

WOSCH, Carolin; ROLFES, Tobias & HEINZE, Aiso
Frankfurt a. M., Frankfurt a. M., Kiel

Trendentwicklung der Mathematikleistungen der Abiturient*innen von 1964 und 1996

1. Einleitung

Steigende Abiturient*innenzahlen verbunden mit Klagen von Hochschulen über unzureichende Mathematikleistungen der Studienanfänger*innen führen schon seit den 1960er Jahren zu Diskussionen über die Leistungsfähigkeit der gymnasialen Oberstufe insbesondere im Fach Mathematik. Bemerkenswert ist dabei allerdings, dass es in Deutschland bisher kein systematisches Bildungsmonitoring der Mathematikleistungen in der gymnasialen Oberstufe gibt (Rolfes et al., 2023). Somit basiert das Urteil über die abnehmenden Mathematikleistungen der Abiturient*innen bzw. der Studienanfänger*innen bisher vor allem auf subjektiven Erfahrungen der Lehrkräfte bzw. der Hochschuldozierenden. In welcher Weise diese Beurteilungen der "Wirklichkeit" entsprechen, ist noch unklar. Nicht ausgeschlossen ist, dass diese Einschätzungen der Trendentwicklung verzerrt sind, da Beurteilungen von vergangenen Ereignissen nicht selten vom Phänomen der Nostalgie geprägt sind (Petersen, 2020).

Um die noch offene Frage der Trendentwicklung der Mathematikleistungen von Abiturient*innen empirisch zu untersuchen, führten Rolfes et al. (2023) eine Sekundäranalyse von Schulleistungsstudien des Instituts for Educational Achievement (IEA) durch. Hierbei wurden Daten aus Deutschland von der First International Mathematics Study (FIMS) des Jahres 1964 und der Third International Mathematics and Science Study (TIMSS) des Jahres 1996 verglichen. Auf der Grundlage neun gemeinsamer Trenditems gab es keine signifikanten Unterschiede zwischen den Leistungen der Abiturient*innen von FIMS und TIMSS. Daher konnte entgegen der weitverbreiteten Meinung keine negative Trendentwicklung von 1964 nach 1996 festgestellt werden. Allerdings bildete die relativ geringe Anzahl an Trenditems eine Limitation der Studie, sodass der Linkfehler vergleichsweise groß war.

Ein alternatives Vorgehen zur Bestimmung der Trendentwicklung von FIMS nach TIMSS ermöglicht ein Linking über die Second International Mathematics Study (SIMS) der Jahre 1980-1982, da FIMS und SIMS mit 30 Trenditems sowie SIMS und TIMSS mit 33 Trenditems verknüpft sind. Zwar hat Deutschland nicht an SIMS teilgenommen, allerdings können die Daten aus den SIMS-Teilnehmerländern verwendet werden, um eine Trendentwicklung über eine größere Anzahl an Trenditems abzuschätzen. Das Ziel der vorliegenden Studie ist folglich, über ein Linking mithilfe der SIMS-Daten

zu ermitteln, in welcher Weise die dargestellten Ergebnisse zu der Trendentwicklung der gymnasialen Oberstufe in Mathematik zwischen 1964 und 1996 (Rolfes et al., 2023) repliziert werden können.

2. Methode

2.1. Stichprobe

Grundlage der Reanalyse waren die Daten der Abiturient*innen aus Deutschland aus FIMS (Husén, 1967) und TIMSS (Baumert et al., 2000) und die internationalen Daten aus SIMS, in welcher der Abschlussjahrgang der Sekundarstufe II getestet wurde (Garden, 1987). Für FIMS bzw. TIMSS lagen Daten von 1 292 bzw. 2 246 Testpersonen aus Deutschland vor. An SIMS nahmen international 42 463 Testpersonen teil, darunter 2 712 Teilnehmerinnen und Teilnehmer aus Schweden, das als eines der wenigen Länder an FIMS, SIMS und TIMSS teilgenommen hat.

2.2. Testmaterial, -design und -administration

Bei FIMS wurden 196 Items auf sechs verschiedene Testhefte verteilt administriert (Husén, 1967). Bei SIMS wurden 136 Items auf acht Testhefte verteilt eingesetzt (Garden, 1987), wohingegen bei TIMSS 68 Items in vier Clustern in einem Multi-Matrix-Design administriert wurden. Wie bereits erwähnt, sind FIMS und SIMS durch 30 Trenditems sowie SIMS und TIMSS durch 33 Trenditems verknüpft. Darunter befinden sich zehn Trenditems, die in allen drei Studien eingesetzt wurden.

2.3 Auswertung

Die folgenden Analysen basieren auf der probabilistischen Testtheorie. Im ersten Schritt wurde ein Land aus SIMS ermittelt, in dem ein möglichst geringes Ausmaß an differenziellem Itemfunktionieren (DIF) zu Deutschland festgestellt wurde. Insbesondere für Schweden und die Vereinigten Staaten konnte dieses überprüft werden, da beide Länder an allen drei IEA-Schulleistungsstudien teilgenommen hatten. Daher wurde jeweils für beide Länder eine Itemskalierung separat für FIMS und TIMSS vorgenommen. Darauf aufbauend wurde die Korrelation zwischen den Itemschwierigkeiten von Schweden bzw. den Vereinigten Staaten und Deutschland berechnet, um das Ausmaß an DIF einzuschätzen.

