

Strukturelle- und externe Validierung eines Experimentiertests

Motivation

Eines der zentralen Ziele der Ausbildung in den Naturwissenschaften ist die Vermittlung von Experimentierkompetenz (KMK, 2005). In der universitären Ausbildung werden dafür Praktika eingesetzt, deren Effizienz allerdings kritisch diskutiert wird (Schumacher & Planinšič, 2007). Während es verschiedene Modernisierungen und Neustrukturierungen zu Praktika gibt (Schumacher & Planinšič, 2007), fehlen bisher gesicherte Erkenntnisse zum Erwerb von Experimentierkompetenz als zentralem Ziel. Dies liegt vor allem darin begründet, dass sich die Erfassung von Experimentierkompetenz im Hinblick auf die Reliabilität und Validität problematisch gestaltet (Shavelson et al., 1999; Emden, 2011; Heidrich et al., 2014).

Experimentierkompetenz

Experimentieren kann als potentiell iterative Abfolge von Handlungen verstanden werden (Klahr & Dunbar, 2000). Diese Abfolge lässt sich prinzipiell in drei Phasen gliedern: Vorbereitung, Durchführung und Auswertung (Emden, 2011). Die erfolgreiche Bearbeitung der einzelnen Phasen erfordert jeweils spezifische Fähigkeiten und Fertigkeiten (vgl. Emden, 2011; z. B. Maiseyken et al., 2011; vgl. Abb. 1). Experimentierkompetenz lässt sich dabei bewerten anhand: (I) der Richtigkeit der Durchführung einzelner Fähigkeiten und Fertigkeiten (Maiseyken et al., 2011); (II) der Strukturiertheit bei der Abfolge von Fähigkeiten und Fertigkeiten (Emden, 2011); und (III) der Zielorientiertheit beim Experimentierprozess mit Blick auf die Lösungsstrategie (Klahr & Dunbar, 2000).

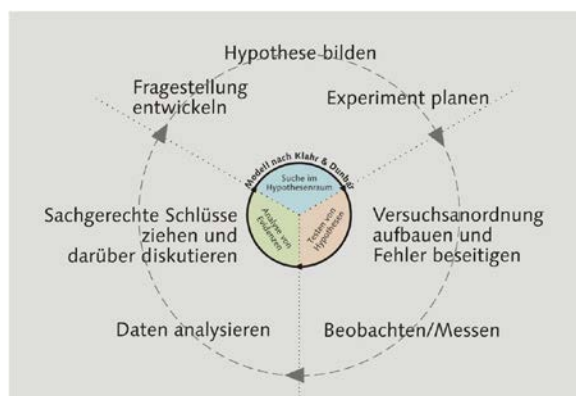


Abbildung 1: Fähigkeiten und Fertigkeiten im Experimentierprozess

Forschungsziel

Das Ziel des vorliegenden Projekts ist es, die Probleme bzgl. der Validität gezielt zu untersuchen und basierend auf den Erkenntnissen ein valides Testinstrument zu entwickeln. Im ersten Schritt wurde basierend auf einer Fachinhaltsanalyse ein Experimentiertest für den Themenbereich Optik des physikalischen Anfängerpraktikums in der universitären Grundausbildung entwickelt und die inhaltliche Validität geprüft. Anschließend wurden mittels einer Videostudie zum lauten Denken produkt- und prozessorientierte Auswerteverfahren bzgl. der kognitiven Validität untersucht (Heidrich et al., 2014). Im vorliegenden Teil des Projekts sollten darauf aufbauend die strukturelle und externe Validität des Instruments untersucht werden (Messick, 1995). Dafür werden theoretisch erwartete Strukturen und Zusammenhänge innerhalb des Modells und zu anderen Konstrukten mit den tatsächlich beobachteten Strukturen und Zusammenhängen aus den vorliegenden Daten verglichen. Die

