaktische Situationen, beinhaltet aber auch fachliche (physikalische) und pädagogische Aspekte. Die fachlichen Inhalte der Unterrichtsstunde, die Wiederholung der Newtonschen Axiome und die Einführung der Impulserhaltung, sind relevant für verschiedene Schulformen und werden üblicherweise im ersten Studiensemester unterrichtet, sodass das Instrument für eine große Gruppe an Lehramtsstudierenden relevant wird. Zudem wurden die Teilnehmenden auf einer vierstufigen Likert-Skala um eine Einschätzung u. a. der Authentizität der verwendeten Videovignetten und ihrer Eignung für die Reflexion von Unterricht gebeten. Mit  $M = 2,96$  (SD: 0,80; 4: volle Zustimmung, 1: volle Ablehnung) wurden die Vignetten insgesamt als authentisch und als geeignet für die Reflexion von Unterricht ( $M = 3,30$ ; SD: 0,62) eingestuft. Die Antwortoptionen in den MC-Aufgaben basieren auf über 150 Reflexionen von Studierenden zu den Unterrichtsausschnitten aus dem Reflexionsperformanztest und sollen so möglichst authentische Optionen darstellen. Antwortoptionen werden dann als „richtig“ bewertet, wenn sie relevante Rückmeldungen gemäß der Experteneinschätzung zu den Videovignetten im Reflexionsperformanztest darstellen. Einschränkend bleibt festzuhalten, dass das Testinstrument Unterrichtsreflexion nur in einem Ausschnitt abbilden kann und die Verwendung von MC-Aufgaben die Studierenden von der Anforderung entlastet, die relevanten Aspekte im Unterricht zu erkennen, die näher betrachtet werden sollten. Dies könnte allerdings gerade für weniger erfahrene Studierende eine Entlastung sein (Lohse-Bossenz et al., 2023).

Im Rahmen von  $N = 7$  Think-Aloud-Interviews mit Physik-Lehramtsstudierenden wurde die *kognitive Validität* des Instruments geprüft (Übersetzungsschritt II zu III). Die Transkripte der Interviews wurden inhaltsanalytisch ausgewertet. Dabei wurde deutlich, dass die Studierenden zur Lösung der Aufgaben Überlegungen anstellen, die sich explizit auf den Unterricht bzw. die Antwortoptionen beziehen oder Reflexionen des Unterrichts darstellen; grundlegende Verständnishürden in der Bearbeitung des Testinstruments traten nicht auf. Zusätzlich erfolgte eine Zweitkodierung von etwa 14,5 % der Segmente mit mittelmäßiger bis sehr guter Übereinstimmung ( $0,54 \leq \kappa_{\text{Cohen}} \leq 0,91$ ,  $p < 0,001$ ; Döring & Bortz, 2016). Eine ausführlichere Darstellung findet sich in Weißbach und Kulgemeyer (2022).

Auf Grundlage von  $N = 189$  Testbearbeitungen angehender Physik-Lehrkräfte (davon 90 im Bachelorstudium, 83 im Masterstudium, 11 im Referendariat und 5 Personen in sonstigen physiklehrschaftsbezogenen Ausbildungssituationen) erfolgt eine Prüfung der statistischen Kennwerte (Übersetzungsschritt III zu IV). Für die 32 MC-Aufgaben zum Bewerten und Vorschlagen von Alternativen ergibt sich eine sehr hohe Reliabilität von  $\alpha_{\text{Cronbach}} = 0,95$  (Döring & Bortz, 2016). Auch die Lösungshäufigkeiten der Aufgaben liegen mit durchschnittlich 0,66 (0,36 bis 0,91) in einem annehmbaren Bereich (ebd.). Ein Vergleich der Ergebnisse von Studierenden im Bachelor- und Masterstudium zeigt signifikant bessere Ergebnisse für Masterstudierende ( $d_{\text{Cohen}} = 0,68$ ,  $p < 0,001$ ). Für eine Kohorte von  $N = 22$  Studierenden zweier Universitäten zeigten sich außerdem signifikante Unterschiede in den Testergebnissen vor und nach dem Praxissemester ( $d_{\text{Cohen}} = 0,46$ ,  $p < 0,01$ ). Beide Unterschiede erscheinen insofern plausibel, da der Studienfortschritt insgesamt mit einer

