

Anna Weißbach<sup>1</sup>  
Christoph Kulgemeyer<sup>1</sup>

<sup>1</sup>Universität Bremen

## **Reflexion von Physikunterricht: Ein Online-Assessment mit Feedback**

Reflexion von Unterricht wird häufig als essentielle Kernaufgabe von Lehrkräften verstanden (von Aufschnaiter et al. 2019). Sie soll einerseits der Verbesserung von Unterricht dienen (ebd.) und andererseits als Mechanismus für die Generierung von Wissen aus Praxiserfahrung fungieren (McAlpine et al., 1999). Allerdings zeigt sich, dass Reflexionen von Studierenden häufig auf niedrigen Niveaus verbleiben (z. B. Hatton & Smith, 1995) und Lerngelegenheiten wie das Praxissemester nicht per se zu einer Verbesserung von Reflexionsfähigkeit führen (Kulgemeyer et al., 2021). Vor diesem Hintergrund stellt sich die Frage, wie Studierende bei der Entwicklung von Reflexionsfähigkeit unterstützt werden können. Im Rahmen des Projekts *ProfiLeP-Transfer* wurde dazu eine Lernumgebung entwickelt, die ein Diagnose-Instrument mit Assessment-Feedback und einem Fördermaterial koppelt. Nachfolgend werden die entwickelten Materialien sowie erste Ergebnisse von Validitätsstudien vorgestellt.

### **Reflexion von Physikunterricht**

Verschiedene Definitionen verstehen Reflexion zwar übereinstimmend als „(höherwertige[n]) Denkprozess“ (von Aufschnaiter, Fraij & Kost, 2019, S. 146), unterscheiden sich darüber hinaus allerdings deutlich z. B. in Bezug auf teilnehmende Personen oder den zeitlichen Bezug zur Handlung (Szogs et al., 2019). In diesem Projekt wird Reflexion von Unterricht verstanden als die „theoriegeleitete Analyse von Unterricht mit dem Ziel der Verbesserung der Unterrichtsqualität und der Entwicklung der Professionalität von Lehrkräften“ (Kempin, Kulgemeyer & Schecker, 2020, S. 439). Ergänzt wird diese Definition durch das von Nowak, Kempin, Kulgemeyer und Borowski (2019) entwickelte Modell zur Bewertung der Qualität von Reflexionen. Das Modell wurde auf Grundlage verschiedener Stufenmodelle (u. a. von Plöger, Scholl & Seifert, 2015) entwickelt und klassifiziert Aussagen in Bezug auf (1) die vorliegenden Aussagetypen bzw. *Elemente der Reflexion* (wird z. B. Unterrichtshandeln beschrieben oder bewertet, werden alternative Handlungsoptionen vorgeschlagen oder Konsequenzen für die Professionalität der Lehrperson gezogen?). Zudem wird eingestuft, ob diese Elemente (2) *begründet* werden und (3) welcher *Wissensbasis des Professionswissens* sie zuzuordnen sind (Sind es fachdidaktische, fachliche oder pädagogische Überlegungen?).

### **Die entwickelte Lernumgebung**

Zur Unterstützung der Studierenden bei der Entwicklung der eigenen Reflexionsfähigkeit wurde eine dreiteilige Lernumgebung bestehend aus einem geschlossenen Diagnose-Instrument, sich anschließendem Assessment-Feedback und einem Fördermaterial entwickelt (einsehbar unter: [www.unterrichtsreflexion.de](http://www.unterrichtsreflexion.de)). Die ausführliche Rückmeldung soll durch die Ableitung von „Verbesserungshinweise[n] die lernförderliche Wirkung“ (Sippel, 2009, S. 9) sicherstellen und fundierte Selbsteinschätzung ermöglichen. Während der Bearbeitung des Diagnose-Instruments sind Studierende dazu aufgefordert, sieben Videovignetten einer zusammenhängenden Physik-Doppelstunde, die ausgewählte physikdidaktische, physikalische und allgemeinpädagogische Probleme beinhalten, im Rahmen einer fiktiven

