

Livia Murer<sup>1</sup>  
 Susanne Metzger<sup>2</sup>  
 Andreas Vorholzer<sup>3</sup>  
 Angela Bonetti<sup>1</sup>  
 Christoph Gut<sup>1</sup>

<sup>1</sup>Pädagogische Hochschule Zürich  
<sup>2</sup>Universität Basel  
<sup>3</sup>Technische Universität München

## **Kognitive Validierung von Aufgaben zum naturwissenschaftlichen Messen**

Der Aufbau von Kompetenzen zum naturwissenschaftlichen Messen ist ein relevantes Ziel des naturwissenschaftlichen Unterrichts (z. B. KMK; 2005; EDK, 2011) und Gegenstand naturwissenschaftsdidaktischer Forschung (z. B. Heinicke, 2012; Hellwig, 2012; Schulz, 2021). Mit der Relevanz dieser Kompetenzen geht ein Bedarf an Instrumenten einher, mit denen sich die Kompetenzen von Schüler/-innen im Bereich des Messens valide erfassen lassen. Zu diesem Zweck wurde im Rahmen des Projekts ExKoNawi u. a. ein hands-on Testverfahren zum Messen entwickelt, dessen kognitive Validität in der hier vorgestellten Studie geprüft wurde.

### **Theoretischer Hintergrund**

#### *Kognitive Validität von Aufgaben*

Ein zentrales Anliegen von Testaufgaben zum naturwissenschaftlichen Messen ist es, zu erfassen, inwiefern Schüler/-innen Konzepte zum Messen intuitiv verstanden haben und für kompetentes Handeln nutzen (vgl. Aufschnaiter & Hofmann; 2014; Vorholzer, Aufschnaiter & Kirschner, 2016). Damit eine Aufgabe valide Schlüsse über die zu erfassenden Kompetenzen zulässt, ist es erforderlich, dass die Konzepte, die in Aufgaben geprüft werden sollen (intendierte Konzepte), auch zum Lösen der Aufgabe verwendet werden. Dieser Aspekt wird als kognitive Validität bzw. ‚substantive aspect‘ (Messick, 1995, S. 745) bezeichnet. Hinweise auf kognitive Validität kann die Untersuchung folgender Fragen liefern (vgl. Baxter & Glaser, 1998):

- Inwiefern ist eine Aktivierung der intendierten Konzepte durch die Aufgabenstellungen naheliegend? (A)
- Inwiefern werden die intendierten Konzepte zum Lösen der Aufgaben verwendet? (B)
- Inwiefern gehen qualitativ hochwertigere Denkprozesse bei der Bearbeitung einer Aufgabe (bzgl. der intendierten Konzepte) mit einer besseren Lösung einher? (C)

#### *Konzepte zum naturwissenschaftlichen Messen*

Das naturwissenschaftliche Messen umfasst eine Reihe unterschiedlicher Fähigkeiten (z. B. Messungen durchführen, Messunsicherheiten abschätzen) und zugehörige Konzepte (siehe z. B. Heinicke, 2012; Hellwig, 2012). Den im Rahmen dieser Studie genutzten Aufgaben liegen folgende Konzepte zu Grunde:

- Messwiederholung und Mittelwertbildung erhöhen die Reliabilität einer Messung (z. B. Gott, Dugan, Roberts & Hussain, 2003; Heinicke, 2012; Kanari & Millar, 2004).
- Für eine präzise Messung braucht es ein möglichst genaues Messinstrument (z. B. Gott et al., 2003; Schulz, 2021).

- Mengenvergrößerung kann die Messgenauigkeit erhöhen (z. B. Suida & Grabowski, 2012). Die Strategie der Mengenvergrößerung kann angewandt werden, wenn die Messung mit nur einem Objekt / Ereignis sehr ungenau wird. Soll z. B. die Schwingungsdauer eines Pendels bestimmt werden, kann zur Erhöhung der Messgenauigkeit mit mehreren Schwingungen gemessen und dann auf die gesuchte Größe zurückgerechnet werden.

Im Rahmen vorliegender Studie wurde entlang der unter (A) bis (C) genannten Fragen untersucht, inwiefern die Ergebnisse der hands-on Aufgaben zum Messen kognitiv valide Schlüsse über das Verständnis von Schüler/-innen bzgl. der oben genannten Konzepte und damit über die Kompetenzen im Bereich des naturwissenschaftlichen Messens zulassen.