größeren Anzahl an Lerngelegenheiten auch für Unterrichtsreflexion einhergeht und das Praxissemester an beiden Standorten eng begleitet wird (z. B. durch Unterrichtsnachbesprechungen mit Universitätsdozierenden) und somit eine Lerngelegenheit für die Studierenden darstellt.

Zur *externen Validierung* der Testwerte (ebenfalls Übersetzungsschritt III zu IV) wurde zunächst in einer Stichprobe von  $N = 41$  Studierenden im Praxissemester zusätzlich zum Reflexionstestinstrument ein Test zum fachdidaktischen Wissen eingesetzt (Jordans et al., 2022). Zwischen diesen beiden Instrumenten konnte kein signifikanter Zusammenhang gefunden werden. Außerdem wurde in einer Stichprobe von  $N = 39$  Studierenden auch ein Reflexionsperformanztest eingesetzt (Kempin, et al., 2018). Da das hier vorgestellte Testinstrument auf dem Reflexionsperformanztest basiert, ermöglicht der Einsatz beider Instrumente die Untersuchung von Zusammenhängen zwischen zwei Maßen für Reflexionsfähigkeit, deren Erfassung sich vorrangig in der Methode (MC-Aufgaben vs. offenes, verbales Antwortformat) unterscheidet. Da beide Instrumente gleiche Unterrichtsausschnitte beinhalten, wurde der Reflexionsperformanztest mit offenem Antwortformat zuerst bearbeitet, um eine potentielle Beeinflussung der freien Reflexionen durch zuvor gelesene Antwortoptionen zu vermeiden. Zwischen den Testwerten beider Instrumente besteht ein Zusammenhang von  $r_{\text{Pearson}} = 0,49$  ( $p < 0,01$ ).

Im Rahmen der *konsequentiellen Validierung* wurden die Studierenden bzgl. ihrer Interpretation des Feedbacks befragt (Übersetzungsschritt IV zu V): Nach einer ersten Interviewstudie mit  $N = 6$  Studierenden wurde das Feedback überarbeitet und es wurden weitere  $N = 6$  Studierende zur überarbeiteten Version interviewt. Die Interviews wurden qualitativ inhaltsanalytisch ausgewertet. In beiden Interviewstudien wurden überwiegend inhaltlich sinnvolle Einordnungen der erreichten Testwerte geäußert, wobei die Studierenden sich zumeist am Median der Vergleichsgruppe orientierten (Ergebnisse darüber werden als positiv eingestuft, solche darunter repräsentierten Entwicklungspotentiale; näheres dazu in Weißbach & Kulgemeyer, 2023). Allerdings ziehen die Studierenden kaum konkrete Schlussfolgerungen aus dem erhaltenen Feedback. Um hier zu unterstützen, wurde das Feedback um eine niedrigschwellige Empfehlung für die Arbeit mit dem Selbstlernmaterial ergänzt (z. B. sich dabei auf das Finden von Alternativvorschlägen zu konzentrieren). Inwiefern diese Empfehlungen tatsächlich umgesetzt werden bzw. einen Mehrwert für die Studierenden bieten, ist allerdings bislang nicht untersucht.

