

Robin GÖLLER, Lüneburg, Larissa ALTENBURGER, Lüneburg,  
Natalie TROPPER, Kassel, Maike HAGENA, Lüneburg &  
Michael BESSER, Lüneburg

## **Zur Validität eines Mathematiktests für die Auswahl von Bewerber\*innen auf ein Lehramtsstudium**

Aufgrund einer steigenden Anzahl von Studienbewerber\*innen, die oft den Umfang freier Studienplätze übersteigt, stehen Universitäten vor der Herausforderung, unter den Bewerber\*innen die „geeignetsten“ auszuwählen. Diese Auswahl darf nach einem Urteil des Bundesverfassungsgerichts aus dem Dezember 2017 nicht allein auf Grundlage der Hochschulzugangsberechtigungsnote getroffen werden. An der Leuphana Universität Lüneburg haben Studienbewerber\*innen auf ein das Fach Mathematik beinhaltendes Lehramtsstudium daher die Möglichkeit, u. a. durch das Absolvieren eines Mathematiktests Zusatzpunkte im Auswahlprozess zu sammeln und hierdurch ihre Chancen auf einen Studienplatz zu erhöhen.

Der dabei eingesetzte Mathematiktest besteht aus 77 Items, die über alle inhaltlichen und prozessbezogenen Kompetenzen der Sekundarstufe I des Modells der Bildungsstandards streuen, wurde an 654 Studierenden pilotiert und weist eine zufriedenstellende interne Konsistenz auf (siehe Besser et al., eingereicht). Die Inhaltsvalidität des Tests wurde durch Experten, die in den Normierungsprozess der Bildungsstandards involviert waren, bestätigt.

Mit Blick auf die Erfassung mathematischen Wissens und Könnens gemäß dem theoretischen Modell der Bildungsstandards sollte zunächst gelten, dass Schüler\*innen der Oberstufe (mit längerer Schulbesuchsdauer) in diesem Test besser abschneiden als Schüler\*innen der Mittelstufe (mit entsprechend kürzerer Schulbesuchsdauer). Hierüber hinaus gilt es, die Konstrukt- bzw. Kriteriumsvalidität des Tests explizit zu prüfen. Schüler\*innenleistungen in existierenden Mathematiktests, die das Modell der Bildungsstandards abbilden, korrelieren in der Regel negativ mit der Mathematiknote (wenn bessere Leistungen mit höheren Testpunkten und einer niedrigeren Note kodiert sind, z. B. Graf et al., 2016), positiv mit dem Selbstkonzept und der Selbstwirksamkeitserwartung in Bezug auf Mathematik, nicht oder leicht positiv mit Mathematikinteresse und negativ mit Mathematikangst (z. B. Lee & Stankov, 2018). Die forschungsleitenden Fragen des vorliegenden Beitrags lauten also wie folgt:

- Frage 1: Wie schneiden Schüler\*innen der Sekundarstufe I im Vergleich zu Studienbewerber\*innen bei einem Mathematiktest zu Inhalten der Sekundarstufe I ab?

- Frage 2: In welchem Zusammenhang steht die Testleistung mit der Mathematiknote, dem Selbstkonzept, der Selbstwirksamkeitserwartung, dem Interesse und der Ängstlichkeit in Bezug auf Mathematik?

Insbesondere sollen dazu folgende Hypothesen getestet werden:

- H1: Studienbewerber\*innen schneiden besser ab als Schüler\*innen der Sekundarstufe I.
- H2: Die Testleistung ist negativ mit der Mathematiknote, positiv mit dem mathematikbezogenen Selbstkonzept und der mathematikbezogenen Selbstwirksamkeitserwartung, schwach positiv oder nicht mit dem Mathematikinteresse und negativ mit der Mathematikangst korreliert.