### Untersuchungsdesign

Die kognitive Validität wurde bei 6 hands-on Aufgaben zum Messen untersucht. In den Aufgaben sollen die Schüler/-innen eine gesuchte Größe mit gegebenen Messinstrumenten möglichst präzise messen, hierfür das für die Messung der gesuchten Größe genauere Messinstrument verwenden sowie Messwiederholungen durchführen und / oder mit einer Menge messen, um die Messgenauigkeit zu erhöhen. Die Aufgaben haben die Schüler/-innen in Einzelarbeit bearbeitet und dabei vorstrukturierte Protokolle ausgefüllt. Nach dem Lösen wurden sie in Einzelinterviews zu ihrem Vorgehen befragt, wobei die ausgefüllten Protokolle als Stimulus dienten. Insgesamt nahmen 27 Schüler/-innen des 8. Schuljahrs an der Studie teil (Alter:  $M = 14$  Jahre und 1 Monat,  $SD = 5$  Monate; 13 Mädchen und 14 Jungen). Sie wurden viermal besucht, wobei sie bei jedem Besuch eine andere Aufgabe zum Messen bearbeiteten. Alle Schüler/-innen haben somit 4 der 6 Aufgaben zum Messen gelöst, in Summe liegen 108 Schüler/-innen-Protokolle und Interviews vor.

Zur Untersuchung der kognitiven Validität wurden 8 Experten/-innen<sup>1</sup> einerseits die Aufgaben und andererseits aufbereitete Daten aus den Protokollen und Interviews vorgelegt:

- Zur Beantwortung von Frage (A) schätzten die Experten/-innen in einem Ratingbogen ein, inwiefern es aus ihrer Sicht naheliegend ist, dass Schüler/-innen beim Lösen der Aufgaben über die intendierten Konzepte (z. B. Messwiederholungen erhöhen die Messgenauigkeit) nachdenken. Von den Einschätzungen wurde ein Mittelwert gebildet.
- Um Frage (B) zu untersuchen, wurden zuerst die Interviews mit einem Kategoriensystem ausgewertet. Hierbei wurden aus den Äußerungen der Schüler/-innen induktiv kategorisierten Aussagen gebildet, in denen inhaltlich ähnliche Äußerungen gebündelt wurden. Zur Prüfung der Zuweisung der Äußerungen der Schüler/-innen zu den kategorisierten Aussagen wurden 30 % der Daten doppelt kodiert, mit zufriedenstellender Interrater-Übereinstimmung (Cohens Kappa  $\geq .70$ ; prozentuale Übereinstimmung  $\geq 78$  %). Die kategorisierten Aussagen wurden anschließend von den Experten/-innen im Hinblick darauf beurteilt, inwiefern sie ein Nachdenken über die intendierten Konzepte nahelegen. Die Experten/-innen wurden darauf hingewiesen, nur einzuschätzen, ob ein Bezug zu den intendierten Konzepten hergestellt wurde, aber nicht, ob die Konzepte richtig angewendet wurden.

---

<sup>1</sup> Professor/-innen und wissenschaftlichen Mitarbeiter/-innen der Naturwissenschaftsdidaktiken aus Deutschland und der Schweiz

- Um Frage (C) nachzugehen, wurden die Experten/-innen im Rahmen des zu Frage (B) beschriebenen Vorgehens außerdem befragt, ob die kategorisierten Aussagen einem eher hohen oder einem eher niedrigen Verständnisniveau des intendierten Konzepts entsprechen. Danach wurde der Zusammenhang dieser Einschätzungen und der Kompetenzdiagnose anhand der Schüler/-innen-Protokolle<sup>2</sup> betrachtet.

### **Ergebnisse** (vgl. Murer, 2022)

Mit Blick auf Frage (A) zeigen die Ergebnisse, dass die Experten/-innen bei den Aufgaben eine Aktivierung der intendierten Konzepte im Bereich Messwiederholung (Aufgaben: Ahorn:  $M = 3.5$ ,  $SD = 0.8$ ; Faden:  $M = 3.5$ ,  $SD = 0.8$ ; Pulver  $M = 3.1$ ,  $SD = 0.6$ )<sup>3</sup> oder Mengenvergrößerung (Aufgaben: Bohne:  $M = 3.9$ ,  $SD = 0.4$ ; Filzstift:  $M = 3.1$ ,  $SD = 0.6$ ; Münze:  $M = 3.8$ ,  $SD = 0.5$ ) für naheliegend halten. Da bei den Aufgaben jeweils nur eines der beiden Konzepte verlangt wurde, wurde dies als Hinweis für kognitive Validität gedeutet.

Bezüglich Frage (B) wurde für alle 6 Aufgaben untersucht, inwiefern es in den Äußerungen der Schüler/-innen Hinweise auf die Aktivierung der intendierten Konzepte *Messwiederholung*, *Mengenvergrößerung* und *Wahl des Messinstruments* gibt. Bei 13 der sich daraus ergebenden 18 Kombinationen hat gemäß Einschätzungen der Experten/-innen die Mehrheit der Schüler/-innen (67 bis 100 %) über das intendierte Konzept beim Lösen der Aufgabe nachgedacht. Bei 4 Kombinationen wurde bei der Mehrheit der Schüler/-innen (67 bis 87 %) nicht ersichtlich, worüber sie beim Lösen der Aufgabe nachdenken. Dieser Befund liefert zwar keine Hinweise für kognitive Validität, aber auch keine Hinweise gegen sie und wird somit nicht als Einschränkung der Validität beurteilt. Da es zudem insgesamt eher wenige Hinweise auf die Nutzung nicht intendierter Konzepte gab<sup>4</sup>, wurde dies insgesamt als Hinweis für kognitive Validität gedeutet.