Insgesamt deuten die Ergebnisse der Validierungsstudien darauf hin, dass das Testinstrument dazu geeignet ist, den Studierenden eine fundierte Einschätzung ihrer Reflexionsfähigkeit zu ermöglichen. Auch wenn das Instrument nur einen Ausschnitt der Unterrichtsreflexion abbilden kann und die MC-Aufgaben Überlegungen der Studierenden im Vergleich zum offenen Antwortformat möglicherweise beeinflussen, stellen die Studierenden relevante Überlegungen im Kontext der Unterrichtsreflexion an und interpretieren das erhaltene Feedback vorwiegend inhaltlich sinnvoll. Auch die hohe interne Konsistenz sowie die externe Validierung mit dem Reflexionsperformanztest stützen die Eignung der Testwerte zu diesem Zweck. Einschränkungen sind in Bezug auf die Eignung des Instruments als explizite Professionalisierungsmöglichkeit festzuhalten, da die Studierenden kaum konkrete Schlussfolgerungen aus dem Feedback ziehen und auch das Selbstlernmaterial nicht umfassend als Lerngelegenheit evaluiert ist. Wegen der hohen internen Konsistenz scheint es aussichtsreich, eine zeitlich effizientere Kurzversion mit weniger Items zu entwickeln. Dies soll zukünftig angegangen werden.

## Literatur

- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Chan, K. K. H., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepert, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., ... Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, S. 77-94. Singapore: Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_2](https://doi.org/10.1007/978-981-13-5898-2_2)
- Dickmann, M. (2016). Messung Von Experimentierfähigkeiten. Validierungsstudien zur Qualität eines Computerbasierten Testverfahrens. <https://doi.org/10.5281/zenodo.168540>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. In: *Review of Educational Research* 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hatton, N. & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. In: *Teaching and Teacher Education* 11(1), 33-49. [https://doi.org/10.1016/0742-051X\(94\)00012-U](https://doi.org/10.1016/0742-051X(94)00012-U)
- Jordans, M., Zeller, J., Große-Heilmann, R. & Riese, J. (2022). Weiterentwicklung eines physikdidaktischen Tests zum Online-Assessment. In S. Habig & H. van Vorst (Hrsg.), *Unsicherheit als Element von naturwissenschaftsbezogenen Bildungsprozessen*. Gesellschaft für Didaktik der Chemie und Physik. Virtuelle Jahrestagung 2021 (Bd. 42). Duisburg-Essen: Universität Duisburg-Essen, 764-767.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement* 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>
- Kempin, M., Kulgemeyer, C. & Schecker, H. (2018). Reflexion von Physikunterricht: Ein Performanztest. In C. Maurer (Hrsg.), *Qualitätsvoller Chemie- und Physikunterricht- normative und empirische Dimensionen*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Regensburg 2017 (Bd. 38). Regensburg: Universität Regensburg, 867-870.
- Lohse-Bossenz, H., Schmitt, M., Lenske, G. & Gold, B. (2023). „The same or different?“ – Effekte von Unterrichtsanalyse und Unterrichtsreflexion auf die Veränderung kognitiver und motivationaler Merkmale professioneller Lehrkompetenz. *Zeitschrift für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-023-01197-1>
- Messick, S. (1995). Validity of Psychological Assessment. *Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning*. *American Psychologist* 50(9), 741-749.
- Oser, F., Heinzer, S., & Salzmann, P. (2010). Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. *Unterrichtswissenschaft*, 38(1), 5–28.
- Von Aufschnaiter, C., Hofmann, C., Geisler, M. & Kirschner, S. (2019). Möglichkeiten und Herausforderungen der Förderung von Reflexivität in der Lehrerbildung. *SEMINAR* 25(1), 49-60.
- Weißbach, A. & Kulgemeyer, C. (2022). Reflexion von Physikunterricht - ein Online-Assessment mit Feedback. *PhyDid B - Didaktik Der Physik - Beiträge Zur DPG-Frühjahrstagung*, 203–210.
- Weißbach, A. & Kulgemeyer, C. (2023). Reflexion von Physikunterricht: Ein Online-Assessment mit Feedback. In H. van Vorst (Hrsg.), *Lernen, Lehren und Forschen in einer digital geprägten Welt*. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Aachen 2022 (Bd. 43). Duisburg-Essen: Universität Duisburg-Essen, 993-996.
- Wyss, C. (2013). *Unterricht und Reflexion. Eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften*. Münster: Waxmann.