kollegialen Beratungssituation zu reflektieren (Fremdreflexion). Zur Reflexion der Vignetten werden 16 verschiedene Aspekte vorgegeben, zu denen jeweils eine Multiple-Choice-Aufgabe zur Bewertung sowie zu alternativen Handlungsoptionen bearbeitet wird. Die Attraktoren und Distraktoren basieren auf realen Reflexionen von insgesamt über 150 Studierenden im Praxissemester zu denselben Videovignetten und sollen so authentische Antwortoptionen abbilden (näheres in Weißbach und Kulgemeyer (2022)). Nach der Bearbeitung des Diagnose-Instruments erhalten Studierende ein individuelles Feedback über ihr Abschneiden. Das Assessment-Feedback beinhaltet neben kurzen Erläuterungen zum Diagnose-Instrument und zur Unterrichtsreflexion Boxplots, in denen die eigenen (Gesamt- und Teil-) Ergebnisse in eine Vergleichsgruppe eingeordnet werden sowie den Hinweis auf ein Fördermaterial. Dieses Material ermöglicht den Studierenden im Sinne des „feed forward“ (Hattie & Timperley, 2007, S. 86) eine selbstständige Beschäftigung mit Unterrichtsreflexion, indem drei weitere Unterrichtsausschnitte angeleitet reflektiert werden können. Dazu stehen allgemeine Leitfragen, spezifische Tipps und eine Musterlösung als Hilfestellungen zur Verfügung.

### Validitätsargumentation zur Evaluation der Lernumgebung

Die einzelnen Bestandteile der Lernumgebung sollen es einerseits ermöglichen, die Fähigkeit, Unterricht zu reflektieren, valide zu erfassen und andererseits ökonomisch (regelmäßiges) Feedback nach Bedarf bereitzustellen, um die weitere Beschäftigung mit Unterrichtsreflexion zu motivieren und die Studierenden bei der Entwicklung ihrer Reflexionsfähigkeit zu unterstützen. Zur Evaluation dieser Anforderungen wird im Sinne des Argument-based-Approach nach Kane (2013) eine Argumentation für die Validität der Testwerte als Maß für die Reflexionsfähigkeit der Studierenden entwickelt. In Anlehnung an Dickmann (2016) werden dazu die Übersetzungsschritte, ausgehend vom Konstrukt der Reflexionsfähigkeit über das entwickelte Diagnose-Instrument, das dadurch evozierte Verhalten, die bestimmten Testwerte bis hin zu den Schlussfolgerungen, die sich für Studierende aus ihrer Rückmeldung ergeben, anhand verschiedener Anforderungen evaluiert (s. Abbildung 1).



Abb. 1: Für die Validitätsargumentation zu evaluierende Anforderungen

### Erste Ergebnisse der Evaluation

Zur Evaluation der Anforderungen an die einzelnen Übersetzungsschritte werden unterschiedliche Erhebungsmethoden genutzt. So wird das bei der Bearbeitung des Diagnose-Instruments evozierte Verhalten im Rahmen von Think-Aloud-Interviews erhoben. Die Analyse der Testwerte erfolgt auf Grundlage statistischer Kennwerte und die Interpretation der Rückmeldung und Wahrnehmung des Fördermaterials durch die Studierenden wird in leitfadengestützten Interviews erfragt. In Bezug auf die Anforderungen zu *II. Verhalten* zeigt sich in  $N = 7$  Think-Aloud-Interviews, dass die Studierenden während der Bearbeitung des Diagnose-Instruments überwiegend konstruktrelevante Überlegungen anstellen, d. h. entweder (Elemente von) Reflexionen oder sonstige Aussagen mit direktem Bezug zum dargestellten Unterricht oder zur eigenen Professionalisierung äußern. Diese Ergebnisse

