

Entwicklung eines interaktiven Videovignetentests für Erklärfähigkeit

Im hier vorgestellten Projekt soll ein interaktives, computerbasiertes Testinstrument entwickelt werden, das die Performanz von Lehrkräften in einer bestimmten Standard-situation des Unterrichtens vorhersagen kann: dem Erklären physikalischer Sachverhalte. Dazu wird ein Videovignetentest entwickelt, der aus zweistufigen Multiple-Choice Items besteht und Large-Scale Erhebungen ermöglicht. Mit diesem Test soll die Lücke zwischen Performanztests (hohe Validität, niedrige Testökonomie) und Papier-Bleistift-Tests (niedrige Validität, hohe Testökonomie) geschlossen werden. Schriftliche Tests sind Standard in der Untersuchung der professionellen Handlungskompetenz von Lehrkräften, ihre Handlungsvalidität ist jedoch fraglich (Vogelsang & Reinhold, 2013, Cauet, 2016).

1. Erklären und Lehrerhandeln

Der Prozess des Erklärens wird in diesem Beitrag als ein konstruktivistischer, dialogischer Prozess zwischen einem Erklärer (z. B. eine Lehrkraft) und einem oder mehreren Adressaten (z. B. Lernende) verstanden und ist von der Vorstellung einer transmissiven Wissensvermittlung abzugrenzen (Kulgemeyer & Schecker, 2013). Ein Kernelement für erfolgreiches Erklären ist dabei die Adaption der Erklärung an die Bedürfnisse der Adressaten (Wittwer & Renkl, 2008). Ein Modell für diese Art des dialogischen Erklärens im Physikunterricht wurde von Kulgemeyer und Schecker (2013) vorgelegt. Es beschreibt zwei Dimensionen, die eine Lehrkraft berücksichtigen muss, wenn sie ihren Lernenden etwas erklärt: Die Sachgerechtigkeit (ist das erklärte Thema fachlich korrekt und vollständig?) und die Adressatengemäßheit (knüpft die Erklärung an Vorwissen und Interesse der Adressaten an?). Dazu stehen vier „Variablen“ zur Verfügung, die die Lehrkraft zur Adaption der Erklärung nutzen kann. Dies sind (1) der Einsatz von Beispielen oder Analogien, (2) der sprachliche Code (z. B. Alltagssprache oder Fachsprache), (3) die Darstellungsform (z. B. ein Foto oder ein Diagramm), und (4) der Mathematisierungsgrad (z. B. der Einsatz von Formeln oder Zahlenbeispielen). Die Adressaten reagieren auf die angebotene Erklärung mit einer (non)verbalen Rückmeldung. Diese Rückmeldung kann der Erklärer nutzen, um seine Erklärung entsprechend der vier Variablen anzupassen und das Verstehen der Adressaten besser zu fördern.

Nach Riese et al. (2015) ist die Erklärfähigkeit von Lehrkräften Teil ihrer professionellen Handlungskompetenz. Diese kann mit verschiedenen Testformaten erhoben werden, die im Sinne von Miller (1990) in vier Kategorien unterteilt werden können: *Beobachtungen realen, beruflichen Handelns* finden in realen Umgebungen mit sehr vielen Freiheitsgraden statt (z. B. Videografieren von Unterrichtsstunden). *Performanztests* versuchen solche Situationen in standardisierten Umgebungen nachzustellen und so die Zahl der Freiheitsgrade zu reduzieren. *Kompetenztests* erfolgen typischerweise schriftlich und untersuchen, ob Testpersonen das professionelle Wissen zum Lösen typischer beruflicher Problemsituationen haben, während (4) *Wissenstests* deklaratives Wissen erheben. Obwohl schriftliche Tests den niedrigsten Grad an Authentizität bezüglich realer unterrichtlicher Handlungen haben, finden sie in der naturwissenschaftsdidaktischen Forschung häufig Anwendung. Mit Blick auf eine durch mehr Authentizität bedingte, höhere prognostische Validität von Unterrichtsqualität werden jedoch alternative Testverfahren gefordert (z. B. von Aufschnaiter und Blömeke, 2010).