## **Methode**

Der Mathematiktest wurde im Spätsommer 2019 im Multimatrixdesign von 140 Studienbewerber\*innen und 294 Schüler\*innen aus 7., 9. und 11. Klassen von Gymnasien und Gesamtschulen bearbeitet (für eine genauere Auflistung siehe Tabelle 1). Für die Tests in den Klassenstufen 7 und 9 wurden nur solche Aufgaben ausgewählt, die laut Lehrplan und Kerncurriculum von Schüler\*innen dieser Jahrgangsstufen bearbeitet werden konnten. Alle Aufgaben sind (Complex-)Multiple-Choice-Aufgaben. Die Antworten wurden dichotom kodiert (0 = nicht korrekt, 1 = korrekt) und anschließend mithilfe von ConQuest eindimensional Rasch-skaliert. Da die Aufgaben bisher noch nicht an Schüler\*innen erprobt waren, wurden die Personenparameter für frei geschätzte Itemschwierigkeiten berechnet. Für diese wurde mit Blick auf Frage 1 Varianzanalysen in SPSS mit sechs Faktoren, die sich aus den drei Klassenstufen und den zwei verschiedenen Schulformen ergeben, durchgeführt. Da sich die Fallzahlen dieser Faktoren deutlich unterschieden, wurde zudem als Post-Hoc-Test der GT2 nach Hochberg zur Bestimmung homogener Gruppen gerechnet.

Zur Beantwortung von Frage 2 wurden beim Einsatz der Tests in den Schulen die Mathematiknote (1 = sehr gut, 6 = ungenügend) sowie das Selbstkonzept in Mathematik (4 Items, Weber & Freund, 2017), die Selbstwirksamkeitserwartung Mathematik (4 Items, Ramm et al., 2006), Interesse und Freude (4 Items, Ramm et al., 2006) und Ängstlichkeit in Mathematik (5 Items, Ramm et al., 2006) mit papierbasierten Fragebögen anhand vierstufiger Likert-Skalen (1 = trifft nicht zu, 4 = trifft zu) erhoben.

## **Ergebnisse**

Eine Übersicht über die deskriptiven Testwerte ist in Tabelle 1 gegeben. Die Testskalierung liefert eine EAP-Reliabilität von 0.709. Die Varianzanalyse

zeigt, dass die Testleistungen der verschiedenen Gruppen nicht alle gleich sind. Der Post-Hoc-Test liefert drei homogene Gruppen, von denen die erste die 7. Klassen und die 9. Klassen der Gesamtschulen, die zweite die 9. und 11. Klassen der Gymnasien und die dritte die Studienbewerber\*innen umfasst (vgl. Tabelle 1). Dabei wurden von Schüler\*innen aus Gruppe 1 im Mittel 20 % der Aufgaben korrekt bearbeitet, von Schüler\*innen aus Gruppe 2 im Mittel 38 % und von den Studienbewerber\*innen im Mittel 48 %.

	Weighted Likelihood Estimates				
	N	M	SD	Min	Max
Gesamt	434	-0.005	1.192	-3.220	3.762
Gruppe 1	139	-0.100	0.936	-3.220	1.293
7. Klasse Gesamtschule	60	-1.073	0.820	-2.979	0.683
7. Klasse Gymnasium	26	-0.616	0.985	-2.979	1.063
9. Klasse Gesamtschule	53	-1.106	1.003	-3.220	1.293
Gruppe 2	155	0.141	0.877	-1.900	2.729
9. Klasse Gymnasium	23	0.234	1.118	-1.900	2.214
11. Klasse Gymnasium	132	0.124	0.832	-1.463	2.729
Gruppe 3 Studienbewerber*innen	140	0.822	1.003	-1.724	3.762

Tab. 1: Übersicht über die Rasch-skalierten Testergebnisse

Die internen Konsistenzen der erhobenen Skalen liegen alle im guten Bereich (Crombachs  $\alpha$  zwischen .80 und .91). Tabelle 2 zeigt die Korrelationen der Testergebnisse mit der Mathematiknote und dem Selbstkonzept, der Ängstlichkeit, dem Interesse und der Selbstwirksamkeitserwartung in Bezug auf Mathematik, sowohl für die Gesamtheit der getesteten Schüler\*innen als auch für die beiden Niveaugruppen.