stützten also die Validitätsargumentation. Etwa 14,5 % der vorliegenden Segmente wurden zusätzlich von einer zweiten Person kodiert, wobei sich für die verschiedenen Kategorien jeweils signifikante mittelmäßige bis sehr gute Übereinstimmungen von  $0,54 \leq \kappa_{\text{Cohen}} \leq 0,91$  ( $p < 0,001$ ) ergeben (Döring & Bortz, 2016). Geringe, im Rahmen der Interviews festgestellte, Überarbeitungsbedarfe in Bezug auf Formulierungen im Testinstrument sind umgesetzt und somit nicht gefährdend für die Validität. Die Prüfung der statistischen Kennwerte zu *III. Testwerte* ist mit dem bisherigen Datenpool ( $N = 89$  Bearbeitungen) aufgrund der noch nicht abgeschlossenen Erhebungen als vorläufig zu betrachten. Auch hier ergeben sich allerdings bislang Ergebnisse, die die Validitätsargumentation stützen: Die Multiple-Choice-Aufgaben zum Bewerten der Unterrichtsausschnitte und zum Vorschlagen von Alternativen weisen eine hohe interne Konsistenz von  $\alpha_{\text{Cronbach}} = 0,955$  auf. Für die Aufgaben ergeben sich außerdem akzeptable Aufgabenschwierigkeiten von  $\bar{\sigma} 0,63$  (0,36 bis 0,88) bei durchschnittlich hoher Trennschärfe von  $\bar{\sigma} 0,62$  (0,26 bis 0,81) (Döring & Bortz, 2016). Gleichzeitig zeigt sich kein signifikanter Zusammenhang zwischen Testwerten und Fachsemester der Studierenden, welcher aufgrund der zunehmenden Anzahl an Lerngelegenheiten im Verlauf des Studiums erwartbar wäre. Die Evaluation von Rückmeldung und Fördermaterial zu *IV. Rückmeldung* erfolgt durch eine qualitative Inhaltsanalyse von  $N = 6$  leitfadengestützten Interviews. Hier zeigt sich, dass die Studierenden das erhaltene Assessment-Feedback überwiegend sinnvoll interpretieren. So werden Ergebnisse primär anhand der Einordnung in die Vergleichsgruppe bewertet, was i. d. R. in einer positiven Bewertung von Ergebnissen über dem Median resultiert (z. B. „[D]iese Rückmeldung ist ja sehr positiv[, w]eil das Ergebnis hier im oberen Mittel liegt“) oder zur Identifikation von Verbesserungspotential bei Ergebnissen unter dem Median führt, die allerdings kaum mit konkreten Schlussfolgerungen verbunden ist (z. B. „Schülvorstellungen würde ich mir dann wahrscheinlich nochmal angucken, hätte ich die Zeit“). Dem Fördermaterial wird von den Studierenden insgesamt ein Mehrwert attestiert (z. B. „[I]ch fand [...] das Material hilfreich, um überhaupt die Situation einzuschätzen.“). Auch im Fördermaterial sind auf Grundlage der Interviews leichte Änderungen zur Reduktion von Redundanz umgesetzt worden. Insgesamt sprechen diese Ergebnisse also dafür, dass das bereitgestellte Assessment-Feedback in Kombination mit dem Fördermaterial die gestellten Anforderungen erfüllt und somit einen Beitrag zur Validität liefert. Zusätzlich gestützt wird diese Schlussfolgerung durch vergleichbare Ergebnisse einer ähnlichen Interviewstudie ( $N = 6$ ), die zu einer früheren Version der Rückmeldung durchgeführt wurde: Auch hier konnte festgestellt werden, dass die Studierenden sich bei der Einordnung ihrer Ergebnisse häufig an den Ergebnissen der Vergleichsgruppe orientieren und als Schlussfolgerung der Rückmeldung Verbesserungspotentiale formulieren (Weißbach & Kulgemeyer, 2022).

### **Ausblick**

Das geschlossene Aufgabenformat im Diagnose-Instrument geht im Vergleich zu realen Reflexionshandlungen mit einer stark reduzierten Authentizität einher. Der im Rahmen der Projekts *ProfiLeP+* entwickelte Performanztest zur Unterrichtsreflexion (Kempin, Kulgemeyer & Schecker, 2018) ermöglicht im Vergleich dazu authentisches Handeln bei gleichzeitiger Standardisierung der Testsituation. Um die Ergebnisse des Diagnose-Instruments mit Ergebnissen authentischeren Reflexionshandelns in Verbindung zu setzen, sollen künftig Zusammenhänge zwischen Ergebnissen beider Instrumente untersucht werden.