Qualität des Experimentierens lässt sich an den drei unabhängigen Aspekten Richtigkeit, Strukturiertheit und Zielorientiertheit beurteilen (vgl. Experimentierkompetenz). Bei der Erfassung von Experimentierkompetenz werden diese drei Aspekte im Sinne von Subskalen für jede Aufgabe erfasst und daraus die Experimentierkompetenz berechnet. Testtheoretisch bedeutet das, es werden ausreichende Reliabilitäten für die drei Subskalen und das daraus zusammengesetzte Maß für die Experimentierkompetenz erwartet (*S1*). Gleichzeitig werden mit den Subskalen verschiedene Aspekte gemessen, entsprechend spricht man von einem formativen Messmodell und eine Korrelation zwischen den Aspekten wird nicht vorausgesetzt (*S2*). Wenn sich mit dem Experimentiertest Experimentierkompetenz erfassen lässt, dann sollte die Testleistung der Experimentierkompetenz mit der Testleistung basierend auf gängigen Auswerteverfahren von Experimentiertests oder der Bewertung durch den Praktikumsleiter im Sinne einer konvergenten Validierung zusammenhängen (auch wenn die gängigen Auswerteverfahren noch Probleme bzgl. der Reliabilität und Validität ausweisen). Vor allem der Aspekt der Richtigkeit sollte die Scores der alternativen Auswerteverfahren vorhersagen können, da diese zumeist auf Modellierungen beruhen, die Richtigkeit als einziges Qualitätskriterium beim Experimentieren beschreiben (*S3*). Der Experimentiertest soll Experimentierkompetenz als Konstrukt messen und kein anderes „nahes“ Konstrukt wie z. B. Fachwissen soll (übermäßig) testrelevant sein (construct-irrelevant variance, Messick, 1995). Entsprechend sollten sich im Sinne einer divergenten Validität höchstens geringe Zusammenhänge der Testleistung mit anderen „nahen“ Konstrukten beobachten lassen (*S4*).

Design

Die Untersuchung der strukturellen und externen Validität wurde im Rahmen des physikalischen Anfängerpraktikums an der Christian-Albrechts-Universität zu Kiel durchgeführt. Einer der sechs regulären Optikversuche wurden durch den Experimentiertest und eine Begleiterhebung ersetzt. In der Begleiterhebung wurden persönliche Informationen, Praktikumsnote in Optik, Fachwissen (Kröger et al., 2014), Selbstwirksamkeit (adaptiert von Rabe et al., 2012) und Strategiewissen (Thillmann, 2007) erhoben. Insgesamt bearbeiteten $N=58$ Studierende im Hauptfachpraktikum den Experimentiertest. Die Daten wurden analog zur Studie zum lauten Denkens mit dem neu entwickelten ökonomischen Auswerteverfahren ausgewertet (vgl. Heidrich et al., 2014). Um zu untersuchen, ob es sich beim gemessenen Konstrukt um Experimentierkompetenz handelt, wurde die Testleistung mit alternativen- in Experimentiertests gängigen- Auswertemethoden bestimmt und mit dem Score der Experimentierkompetenz verglichen. Dazu wurden die folgende Verfahren gewählt: (I) Es wurde das Endergebnis bewertet; (II) es wurde jeweils ein globales Maß für die Aspekte der Experimentierkompetenz kodiert; (III) es wurden die Aufzeichnungen qualitativ bewertet; (IV) es wurde eine Checkliste als Vergleich zur Musterlösung eingesetzt.

Ergebnisse

Zur Untersuchung der Struktur (*S1*) wurde Cronbachs α bestimmt. Es zeigt sich, dass das Gesamtmaß ($\alpha=.67$) sowie die Subskalen für Richtigkeit ($\alpha=.60$) und Zielorientiertheit ($\alpha=.73$) reliabel abgebildet werden konnten. Die Subskala für die Strukturiertheit muss noch weiter angepasst werden ($\alpha=.54$). Für (*S2*) wurden die Korrelationen der Aspekte untereinander untersucht. Dabei konnten maximal geringe Korrelationen von .33, $p < .01$ gefunden werden. Zur Untersuchung von (*S3*) wurde analysiert wieviel Varianz durch die Aspekte der Experimentierkompetenz aufgeklärt werden konnten. Die Ergebnisse finden sich zusammen mit den Faktorladungen in Tabelle 1. Zur Untersuchung von (*S4*) wurden die Zusammenhänge der externen Maße Praktikumsnote, Fachwissen, Selbstwirksamkeit und Strategiewissen mit den Aspekten der Experimentierkompetenz verglichen. Die Praktikumsnote korreliert mit .47, $p < .001$ mit der Richtigkeit. Das Fachwissen weist keine ausreichende interne

Konsistenz ($\alpha=.40$) auf, um Zusammenhänge zu untersuchen. Selbstwirksamkeit ($\alpha=.81$) und Strategiewissen ($\alpha=.80$) weisen keine Zusammenhänge zur Experimentierkompetenz auf.

