

Angela Bonetti^{1,3}
 Christoph Gut¹
 Susanne Metzger²
 Maik Walpuski³

¹PH Zürich

²PH FHNW

³Universität Duisburg-Essen

Performanz beim Experimentieren mit und ohne Experimentiermaterial

Einleitung & Übersicht

Das praktische Experimentieren mit Experimentiermaterial im Sinne der Untersuchung kausaler Zusammenhänge stellt an Schülerinnen und Schüler Anforderungen, die über das reine Wissen der dazu notwendigen Experimentierstrategien, wie beispielsweise die Variablenkontrollstrategie beim Planen von Versuchsanordnungen, hinausgehen. U. a. erweist sich der Aufgabenkontext als ein Faktor, der individuell die Performanz beim praktischen Experimentieren wesentlich beeinflusst (z. B. Shavelson et al., 1999). Dieses Ergebnis wirft die Fragen auf, inwiefern das praktische Experimentieren eine eigenständige Kompetenz ist, bzw. welche Wissensarten die Performanz beim praktischen Experimentieren bestimmen.

Im Projekt „ExKoNawi“ (Experimentelle Kompetenzen in den Naturwissenschaften) werden die experimentellen Kompetenzen nach dem Problemtypen-Ansatz modelliert (Gut et al., 2014). Ziel der laufenden Studie ist es unter anderem, das entstandene Messinstrument über drei Problemtypen extern zu validieren (Messick, 1996; Leuders, 2014). In der folgenden Teilauswertung werden nur die Daten aus den Tests mit dem Problemtyp Untersuchen, der die Variablenkontrollstrategie mit beinhaltet, dargelegt. (Für einen ausführlicheren Beschrieb der gesamten Studie siehe Bonetti et al., 2017.) Im Rahmen des Projekts bearbeiteten Schülerinnen und Schüler (aller Leistungsniveaus) jeweils vier von insgesamt sechs Hands-on Aufgaben zum „Untersuchen von Zusammenhängen“ aus nicht curricularen biologischen, chemischen und physikalischen Kontexten. Neben weiteren Begleitskalen wurden kognitive Fähigkeiten, die Lesegeschwindigkeit und das Strategiewissen der naturwissenschaftlichen Arbeitsweisen erhoben. Um obenstehende Fragen zu beantworten, werden erste ausgewählte Resultate der Studie präsentiert und diskutiert.

Design & Stichprobenziehung

Während vierer Testzeitpunkte (TZP) lösten rund 470 Zürcher Schülerinnen und Schüler der 8. Jahrgangsstufe in Einzelarbeit insgesamt 12 experimentelle Hands-on Aufgaben (zu jedem TZP drei Aufgaben, jeweils zu einer Unit zusammengefasst). Da zu jedem TZP auch eine Untersuchungsaufgabe bearbeitet wurde, bearbeitete jede/r Schüler/in vier von sechs Untersuchungsaufgaben, welche in einem Multi Matrix Design verteilt wurden. Zusätzlich wurden, neben weiteren hier nicht weiter ausgeführten Fragebögen, zu jedem TZP externe Faktoren miterhoben. Zum ersten TZP wurde für das Strategiewissen der naturwissenschaftlichen Arbeitsweisen der NAW-Test (Koenen, 2014; Mannel, 2011) eingesetzt, zum dritten TZP für die kognitiven Fähigkeiten die Subskalen N1 und Q2 des KFTs (Heller & Perleth, 2000) und zum vierten TZP die Subskala V1 zusammen mit dem Salzburger Lesescreening (SLS) (Mayringer & Wimmer, 2014) für die Messung der sprachlichen Fähigkeiten (Lesegeschwindigkeit). Beim NAW-Test wurde für jedes Leistungsniveau ein eigenes Testheft verwendet, wobei insgesamt 25 Items unterschiedlich zu drei Testheften à 19 (Sek B/C), 20 (Sek A), respektive 21 Items (Gymnasium) zusammengestellt wurden.