	Mathenote	MSK	SWE	Interesse	Angst
Alle SuS	-.362**	.239**	.206**	n. s.	-.235**
Gruppe 1	-.266**	.216*	.178*	n. s.	-.235**
Gruppe 2	-.418**	.575**	.527**	.489**	-.490**

Tab. 2: Korrelationen der Testleistung mit Mathematiknote, mathematikbezogenem Selbstkonzept, mathematikbezogener Selbstwirksamkeitserwartung, Mathematikängstlichkeit und Mathematikinteresse. \*\*  $p < 0.01$ , \*  $p < 0.05$  (2-seitig).

## Diskussion

Die Ergebnisse zeigen, dass die Studienbewerber\*innen besser abschneiden als die Schüler\*innen der 9. und 11. Gymnasialklassen und diese wiederum besser als die Schüler\*innen der 7. Klassen und der 9. Gesamtschulklassen. Dies bestätigt Hypothese 1. Die Korrelationen in Tabelle 2 bestätigen Hypothese 2, insbesondere für Gruppe 2. Bemerkenswert ist hier die deutliche Korrelation des Interesses mit der Testleistung für Gruppe 2, die höher als in der einleitend erwähnten Literatur ausfällt. Die niedrigeren Korrelationen für Gruppe 1 lassen sich möglicherweise durch die geringeren Lösungshäufigkeiten in dieser Gruppe erklären, aufgrund derer einzelne richtig geratene Aufgaben das Testergebnis deutlich beeinflussen können.

Offensichtliche Limitationen sind die kleinen Gruppengrößen (insbesondere der getesteten 7. und 9. Gymnasialschüler\*innen) und mögliche Selektionseffekte, die sich aus den verschiedenen Schulformen oder der Bewerbung auf einen zulassungsbeschränkten Studiengang ergeben. Zudem ist zu bedenken, dass die Studienbewerber\*innen den Test in einer Prüfungssituation, in der sie Zusatzpunkte für ihre Bewerbung erreichen konnten, und die Schüler\*innen den Test freiwillig, anonym und ohne jegliche Aufwandsentschädigung bearbeiteten.

Insgesamt sprechen die Ergebnisse allerdings für die Güte des Testinstruments, das verschiedene Kompetenzniveaus valide erfasst und somit zur Auswahl von Studienbewerber\*innen eingesetzt werden kann. Nächste Schritte zur Überprüfung der Validität des Tests sind sein Einsatz in weiteren Stichproben mit zusätzlichen externen Konstrukten sowie insbesondere die Untersuchung seiner prädiktiven Validität.

## Literatur

- Besser, M., Ehmke, T., Leiss, D. & Hagen, M. (eingereicht). Mathematisches Vorwissen von Bewerber\*innen auf ein Mathematik-Lehramtsstudium: Entwicklung eines Testinstruments und Ergebnisse eines universitären Auswahlverfahrens. *Journal für Mathematik-Didaktik*.
- Graf, T., Harych, P., Wendt, W., Emmrich, R. & Brunner, M. (2016). Wie gut können VERA-8-Testergebnisse den schulischen Erfolg am Ende der Sekundarstufe I vorher sagen? *Zeitschrift für Pädagogische Psychologie*, 30(4), 201–211.
- Lee, J. & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, 65, 50–64.
- Ramm, G., Adamsen, C., Neubrand, M. & Deutsches PISA-Konsortium (Hrsg.). (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Weber, K. E. & Freund, P. A. (2017). Erfassung des Selbstkonzepts von Kindern im Grundschulalter: Validierung eines deutschsprachigen Messinstruments. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 49(1), 38–49.