In diesem Sinne haben Kulgemeyer und Tomczyszyn (2015) den *Dialogischen Erklärtest* (DET) entwickelt, einen Performanztest, der mit Videoauswertung arbeitet und die Erklärfähigkeit einzelner Probanden in einem standardisierten Setting misst. Sie konnten damit zeigen, dass Erklärfähigkeit sowohl durch das Fachwissen als auch das fachdidaktische Wissen der Probanden beeinflusst wird (s. Beitrag von C. Kulgemeyer in diesem Band). Der DET ermöglicht aufgrund der Authentizität der Erhebungssituation eine valide Abbildung von Lehrerklärungen. Die Datenerhebung und Auswertung sind jedoch sehr zeitaufwändig, was Large-Scale Erhebungen fast unmöglich macht. Diese sind jedoch von Interesse, soll der Effekt von Elementen der Lehrerbildung (z. B. Praxissemester) auf die Erklärfähigkeit von Lehramtsstudierenden untersucht werden. Ein vielversprechender Ansatz zur testökonomischen und validen Messung der Handlungskompetenz von Lehrkräften ist der Einsatz von Videovignettentests (Rehm und Bölsterli, 2014). Bei typischen Instrumenten handelt es sich um Kompetenztests, die den Probanden durch kurze Videosequenzen, die an relevanten Stellen unterbrochen werden, in eine bestimmte Situation versetzen und das anschließende Verhalten des Probanden erheben. Im Gegensatz zu Performanztests weisen bisherige Verfahren jedoch einen geringen Grad an Interaktivität auf (Lindmeier, 2013).

2. Anlage des Projekts

Im hier vorgestellten Projekt wird ein Videovignettentest entwickelt, der interaktives Handeln ermöglicht. Er soll die Erklärfähigkeit von Lehrkräften in Erklärsituationen im Unterricht vorhersagen können und als standardisiertes Instrument mit geschlossenen Antworten die Umsetzung von Large-Scale Assessments ermöglichen. Dazu werden Videovignetten erstellt, die an Erklärungen angelehnt sind, welche mit dem DET aufgenommen wurden. Sie zeigen eine Lehrkraft, die im Dialog versucht, einem Lernenden einen physikalischen Sachverhalt zu erklären. Bei den Sachverhalten handelt es sich um Themen aus der Mechanik, die in einen Kontext eingebettet sind, z. B. „Wie kann die Erde vor einem herannahenden Asteroiden durch dessen Sprengung gerettet werden?“ (physikalischer Inhalt: Impulserhaltung). Beispiele sind in Bartels & Kulgemeyer (2016) aufgeführt. Jeder Sachverhalt wird mit etwa 15 Videovignetten realisiert, die eine zusammenhängende Erklärsituation im Dialog zeigen. Der Test ist online zugänglich. Während der Proband den Test bearbeitet, stoppen die Videovignetten an Stellen, an der die Lehrkraft eine Entscheidung über das weitere Vorgehen in der Erklärung treffen muss. Der Proband entscheidet dann in zweistufigen Items über den Fortgang der Erklärung (*Two-Tier-Multiple-Choice Items* vgl. z.B. Tan, et al., 2002). In der ersten Stufe muss er aus jeweils vier gegebenen Möglichkeiten zum Fortführen der Erklärung diejenige auswählen, die er für am besten geeignet hält. Dabei sind alle Antwortmöglichkeiten fachlich korrekt – sie unterscheiden sich nur in der Art der Erklärung (z. B. Unterschiede im Grad der eingesetzten Fachsprache). Für seine Auswahl in dieser Stufe hat der Proband nur begrenzt Zeit, was den Handlungsdruck in einer realen Unterrichtssituation simuliert und damit die Authentizität des Tests erhöht (vgl. Rehm und Bölsterli, 2014). In der zweiten Stufe wird der Proband gebeten, seine Auswahl zu begründen. Diese Stufe besteht ebenfalls aus geschlossenen Antwortmöglichkeiten, die durch einen adaptiven Testverlauf zur in der ersten Stufe getroffenen Auswahl passen.

