

Angela Bonetti<sup>1</sup>  
 Susanne Metzger<sup>1</sup>  
 Christoph Gut<sup>1</sup>

<sup>1</sup>Pädagogische Hochschule Zürich

## **Validierung des ExKoNawi-Modells (Experimentelle Kompetenzen in den Naturwissenschaften)**

### **Überblick**

Im Rahmen einer SNF-finanzierten Studie wird ein reduziertes *ExKoNawi*-Modell (*Experimentelle Kompetenzen in den Naturwissenschaften*, siehe Beitrag von Metzger und Gut, in diesem Band) validiert. Hierbei werden auf der Grundlage von drei Pilotstudien (siehe auch Beitrag von Gut et al., in diesem Band) die drei Problemtypen „skalenbasiertes Messen“, „effektbasiertes Vergleichen“ und „fragengeleitetes Untersuchen“ für den Haupttest zusammengefasst und ab Herbst 2016 untersucht. 24 Klassen der Sekundarschule A, B und C und des Langzeitgymnasiums werden hierfür im Fach Natur und Technik (Sekundarschulen) oder in einem der Disziplinen Biologie, Chemie und Physik (Langzeitgymnasium) besucht. Ziel der Studie ist die Untersuchung der internen und der externen Validität des ExKoNawi-Modells (Messick, 1996; Leuders, 2014).

### **Die Items: Auswahl und Konstruktion**

Die Auswahl für den Haupttest umfasst pro Problemtyp sechs Experimentieraufgaben (siehe Tab. 1), wobei die Fachkontexte über die drei naturwissenschaftlichen Disziplinen äquivalent verteilt sind. Jeder Schüler / jede Schülerin wird neben 4 Begleittests (KFT, Sprach- und Strategietest (NAW), Fragebogen zu Motivation und Selbstwirksamkeit) 12 Experimentieraufgaben bearbeiten. Hierbei wird das Protokoll der Schülerin / des Schülers analysiert, welches in vordruckten Testheften festgehalten wird. Auf die Aufgaben zum Problemtyp „kategoriegeleitetes Beobachten“ wurde aufgrund messtechnischer Mängel verzichtet.

<b>Problemtypen</b>	skalenbasiertes <b>Messen</b>	quantitative Grössen mit gegebenen Messinstrumenten (Skala) genau messen <b>6 Experimentieraufgaben</b>	<b>Haupttest (2016-2018)</b>
	fragengeleitetes <b>Untersuchen</b>	korrelative Zusammenhänge zwischen gegebenen Variablen (Frage) untersuchen <b>6 Experimentieraufgaben</b>	
	effektbasiertes <b>Vergleichen</b>	Objekte anhand einer gegebenen Eigenschaft (ohne direkte Messung) experimentell vergleichen <b>6 Experimentieraufgaben</b>	

*Tab. 1: Strukturmodell und Übersicht über die eingesetzten Experimentieraufgaben im Haupttest zu ExKoNawi. Einteilung nach Lubben & Millar (1996), RuizPrimo & Shavelson (1996), Hammann et al. (2008), SolanoFlores (1994)*

Bei der Itemkonstruktion über alle Experimentieraufgaben wurde ein Minimum an technischen Begriffen und möglichst eine alltägliche Sprache bzw. Kontext verwendet. Außerdem steht am Anfang jedes Items eine kurze Erklärung des dahintersteckenden Phänomens, sodass die Aufgabe ohne nötiges Vorwissen gelöst werden kann. Als letztes werden über alle Problemtypen hinweg hochstandardisierte Aufgabeformate angewandt. Jedem Problemtypen liegt somit ein Itemkonstruktionsschema zugrunde, mittels welchem

die Prozessqualität in partial credits erhoben werden kann; es können pro Problemtyp bis zu vier Qualitätsstandards dichotom bewertet werden. Dieses Itemkonstruktionsschema spiegelt sich in den Testheften, in welchen die Schülerinnen und Schüler aufgefordert werden die gewonnenen Resultate und Beobachtungen bei den Messungen, Untersuchungen oder Vergleiche in Form eines Eigenrapports festzuhalten und genau zu erklären, wie sie dabei vorgegangen sind.

### Geplantes Testdesign

Die 18 ausgewählten Experimentieraufgaben werden auf 6 Units eingeteilt, wobei pro Unit jeweils eine Mess-, eine Vergleichs- und eine Untersuchungsaufgabe enthalten sind. Außerdem beinhaltet jede Unit eine Aufgabe zu einem chemischen, biologischen und physikalischen Kontext, wobei auch die Disziplinen pro Problemtyp gleich oft auftreten und durchmischt werden (Abb. 1 unten rechts).

An jedem von vier Testtagen werden eine Unit und ein zusätzlicher Test (Fragebogen) erhoben. Beispielsweise wird an Tag 2 neben einer Experimentieraufgabe zu jedem der drei Problemtypen auch noch der KFT figurale Test (Heller & Perleth, 2000) durchgeführt (Abb. 1 linke Seite). Dieses Testdesign folgt dem Youden Square Design (Frey, Hartig & Rupp, 2009) und ermöglicht, dass von den 450 Schülerinnen und Schülern jede Unit (und somit jede Experimentieraufgabe) von mindestens 300 Schülerinnen und Schülern bearbeitet wird. Außerdem wird jede Unit-Kombination von mindestens 150 Schülerinnen und Schülern gelöst.

