

Larissa BARTOK, Philipp GEWESSLER, Jan STEINFELD, Bundesministerium für Bildung, Ivo PONOCNY, Modul University Vienna, Elisabeth PONOCNY-SELIGER, Coaching, empirische Sozialforschung und Gender Research, Eva SATTLBERGER, Michael THEMESSEL-HUBER, Bundesministerium für Bildung, AT

## **Geschlechtsspezifische Unterschiede in Mathematik: Welchen Einfluss haben Persönlichkeitseigenschaften? Eine Analyse im Rahmen der Österreichischen Matura.**

Bei der 2015 flächendeckend eingeführten standardisierten schriftlichen Reifeprüfung (SRP) an Gymnasien in Österreich konnte bei allen Klausurterminen im Fach Mathematik ein geschlechtsspezifischer Unterschied hinsichtlich der Lösungsquote und Notenverteilung zuungunsten der Mädchen festgestellt werden. So scheiterten z.B. zum Haupttermin 2016 rund 25 Prozent der Mädchen und rund 17 Prozent der Burschen an der schriftlichen Reifeprüfung in Mathematik. In PISA 2015 und in internationalen Vergleichstests davor (vgl. Suchań & Breit, 2016; OECD, 2014) sticht Österreich mit einem großen Geschlechtsunterschied zwischen Schülerinnen und Schülern (der 8. Schulstufe) hinsichtlich der Mathematikleistungen unter allen teilnehmenden Ländern hervor.

Neben bereits bekannten und manifesten Ursachen für diese Unterschiede, wie der unterschiedlichen Grundgesamtheit der maturierenden Mädchen und Burschen (vgl. Neuwirth, 2015), müssen auch unterschiedliche Ausprägungen von Persönlichkeitseigenschaften und deren Auswirkung auf die Fähigkeiten und der Entwicklung von mathematischen Fertigkeiten in Betracht gezogen werden. Im Rahmen der Feldtestung von Prüfungsaufgaben für die SRP in Österreich wurde in Mathematik der allgemeinbildenden höheren Schule (AHS) 2016 daher ein Fragebogen vorgegeben, der zusätzlich zu technischen Aspekten auch Fragen bezüglich des akademischen Selbstkonzepts enthielt sowie des Interesses an Mathematik und der Motivation, gute Noten zu erhalten. Feldtestungen sind ein Instrument, die die gleichbleibend hohe Qualität der für die SRP erstellten Aufgaben sicherstellen sollen. Unter prüfungsähnlichen Bedingungen bearbeiten Schülerinnen und Schüler der Abschlussklassen eine Reihe von Aufgaben. Diese werden nach dem Rasch-Modell (1PL) skaliert, wobei überprüft wird, ob Subgruppen von Personen (z.B. differenziert nach Geschlecht) durch die Aufgaben systematisch benachteiligt werden (differential item functioning). In den Maturadaten 2016 ist kein differential item functioning nachweisbar. Alle folgenden Analysen wurden mit der freien Statistiksoftware R (R Core Team, 2015) durchgeführt.

## Fragebogen und Datenerhebung

Im Anschluss an die Aufgabenbearbeitung wurde in der Feldtestung 2016 ein Fragebogen zur Sammlung von Persönlichkeitseigenschaften vorgegeben. Die darin enthaltenen Fragen beziehen sich auf das Interesse an Mathematik, auf motivationale Aspekte und auf das akademische Selbstkonzept im Hinblick auf Mathematik (hierarchisches Modell von Shavelson, Hubner & Stanton, 1976). In Summe nahmen 3027 Schülerinnen und Schüler an den Feldtestungen und damit an der Fragebogenuntersuchung teil. Davon bearbeiteten 93.3% sowohl die Feldtestungsaufgaben als auch den anschließenden Fragebogen vollständig. Für 2823 Schülerinnen und Schüler (41.48% Buben, 58.52% Mädchen) war es möglich, Personenfähigkeitsparameter (WLE) mithilfe des R Pakets PP (Reif, 2014) zu schätzen, wobei die Daten sich zuvor als konform mit dem Rasch-Modell herausstellten. Zur Anwendung kam dabei das R Paket eRm (Mair, Hatzinger, & Maier, 2015).