Im zweiten Schritt wurde die Trendentwicklung anhand der 30 bzw. 33 Trenditems mithilfe zweier Linking-Methoden ermittelt. Zum einen wurde eine Fixed-Parameter-Kalibrierung genutzt. Das bedeutet, dass zunächst die SIMS-Daten skaliert und darauf aufbauend die Itemschwierigkeiten der Trenditems in der FIMS- und TIMSS-Auswertung auf die SIMS-Itemschwierigkeiten fixiert wurden. Durch eine FIMS- bzw. TIMSS-Skalie-

rung mit fixierten SIMS-Itemschwierigkeiten konnten nun die Personenfähigkeiten der Abiturient*innen in Deutschland auf einer gemeinsamen Skala geschätzt und miteinander verglichen werden.

Zum anderen wurde ein Mean-Mean-Linking umgesetzt, wofür die FIMS-, SIMS- und TIMSS-Daten ohne fixierte Itemschwierigkeiten separat skaliert wurden. Anschließend wurden jeweils für die Trenditems in FIMS und TIMSS die mittlere Itemschwierigkeit in der deutschen Stichprobe und entsprechend die mittlere Itemschwierigkeit der Trenditems für die SIMS-Skalierung bestimmt. Mit diesen mittleren Itemschwierigkeiten der Trenditems konnten nun die Mathematikleistungen von FIMS, SIMS und TIMSS in Beziehung gesetzt werden.

3. Ergebnisse

Zwischen den Itemschwierigkeiten der deutschen und amerikanischen Stichprobe ergab sich für FIMS nur eine schwache Korrelation, wohingegen die Itemschwierigkeiten für TIMSS stark korrelierten. Dagegen bestand für Schweden sowohl für die FIMS- als auch für die TIMSS-Itemschwierigkeiten eine hohe Korrelation. Dies bedeutete, dass die Itemschwierigkeiten Schwedens für FIMS und TIMSS vergleichbarer mit denen Deutschlands waren und Schweden daher als geeigneteres Land für das oben beschriebene Linking über SIMS angenommen werden konnte.

Sowohl bei der Fixed-Parameter-Kalibrierung als auch bei dem Mean-Mean Linking zeigte sich eine Verschlechterung der Leistungen der Abiturient*innen von 1964 bis 1996. Diese Differenz lag in den beschriebenen Analysen zwischen 0.3 und 0.5 Logit, was in etwa einer Verringerung der Lösungsrate um sechs bis zehn Prozentpunkte entspricht. Es führte zu sehr ähnlichen Ergebnissen, wenn nur Schweden oder die gesamte internationale SIMS-Stichprobe als Datengrundlage verwendet wurde. Insbesondere die Differenz von SIMS nach TIMSS im Mean-Mean-Linking deutet auf einen überzufälligen Leistungsunterschied hin.

4. Diskussion

Im Rahmen der Reanalyse wurden die Mathematikleistungen der deutschen Abiturjahrgänge bei den Schulleistungsstudien FIMS und TIMSS anhand der SIMS-Daten verglichen. Dadurch konnte ein Linking mit einer größeren Anzahl an Trenditems durchgeführt werden. Hierbei wurde eine Verschlechterung der Mathematikleistungen von Abiturient*innen von 1964 nach 1996 festgestellt. Daher zeichnete sich eine abweichende Trendentwicklung als in der vorangegangenen Studie mit weniger Trenditems und entsprechend höherem Linkfehler (Rolfes et al., 2023) ab. Allerdings hält sich auch bei den

hier berichteten Analysen die negative Trendentwicklung in Anbetracht einer Verdreifachung der Abiturient*innenrate (FIMS: 8%; TIMSS: 25%) noch in überschaubaren Grenzen.

Auch wenn die vorliegende Reanalyse die Limitation der geringen Anzahl an Trenditems in der Studie von Rolfes et al. (2023) verringert hat, sind auch bei der hier berichteten Reanalyse Grenzen zu beachten. Schweden wurde als Land mit dem geringsten DIF zu Deutschland bei FIMS und TIMSS identifiziert. Allerdings ist schwer abschätzbar, in welcher Weise das trotzdem vorhandene DIF die Ergebnisse beeinflusst, auch wenn die Verwendung der gesamten internationalen Stichprobe von SIMS zu ähnlichen Ergebnissen führt.

Weiterer Forschungsbedarf besteht in der Nutzung alternativer Skalierungs- und Linking-Methoden, um die Variabilität der Trendergebnisse von FIMS nach TIMSS noch besser abschätzen zu können. Zudem stellt sich weiterhin die Frage, inwieweit sich die Trendentwicklung bis in die Gegenwart fortsetzt. Außerdem ist unklar, in welcher Weise sich die Mathematikleistungen innerhalb der gymnasialen Oberstufe von der Eingangsstufe bis zum Abitur entwickeln und wie gezielte Unterstützungsangebote zum Beispiel im Übergang von der Sekundarstufe I zur Sekundarstufe II zu einer positiven Leistungsentwicklung im Fach Mathematik beitragen können.

Literatur

- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000). *Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie—Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 2). VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-322-83411-9>
- Garden, R. A. (1987). *Second IEA Mathematics Study. Sampling Report*. International Association for the Evaluation of Educational Achievement; New Zealand Department of Education.
- Husén, T. (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries* (Bd. 1). Almqvist & Wiksell.
- Petersen, T. (2020). Früher war das Wetter auch schon früher besser. *Zeitschrift für Politik*, 67(2), 194–213. <https://doi.org/10.5771/0044-3360-2020-2-194>
- Rolfes, T., Robitzsch, A., & Heinze, A. (2023). Früher war alles besser? Mathematikleistungen von Abiturientinnen und Abiturienten von 1964 und 1996 im Vergleich. *Zeitschrift für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-023-01176-6>