Datenaufbereitung, Datenanalyse & Resultate

Untersuchungsaufgaben: Als Basis für die Kodierung der Protokolle dienten die Manuale der vorausgegangenen Pilotstudien. Pro Aufgabe konnten maximal 4 Punkte erreicht werden, welche mittels 16 Indikatoren erhoben wurden. An der Kodierung waren vier Personen beteiligt, welche nach einer Trainingsphase 15% der Testhefte doppelt kodiert haben. Dabei wurden folgende Kriterien für die Objektivität erreicht: $\kappa \geq 0.61$ oder Gwet's AC1 ≥ 0.81 bei PÜ $\geq 85\%$. Die Daten wurden anschließend von einer Person in der Masterspur kodiert. Die erhaltenen Daten der experimentellen Aufgaben (*partial credit*) wurden mittels Rasch-Skalierung (ConQuest) auf die Reliabilität geprüft. Für eine Übersicht der erhaltenen Werte und die DIF-Analysen zum Geschlecht und zum Leistungsniveau siehe Tab. 1a. Es zeigt sich anhand der Reliabilität, dass mit den 6 Untersuchungs-Aufgaben eine experimentelle Kompetenz gemessen werden kann. Dabei weisen die Items gute Infit- / Outfit-Werte auf. Die DIF-Analyse zum Geschlecht ist hochsignifikant, wobei die Mädchen besser abschneiden als die Jungen. Jedoch besteht kein DIF-Effekt auf der Ebene der Items. Auch bei der DIF-Analyse zur Leistungsstufe besteht ein hochsignifikanter Leistungsunterschied, wobei die Schülerinnen und Schüler in der folgenden Reihenfolge von stärker zu schwächer abschneiden: Gymnasium, Sek A, Sek B/C. Auch hier zeigt sich kein DIF-Effekt auf der Ebene der Items.

Externe Faktoren: Die verwendeten Begleittests SLS, NAW und KFT sind Multiple-Choice Formate, womit sich die Objektivitätsüberprüfung erübrigt. Die mittels Rasch-Analyse (ConQuest) erreichten Reliabilitätsmasse des NAW-Tests sind in Tab. 1b zusammengefasst. Dieser wurde aufgrund eines Infit-Werts um eine Aufgabe gekürzt. Die Reliabilität der drei verwendeten Subskalen des KFTs liegt bei einem Cronbach's α von 0.733 (SPSS).

Tab. 1a: Statistische Prüfung der sechs Untersuchungsaufgaben (ConQuest)

Untersuchungsaufgaben (6 Items) <i>partial credit</i>			
EAP/PV-Reliabilität	0.637	Varianz	0.407
Infit / Outfit	0.94 – 1.07	T-Werte	< 1
DIF-Analyse	Geschlecht**	Item*Geschlecht	< 0.068
Jungen	-0.195 \pm 0.023		
Mädchen	0.195 \pm 0.023		
Infit / Outfit	0.87 – 1.2	T-Werte	< 1.5
DIF-Analyse	Leistungsstufe**	Item*Stufe	< 0.143
Gymnasium	0.543 \pm 0.032		
Sek A	0.072 \pm 0.026		
Sek B/C	-0.614 \pm 0.042		
Infit / Outfit	0.80 – 1.19	T-Werte	< 1.8

Tab. 1b: Statistische Prüfung der NAW-Aufgaben (ConQuest)

NAW-Aufgaben (24 Items) <i>dichotom</i>			
EAP/PV Reliabilität	0.698	Varianz	1.177
Infit	0.82 – 1.15	T-Werte	< 2 (excl. 2 Items \leq 2.9)
Outfit	0.64 – 1.56	T-Werte	\leq 3.8 (excl. 2 Items \leq 6.2)

Regressionsanalyse: Um den Einfluss der externen Faktoren auf die Performanz der Übungsaufgaben zu prüfen, wurde eine Regressionsanalyse durchgeführt (siehe Abb. 1). Darin zeigt sich ein hochsignifikanter, mittlerer Einfluss des NAW-Tests und ein hochsignifikanter, eher kleiner Einfluss des KFTs auf die Performanz des Untersuchens. Es

kann kein Einfluss des SLS auf die experimentelle Kompetenz (Untersuchen von Zusammenhängen) gemessen werden.

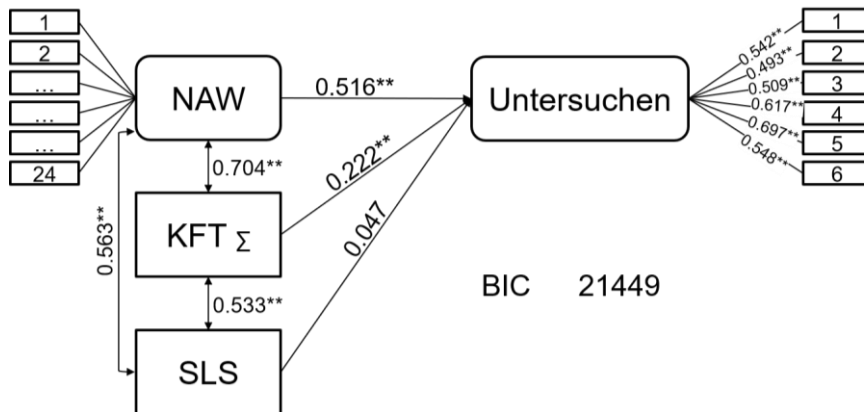


Abb. 1: Regressionsanalyse (Mplus) als Modell dargestellt, wobei ** = $p < 0.01$.

Diskussion

Untersuchen: Die Resultate zeigen, dass die Bearbeitung von 4 von insgesamt 6 Hands-on Aufgaben ausreicht, um eine experimentelle Kompetenz (Untersuchen von Zusammenhängen) mit mäßiger Reliabilität (0.63) messen zu können.