## Literatur

- Dickmann, M. (2016). *Messung Von Experimentierfähigkeiten. Validierungsstudien zur Qualität eines Computerbasierten Testverfahrens*. DOI: <https://doi.org/10.5281/zenodo.168540>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer. DOI: <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), S. 81-112. DOI: <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), S. 33-49. DOI: [https://doi.org/10.1016/0742-051X\(94\)00012-U](https://doi.org/10.1016/0742-051X(94)00012-U)
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement* 50(1), S. 1-73. DOI: <https://doi.org/10.1111/jedm.12000>
- Kempin, M., Kulgemeyer, C. & Schecker, H. (2018). Reflexion von Physikunterricht: Ein Performanztest. In Maurer, C. (Hrsg.), *Qualitätsvoller Chemie- und Physikunterricht – normative und empirische Dimensionen. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Regensburg 2017*. Regensburg: Universität Regensburg, S. 867-870.
- Kempin, M., Kulgemeyer, C. & Schecker, H. (2020). Wirkung von Professionswissen und Praxisphasen auf die Reflexionsfähigkeit von Physiklehramtsstudierenden. In: Habig, S. (Hrsg.), *Naturwissenschaftliche Kompetenz in der Gesellschaft von morgen. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Wien 2019*. Essen: Duisburg-Essen, S. 439-442.
- Kulgemeyer, C., Kempin, M., Weißbach, A., Borowski, A., Buschhüter, D., Enkrott, P., Reinhold, P., Riese, J., Schecker, H., Schröder, J. & Vogelsang, C. (2021). Exploring the impact of pre-service science teachers' reflection skills on the development of professional knowledge during a field experience. *International Journal of Science Education*, 43(18), S. 3035-3057. DOI: <https://doi.org/10.1080/09500693.2021.2006820>
- McAlpine, L., Weston, C., Beauchamp, C., Wiseman, C., & Beauchamp, J. (1999). Monitoring student cues: Tracking student behaviour in order to improve instruction in Higher Education. *The Canadian Journal of Higher Education*, 29(3), S. 113-144. DOI: <https://doi.org/10.47678/cjhe.v29i3.183335>
- Nowak, A., Kempin, M., Kulgemeyer, C & Borowski, A. (2019). Reflexion von Physikunterricht. In: Maurer, C. (Hrsg.). *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe*. Regensburg: Universität Regensburg, 838-841.
- Plöger, W., Scholl, D. & Seifert, A. (2015). Analysekompetenz - ein zweidimensionales Konstrukt?! *Unterrichtswissenschaft. Zeitschrift für Lernforschung* 43(2), S. 166-184.
- Sippel, S. (2009). Zur Relevanz von Assessment-Feedback in der Hochschullehre. In: *Zeitschrift für Hochschulentwicklung* 4(1), S. 1-22.
- Szogs, M., Kobl, C., Volmer, M. & Korneck, F. (2019). Bedeutsamkeit von Reflexion und Reflexivität in der Professionalisierung von Lehrkräften sowie ihre Beziehung zu anderen Prozessen und Konstrukten. In: Maurer, C. (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018*. Regensburg: Universität Regensburg, S. 317-320.
- Von Aufschnaiter, C., Fraij, A. & Kost, D. (2019). Reflexion und Reflexivität in der Lehrerbildung. In: *Herausforderung Lehrer\_innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion* 2(1), S. 144-159. DOI: <https://doi.org/10.4119/UNIBI/hlz-144>
- Von Aufschnaiter, C., Hofmann, C., Geisler, M. & Kirschner, S. (2019). Möglichkeiten und Herausforderungen der Förderung von Reflexivität in der Lehrerbildung. In: *SEMINAR* 25(1), S. 49-60. Weißbach, A. & Kulgemeyer, C. (2022). Reflexion von Physikunterricht – ein Online-Assessment mit Feedback. In: Habig, S. & van Vorst, H. (Hrsg.), *Unsicherheit als Element von naturwissenschaftsbezogenen Bildungsprozessen. Gesellschaft für Didaktik der Chemie und Physik, Virtuelle Jahrestagung 2021*. GDCCP, S. 756-759.