Nachdem der Proband seine Auswahl getroffen hat, zeigt die nächste Vignette das tatsächliche (und nicht immer optimale) Verhalten der Lehrkraft und die anschließende Reaktion des Lernenden. Der Dialog wird so fortgeführt. Jede Videovignette bezieht sich auf eine der vier Variablen des Erklärens. Im Verlauf jedes Sachverhalts werden jedoch alle vier Variablen mehrfach behandelt.

Mit diesem Instrument sollen folgende Forschungsfragen beantwortet werden:

FF1: Wie valide und reliabel kann Erklärfähigkeit mit einem standardisierten, interaktiven Instrument computerbasiert gemessen werden?

FF2: Inwieweit steht die durch das Instrument erfasste Erklärfähigkeit von Lehrkräften mit dem Lernzuwachs ihrer Lernenden in Zusammenhang?

3. Stand der Testentwicklung

Zurzeit läuft eine Pilotierung des vorläufigen Testinstruments. Dafür wurde ein Videovignettentest mit 15 Vignetten realisiert. In einer Expertenrunde wurden alle Items des Tests bezüglich der Frage diskutiert, ob die Antwortmöglichkeiten mit Blick auf das Verhalten der Schülerin in der zuvor gezeigten Videovignette eindeutig richtig oder falsch seien. Der so entwickelte Test wurde an 20 Probanden erprobt. Um möglichst große Unterschiede in der Erklärfähigkeit zu haben, enthielt die Stichprobe sowohl Physiklehramtsstudierende aus dem ersten und letzten Studienjahr, als auch Studierende der Fachphysik. Die zweite Stufe der Items wurde in einem offenen Format durchgeführt, um möglichst alle Überlegungen der Probanden zu erfassen. Mit Blick auf die Inhaltsvalidität wurde ferner mit 10 Probanden ein Stimulated-Recall-Interview durchgeführt, in dem die Testpersonen ausführlich zu ihren Erwägungen während des Entscheidungsprozesses bei der Beantwortung der Items befragt wurden. Aus den so gewonnenen Ergebnissen wurde eine überarbeitete Version des Testinstruments mit angepassten Videovignetten erstellt.

4. Ausblick

Aktuell werden zwei weitere Sachverhalte entwickelt und pilotiert. Hier sollen ebenfalls eine Expertenrunde und Stimulated-Recall-Interviews eingesetzt werden. Ferner wird eine Untersuchung zur kongruenten Validität vorbereitet. Hierzu sollen Probanden sowohl den DET als auch den Videovignettentest durchführen. Da Studien zum DET bereits sehr belastbare Argumente für die Validität des Verfahrens erbrachten (Kulgemeyer und Tomczyszyn, 2015), wäre eine hohe Korrelation der Ergebnisse auch ein starkes Argument für die Validität im Zusammenhang mit dem Vignettentest. Anschließend soll die Haupterhebung mit mindestens 100 Probanden (Lehramtsstudierende und Lehrkräfte) durchgeführt werden, um umfangreiche statistische Analysen zu ermöglichen. Falls sich das Instrument hier als belastbar hinsichtlich Validität und Reliabilität erweist, wäre dies ein starkes Argument für die Generalisierbarkeit des Testverfahrens und würde die Beantwortung der FF 1 ermöglichen.