## Resultate

Bei einigen Indikatoren zeigten sich kaum Geschlechtsunterschiede (Item 10: „Die Bearbeitung mathematischer Aufgaben bereitet mir Freude“), einige Indikatoren wiesen hingegen relativ starke Geschlechtsunterschiede auf. So berichteten Mädchen beispielsweise gründlichere Vorbereitung auf Schularbeiten, gaben aber gleichzeitig höhere Überforderung mit mathematischen Inhalten an. Um sowohl explorative als auch konfirmatorische Analysen mit dem Datensatz durchführen zu können, wurde die Gesamtstichprobe randomisiert in eine Konstruktionsstichprobe zur Modellgenerierung ( $N = 1411$ , Geschlechterverhältnis [m:w] = 1:1.45) und eine Validierungsstichprobe zur Modellevaluation ( $N = 1412$ , Geschlechterverhältnis = 1:1.37) aufgeteilt. Zum Finden jenes Faktorenmodell mit der besten Modell-Daten-Passung wurden explorative Faktorenanalysen (geomin Rotation; R Paket psych, Revelle, 2016) durchgeführt. Es resultierte eine 3-Faktoren-Lösung (TLI = .965, RMSEA = .063), wobei die Faktoren entsprechend ihren Ladungen als Interesse, akademisches Selbstkonzept und Motivation identifiziert wurden. Die Faktorkorrelation zwischen Interesse und akademischem Selbstkonzept war hoch,  $r = .69$ , während die Korrelation zwischen Motivation und den anderen Faktoren niedriger ausfiel, beide  $r = .22$ .

Die Zulässigkeit der gefundenen Faktorenlösung wurde in der Validierungsstichprobe mithilfe konfirmatorischer Faktorenanalysen überprüft. Da das Modell unter Berücksichtigung des Faktors Motivation nicht identifiziert war, wurde das Modell ohne diesen Faktor geschätzt. Die Modellierung des 2-Faktoren-Modells ergab eine zufriedenstellende Anpassungsgüte (TLI

= .985, RMSEA = .092) mit hoher Faktorkorrelation,  $r = .76$ . In einem weiteren Schritt wurde überprüft, ob der Fragebogen messinvariant bezüglich des Geschlechts ist. Diese Analysen zeigten, dass in der Stichprobe starke Messinvarianz (invariante Faktorladungen und Schwellenwerte) angenommen werden kann und somit kein erkennbares differential item functioning vorliegt. Durch die Annahme starker Messinvarianz konnten Mittelwertsvergleiche latenter Faktoren vorgenommen werden, mit dem Ergebnis, dass Schüler sowohl über ein höheres Interesse an Mathematik,  $d = 0.23$ , als auch ein höheres akademisches Selbstkonzept verfügen als Schülerinnen,  $d = 0.38$ .

Zur Prüfung der Hypothese, ob Persönlichkeitsmerkmale Einfluss auf die Testleistung von Schülerinnen und Schülern haben, wurden separat lineare Modelle in der Trainings- und Validierungsstichprobe geschätzt. Dazu wurde in der Trainingsstichprobe die bestmögliche Kombination von Prädiktoren auf Basis inhaltlichen Kriterien und Informationskriterien (AIC, BIC) bestimmt (Evaluation in der Validierungsstichprobe). Die abhängige Variable stellte der geschätzte Personenparameter aus der Feldtestung dar. Als Prädiktoren standen das Geschlecht und die extrahierten gewichteten Faktorscores für Motivation, Selbstkonzept und Interesse zur Verfügung. Die größte Assoziation bestand zwischen Selbstkonzept und Testleistung ( $b = 0.75 [0.63, 0.87]$ ). Der Effekt des Selbstkonzeptes wirkte sich bei Buben leicht stärker aus als bei Mädchen ( $b = -0.16 [-0.28, -0.04]$ ). Sowohl Geschlecht als auch das Interesse an Mathematik schienen keinen Einfluss auf die Testleistung in der Trainingsstichprobe zu haben. Das erneute Schätzen des Modells in der Validierungsstichprobe bestätigte den großen Effekt des Selbstkonzeptes auf die Testleistung ( $b = 1.06 [0.86, 1.26]$ ). Den Haupteffekten von Interesse und Geschlecht kann in diesem Modell keine Bedeutsamkeit zugeschrieben werden ( $b = 0.03 [-0.08, 0.13]$  und  $b = 0.01 [-0.10, 0.12]$ ). Der Interaktionseffekt zwischen Geschlecht und akademischem Selbstkonzept aus der Trainingsstichprobe konnte nicht reproduziert werden ( $b = -0.10 [-0.29, 0.09]$ ).