Tabelle 1: Vorhersage alternativer Auswerteverfahren durch Experimentierkompetenz

	Aufgeklärte Varianz	Faktorladungen
Endergebnis	.518	$b_R=.566^{***}$
Global Richtigkeit	.468	$b_R=.527^{***}$
Global Strukturiertheit	.527	$b_R=.482^{***}$; $b_Z=.413^{***}$
Global Zielorientiertheit	.506	$b_R=.403^*$; $b_Z=.503^{**}$
Bewertung Prüfer	.701	$b_R=.816^{***}$
Checkliste	.622	$b_R=.646^{***}$; $b_S=.249^*$

Diskussion

Das Ziel, strukturell valide Daten zu generieren, wurde fast erreicht. Die Reliabilität (ausschließlich Strukturiertheit) ist ausreichend, das latente Messmodell konnte bestätigt werden und die alternativen Auswerteverfahren konnten zu 47-70 % aufgeklärt werden. Zum vollständigen Erreichen des Ziels muss der Aspekt der Strukturiertheit weiter ausgeschärft werden. Erwartungskonform war der Stellenwert der Richtigkeit bzgl. anderer Auswerteverfahren zentral. Allerdings gibt es bisher, auch wegen fehlender Vergleichsmöglichkeiten, kaum Erkenntnisse zur Passung der Strukturen der Aspekte Strukturiertheit und Zielorientiertheit. Bei der externen Validität konnte ein Zusammenhang zur Praktikumsnote, als wichtigstem Indikator gefunden werden. Es konnten keine Zusammenhänge zu Fachwissen, Selbstwirksamkeit und Strategiewissen gefunden werden. Eine nähere Analyse zeigte, dass die Erfassung des Fachwissens (fehlende Reliabilität) und das Strategiewissen (zu leichter Test) nicht adäquat gelungen sind. Entsprechend kann nur bestätigt werden, dass Selbstwirksamkeit nicht testrelevant war und bei der externen Validität weitere Evidenz generiert werden muss.

Literatur

- Emden, M. (2011). Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens: Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I. Univ., Diss.--Duisburg-Essen, 2011. Studien zum Physik- und Chemielernen: Vol. 118. Berlin: Logos.
- Heidrich, J., Petersen, S., & Neumann, K. (2014). Kognitive Validierung eines Experimentiertests durch Think Alouds. In S. Bernholt (Ed.), Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht. GDCP Jahrestagung in München 2013 (pp. 180–182). Kiel: IPN.
- Klahr, D., & Dunbar, K. (2000). Exploring science: The cognition and development of discovery processes. Cambridge, Mass.: MIT Press.
- Kröger, J., Neumann, K., & Petersen, S. (2014). Erfassung des Professionswissens angehender Physiklehrkräfte im Rahmen des Projekts KiL. In S. Bernholt (Ed.), Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht. GDCP Jahrestagung in München 2013 (pp. 117–119). Kiel: IPN.
- Maiseyenko, V., Nawrath, D., & Schecker, H. (2011). Modellbasierte Diagnose und Förderung experimenteller Kompetenz. In D. Höttecke (Ed.), Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. GDCP Jahrestagung in Potsdam 2010. Münster, Westf: LIT-Verl.
- Messick, S. (1995). Validity of psychological assessment: Validation of interferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Rabe, T., Krey, O., & Meinhardt, C. (2013). Physikdidaktische Selbstwirksamkeitserwartungen zukünftiger Physiklehrkräfte I. In S. Bernholt (Ed.), Inquiry-based Learning - Forschendes Lernen. GDCP Jahrestagung in Hannover 2012 (pp. 635–637). Kiel: IPN.
- Schumacher, D., & Planinšič, G. (2007). Student undergraduate laboratory and project work. *European Journal of Physics*, 28(3).
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2005). Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. München: Luchterhand.
- Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (1999). Note on Sources of Sampling Variability in Science Performance Assessments. *Journal of Educational Measurement*, 36(1), 61–71.
- Thillmann, M. (2007). Selbstreguliertes Lernen durch Experimentieren: Von der Erfassung zur Förderung.