Ferner werden Untersuchungen mit mindestens 10 Schulklassen von Lehrkräften durchgeführt, die zuvor am Vignettentest teilgenommen haben. Mit einem Pre-Post Test wird der Lernzuwachs der Lernenden durch eine kontrollierte Erklärsituation mit den Lehrkräften untersucht. Die Lehrkräfte erklären dabei jeweils ein zuvor festgelegtes Szenario, das – wie die Videovignettentests – thematisch im Bereich Mechanik angesiedelt ist. Als Testinstrument sollen hier Subskalen des Force Concept Inventory (vgl. z. B. Hestenes et al., 1992) eingesetzt werden. Von dieser Untersuchung werden erste Hinweise auf einen positiven Effekt der Erklärfähigkeit von Lehrkräften auf den tatsächlichen Lernzuwachs ihrer Lernenden erwartet. Sie soll die Beantwortung der FF 2 ermöglichen und Argumente für die prognostische Validität des Vignettentests liefern.

Das so entwickelte Instrument soll einen Beitrag zur validen Messung der Handlungskompetenz von Lehrkräften leisten. Langfristig wird mit solchen Verfahren die Untersuchung der Auswirkungen von Studienelementen der Lehrerbildung auf die tatsächliche Handlungskompetenz von zukünftigen Lehrkräften im Klassenraum angestrebt.

Literatur

- Von Aufschnaiter & C.; Blömeke, S. (2010): Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata. In: Zeitschrift für Didaktik der Naturwissenschaften 16, S. 361–367.
- Bartels, H. & Kulgemeyer, C. (2016): Entwicklung eines interaktiven computerbasierten Tests für Erklärbarkeit. In PhyDid B - Didaktik der Physik - Beiträge zur DPG Frühjahrstagung.
- Cauet, E. (2016): Testen wir relevantes Wissen? Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten. Berlin: Logos Berlin (Studien zum Physik- und Chemielernen, 204).
- Hestenes D., Wells M. & Swackhamer G. 1992: Force Concept Inventory. *The Physics Teacher* 30: 141-166.
- Kulgemeyer, Christoph; Schecker, Horst (2013): Students Explaining Science - Assessment of Science Communication Competence. In: *Res Sci Educ* 43 (6), S. 2235–2256.
- Kulgemeyer, C. & Tomczyszyn, E. (2015): Physik erklären. Messung der Erklärbarkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. In: *Zeitschrift für Didaktik der Naturwissenschaften* 21 (1), S. 111–126.
- Lindmeier, A. (2013): Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In: Klaas Macha und Ulrich Riegel (Hg.): *Videobasierte Kompetenzforschung in den Fachdidaktiken*. Münster: Waxmann, S. 45–61.
- Miller, G. (1990): The Assessment of Clinical Skills / Competence / Performance. In: *Journal of the Association of American Medical Colleges* 65 (9), S. 563–567.
- Rehm, M. & Bölsterli, K. (2014): Entwicklung von Unterrichtsvignetten. In: Dirk Krüger, Ilka Parchmann und Horst Schecker (Hg.): *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin, Heidelberg: Springer Spektrum, S. 213–225.
- Riese, J.; Kulgemeyer, C.; Zander, S.; Borowski, A.; Fischer, H. E.; Gramzow, Y. et al. (2015): Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In: *Zeitschrift für Pädagogik* (61), S. 55–79.
- Tan, K.; Goh, N. K.; Chia, L. S. & Treagust, D. (2002): Development and Application of a Two-Tier Multiple Choice Diagnostic Instrument To Assess High School Students' Understanding of Inorganic Chemistry Qualitative Analysis. In: *Journal of Research in Science Teaching* 39 (4), S. 283–301.
- Vogelsang, C. & Reinhold, P. (2013): Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. In: *Zeitschrift für Didaktik der Naturwissenschaften* 19, S. 103–128.
- Wittwer, J. & Renkl, A. (2008): Why Instructional Explanations Often Do Not Work: A Framework for Understanding the Effectiveness of Instructional Explanations. In: *Educational Psychologist* 43 (1), S. 49–64.