## Diskussion

Ziel der vorliegenden Studie war, a priori bestehende Geschlechtsunterschiede in der Testleistung im Zusammenhang mit Persönlichkeitsmerkmalen zu untersuchen. Burschen berichteten über höhere Werte sowohl im akademischen Selbstkonzept als auch im Interesse für mathematische Inhalte. Die Analysen zeigten, dass ein konsistent großer Effekt des akademischen Selbstkonzeptes auf die Testleistung vorliegt. Der direkte Einfluss des Geschlechts und des Interesses an mathematischen Inhalten scheint den Ergeb-

nissen nach nur eine untergeordnete Rolle zu spielen. Wird also das Selbstkonzept berücksichtigt, können die vorhandenen Geschlechtsunterschiede in der Feldtestungsleistung bereits ausreichend erklärt werden. Der moderierende Effekt des Geschlechts auf das akademische Selbstkonzept hat sich in der Validierungsstichprobe nicht gezeigt. Da in dieser Studie Schülerinnen und Schüler der Abschlussklassen untersucht wurden, ist bei der Interpretation der Ergebnisse zu berücksichtigen, dass die Stichprobe nicht repräsentativ bezüglich des Mädchenanteils in Bezug auf die Gesamtpopulation an gleichaltrigen Schülerinnen und Schülern ist. Die Generalisierung der Ergebnisse auf andere Schultypen ist somit nicht zulässig. Zusätzlich ist davon auszugehen, dass bereits vor dem Eintritt in die Oberstufe der AHS eine Selbstselektion (vgl. Neuwirth, 2015) stattgefunden hat.

Die vorliegende Studie bietet Ansatzpunkte für weitere Untersuchungen und zeigt, dass das akademische Selbstkonzept einen jedenfalls nicht zu vernachlässigenden Einfluss auf die Mathematikleistungen von Schülerinnen und Schülern hat. Basierend auf den Ergebnissen scheint ein Mediationsmodell mit Selbstkonzept als Mediator für zukünftige Untersuchungen angemessen.

## Literatur

- Mair, P., Hatzinger, R., & Maier M. J. (2015). *eRm: Extended Rasch Modeling*. 0.15-6. <http://erm.r-forge.r-project.org/>
- Neuwirth, E. (2015). *Was die Ergebnisse der Zentralmatura wirklich zeigen*. Abgerufen am 21.12.2016, von <http://www.neuwirth.priv.at/forblogs/BLErgebnisse.html>
- OECD (2014). PISA 2012 results: What students know and can do – Student performance in mathematics, reading and science (Volume I, Revised Edition, February 2014). *PISA*, OECD Publishing.
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reif, M. (2014). *PP: Estimation of person parameters for the 1,2,3,4-PL model and the GPCM*. R package version 0.5.3. <https://github.com/manuelfreif/PP>
- Revelle, W. (2016) *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, Illinois, USA, <http://CRAN.R-project.org/package=psych> Version = 1.6.6.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. <http://www.jstatsoft.org/v48/i02/>.
- Suchań B. & Breit S. (Hrsg.). (2016). PISA 2015. *Grundkompetenzen am Ende der Pflichtschulzeit im internationalen Vergleich*. Graz: Leykam.
- Shavelson, R., Hubner, J. & Stanton, G. (1976). Self-Concept: Validation of construct interpretations. *Review of Educational Research*, 46 (3), 407–441.