Naturwissenschaftliche Arbeitsweisen: Zwischen dem Strategiewissen zu den naturwissenschaftlichen Arbeitsweisen und dem praktischen Untersuchen kann ein hoch signifikanter mittlerer Zusammenhang (0.516**) gemessen werden. Dieser kann damit erklärt werden, dass beide Tests ein ähnliches Konstrukt messen und das Verständnis der Variablenkontrollstrategie voraussetzen. Das mit einem Paper-and-Pencil gemessene „Wissen wie“ (NAW) als Disposition erklärt somit rund 25 % der Varianz des in einem Hands-on Setting gemessenen „Könnens“ (Untersuchen) als Performanz.

Kognitive Fähigkeiten: In den Pilotstudien zum Projekt ExKoNawi wurden jeweils hoch signifikante und starke Leistungsunterschiede beim Experimentieren zwischen den Leistungsstufen gefunden. Da die Aufgabeninhalte wenig curricular sind, wird deshalb angenommen, dass mit den Experimentieraufgaben von ExKoNawi zu einem gewissen Teil allgemeine kognitive Fähigkeiten gemessen werden. Diese Vermutung wird zwar durch die Literatur gestützt, wonach ein mittlerer Einfluss der kognitiven Fähigkeiten (KFT) auf die Performanz bei experimentellen Paper-and-Pencil-Aufgaben besteht (z. B. Wellnitz, 2012; Leutner et al., 2004) wird. Jedoch wird die Vermutung durch die vorgestellte Studie nicht bestätigt: Die kognitiven Fähigkeiten machen nur eine Varianzaufklärung von knapp 5 % aus.

Sprache: Das Testformat von ExKoNawi stellt an die Schülerinnen und Schülern Lese- und Schreibanforderungen, da mit schriftlichen Testheften und Schülerprotokollen gearbeitet wird (vgl. Stecher et al., 2000; Gut, 2012). Trotz dieser sprachlichen Anforderungen zeigt die Lesefähigkeit, gemessen mit dem SLS-Test, keinen signifikanten Einfluss auf die experimentelle Kompetenz.

Zusammenfassend ergeben die Resultate eine positive Validierung des experimentellen Tests. Mit einem Test von sechs Items, wovon jeweils vier bearbeitet wurden, kann eine experimentelle Kompetenz mit mäßiger Reliabilität gemessen werden, die wie erwartet stark vom Wissen der naturwissenschaftlichen Arbeitsweisen und in kleinerem Mass von den allgemeinen kognitiven Fähigkeiten, jedoch nicht von der Lesefähigkeit abhängt.

Literatur

- Bonetti, A., Gut, C. & Metzger S. (2017): *Validierung des ExKoNawi-Modells* (Experimentelle Kompetenzen in den Naturwissenschaften). In C. Maurer, *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis. GDCP Jahrestagung 2016 Zürich* (S. 336-339). Regensburg: Universität von Regensburg. http://www.gdcp.de/images/tagungsbaende/GDCP_Band37.pdf
- Gut, C. (2012). *Modellierung und Messung experimenteller Kompetenz: Analyse eines large-scale Experimentiertests*. Berlin: Logos.
- Gut, C., Metzger, S., Hild, P., & Tardent, J. (2014). *Problemtypenbasierte Modellierung und Messung experimenteller Kompetenzen von 12- bis 15-jährigen Jugendlichen*. PhyDid B, Didaktik Der Physik, Beiträge Zur DPG-Frühjahrstagung 2014, 9 S.
- 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 Mathematik-Didaktik, 35(1), 7-48.
- Leutner D., Klieme E., Meyer K. & Wirth J. (2004). *Problemlösen*. In: M. Prenzel et al., *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147-175). Münster: Waxmann.
- Mannel S. (2011). *Assessing scientific inquiry – Development and evaluation of a test for the low-performing stage*. Berlin: Logos.
- Mayringer, H. & Wimmer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2-9*. Bern: Hogrefe Verlag.
- Messick, S. (1996). *Validity of Performance Assessments*. In G.W. Philips (Ed.), *Technical Issues in Large-Scale Performance Assessment*. NCES: Washington DC.
- Shavelson, R. J., Ruiz-Primo, A. M., & Wiley, E. W. (1999). *Note on sources of sampling variability in science performance assessments*. Journal of Educational Measurement, 36(1), 61-71.
- Stecher, B. M., Klein, S. P., Solano-Flores, G., McCaffrey, D., Robyn, A., Shavelson, R. J. & Haertel, E. (2000). *The effects of content, format, and inquiry level on science performance assessment scores*. Applied Measurement in Education, 13(2), 139-160.
- Wellnitz N. (2012). *Kompetenzstruktur und -niveaus von Methoden naturwissenschaftlicher Erkenntnisgewinnung*. Biologie lernen und lehren. Band 2. Berlin: Logos.