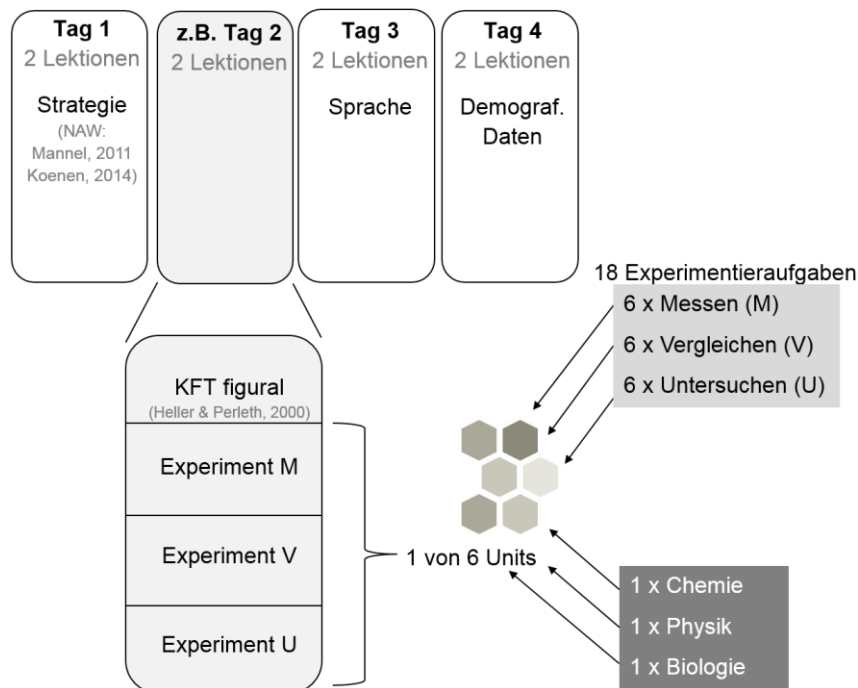


Abb. 1: Geplantes Testdesign am Beispiel von Tag 2.

Durch diese Testanordnung kann eine strukturelle Dimensionsanalyse durchgeführt werden. Außerdem werden weitere strukturelle Einflüsse wie Positionseffekte oder Klasseneffekte durch das Design ausbalanciert.

### Forschungsfragen

Mit dem beschriebenen Projektdesign wird die strukturelle Validität des ExKoNawi-Modells untersucht (Abb. 2 oberer Pfeil & graue Farbe). Hierbei wird verglichen, ob eine übergeordnete Experimentierkompetenz (links) oder die drei untersuchten Problemtypen (rechts) als unabhängige Teilkompetenzen die Resultate besser erklären können. Geplant ist ein Rasch-analytischer dimensionaler Modellvergleich. Als weitere Forschungsfrage wird untersucht, inwiefern die miterhobenen Größen des naturwissenschaftlichen Strategiewissens, der kognitiven und der sprachlichen Fähigkeiten mit den erhobenen Daten zum Problemlösen korrelieren. Hierbei wird die externe Validität des ExKoNawi-Modells mittels Hintergrundmodell und DIF-Analysen untersucht (Abb. 2 linker Pfeil & schwarze Farbe). Zusätzlich folgen weitere Datenanalysen, wie z. B. der Vergleich über die Schulniveaus (Sek A, B, C bzw. Gymnasium).

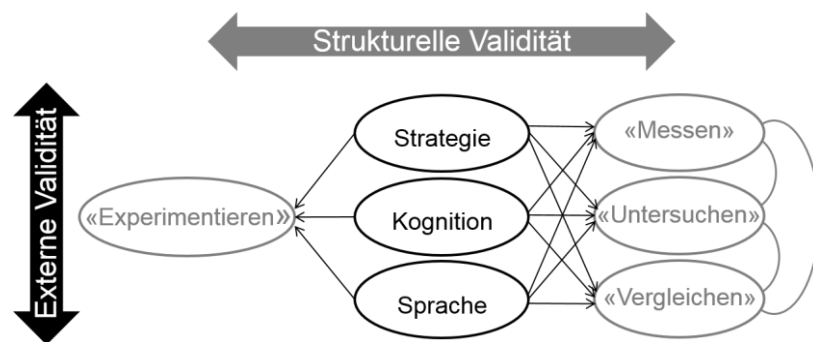


Abb. 2: Forschungsfragen: Überprüfung der strukturellen und der externen Validität

**Literatur**

- Frey, A., Hartig, J. & Rupp A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: theory and practice. *Educational Measurement: Issues and Practice*, 28(3), 39-53
- Hammann, M., Phan, T. T. H., Ehmer, M. & Grimm, T. (2008). Assessing pupil's skills in experimentation. *Journal of Biological Education*, 42(2), 66-72
- Heller, K. A. & Perleth, C. (2000). KFT 4-12+ R. Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision. Göttingen: Beltz Test
- Koenen, J. (2014). Entwicklung und Evaluation von experimentunterstützten Lösungsbeispielen zur Förderung naturwissenschaftlich-experimenteller Arbeitsweisen. Berlin: Logos
- Leuders, T. (2014). Modellierungen mathematischer Kompetenzen - Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. *Journal für Mathematikdidaktik*
- Lubben, F. & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education*, 18(8), 955-968
- Mannel, S. (2011). Assessing scientific inquiry. Development and evaluation of a test for the low-performing stage. Berlin: Logos
- Messick, S. (1996). Validity of Performance Assessments. In G.W. Philips (Ed.), *Technical Issues in Large-Scale Performance Assessment*. Washington DC: NCES
- Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessments: an update. *Journal of Research in Science Teaching*, 33(10), 1045-1063
- Solano-Flores, G. (1994). A logical model for the development of science performance assessments. University of California, Santa Barbara