Mit Blick auf die Qualität der Lösung (Frage (C)) konnte ein positiver Zusammenhang zwischen dem Ergebnis der Kompetenzdiagnose anhand der Schüler/-innen-Protokolle und der Experten/-innen-Einschätzung bzgl. der Qualität bei den kategorisierten Aussagen festgestellt werden ( $r = .46$ ;  $p \leq .001$ ;  $N = 108$ ). Somit kann das Ergebnis der Diagnose anhand der Schüler/-innen-Protokolle die Qualität der Denkprozesse bei den intendierten Konzepten abbilden. Dies wurde als weiterer Hinweis für kognitive Validität gedeutet.

### **Fazit**

Die Ergebnisse deuten in Summe darauf hin, dass das ExKoNawi-Testinstrument kognitiv valide Schlüsse über die experimentellen Kompetenzen von Schüler/-innen des 8. Schuljahrs im Bereich des naturwissenschaftlichen Messens zulässt. Gleichzeitig ist jedoch festzuhalten, dass die Betrachtung eines einzelnen Validitätsaspekts in der Regel nur einer von mehreren Bausteinen der Validierung ist und es somit weitere Aspekte zu prüfen gilt (siehe z. B. Bonetti et al. in diesem Band).

<sup>2</sup> Die Schüler/-innen-Protokolle wurden im Rahmen der Gesamtstudie von ExKoNawi mit Kodiermanualen ausgewertet (Doppelkodierung: 15 % der Daten; Cohens Kappa  $\geq .61$ ; prozentuale Übereinstimmung  $\geq 81$  %).

<sup>3</sup> Skala: 1: nicht naheliegend; 2: eher nicht naheliegend; 3: eher naheliegend; 4: naheliegend.

<sup>4</sup> Bei 17 von 18 Kombinationen haben lediglich 0 bis 33 % der Schüler/-innen beim Lösen der Aufgabe über ein nicht intendiertes Konzept nachgedacht. Nur bei einer Kombination hat die Mehrheit der Schüler/-innen (67 %) über ein nicht intendiertes Konzept nachgedacht. Im Bereich dieser Kombination lässt die Aufgabe folglich keine kognitiv validen Schlüsse bzgl. der Kompetenzen der Schüler/-innen im untersuchten Bereich zu.

**Literatur**

- Aufschnaiter, C. v. & Hofmann, J. (2014). Kompetenz und Wissen – Wechselseitige Zusammenhänge und Konsequenzen für die Unterrichtsplanung. *Der mathematische und naturwissenschaftliche Unterricht (MNU)*, 67 (1), 10-16.
- Baxter, G. P., & Glaser, R. (1998). Investigating the Cognitive Complexity of Science Assessment. *Educational Measurement: Issues and Practice*, 17 (3), 37–45.
- Erziehungsdirektorenkonferenz (EDK) (Hrsg.). (2011). Grundkompetenzen für die Naturwissenschaften - Nationale Bildungsstandards. Zugriff: 01.03.2022. Verfügbar unter: <https://www.edk.ch/de/themen/harmos/nationale-bildungsziele>
- Gott, R., Dugan, S., Roberts, R., & Hussain, A. (2003). Research into understanding scientific evidence. Zugriff 10.02.22, unter: <http://community.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm>
- Heinicke, S. (2012). Aus Fehlern wird man klug - Eine Genetisch-Didaktische Rekonstruktion des «Messfehlers». Berlin: Logos Verlag.
- Hellwig, J. (2012). Messunsicherheiten verstehen: Entwicklung eines normativen Sachstrukturmodells am Beispiel des Unterrichtsfaches Physik. Dissertation, Ruhr-Universität Bochum.
- Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, 41 (7), 748–769.
- 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.
- Murer, L. (2022). Diagnose experimenteller Kompetenzen beim praktisch-naturwissenschaftlichen Arbeiten: Vergleich verschiedener Methoden und kognitive Validierung eines Testverfahrens. Dissertation, Institut für Bildungswissenschaften der Universität Basel.
- Schulz, J. (2021). Entwicklung eines Testinstrumentes zur Erfassung von Kompetenzen im Umgang mit Messunsicherheiten. Dissertation, Humboldt-Universität zu Berlin.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). (2005). Bildungsstandards im Fach Physik für den Mittleren Schulabschluss (Jahrgangsstufe 10). München: Luchterhand
- Suida, R., & Grabowski, A. (2012). Combined measurements – A way to improve the measurement accuracy of an additive quantity. *Measurement*, 45, 1165–1169.
- Vorholzer, A., Aufschnaiter, C. v., & Kirschner, S. (2016). Entwicklung und Erprobung eines Tests zur Erfassung des Verständnisses experimenteller Denk- und Arbeitsweisen. *Zeitschrift für Didaktik der Naturwissenschaften*, 22, 25–41.