

Bessere PISA-Ergebnisse durch selektives Testen

Von Rainer Bölling

Ein wenig beachteter, aber gravierender Schwachpunkt der PISA-Studie liegt darin, dass Länder bzw. Volkswirtschaften miteinander verglichen werden, in denen die für den Test relevante Schülerpopulation sehr unterschiedlich erfasst wird. Welche Folgen das haben kann, sei an einem einfachen Beispiel verdeutlicht.

Wir nehmen an, dass in einem Land idealerweise 100 Prozent aller für eine repräsentative Stichprobe gezogenen Schülerinnen und Schüler am Test teilnehmen und auf einer siebenstufigen Skala folgende Punktverteilung erreichen:

Punkteskala	300	350	400	450	500	550	600	650	700	Gesamt
Messwerte	5	11	12	14	16	14	12	11	5	100
Punktzahl	1500	3850	4800	6300	8000	7700	7200	7150	3500	50000
Mittelwert										500

Diese Punkteverteilung entspricht ziemlich genau der Gaußschen Normalverteilung, die PISA zugrunde liegt: der Mittelwert aller Messwerte beträgt 500, und 68 Prozent aller Werte liegen nicht mehr als eine Standardabweichung (100 PISA-Punkte) vom Mittelwert entfernt, also zwischen 400 und 600 Punkten. Ein Land mit diesen Testwerten landet im internationalen Vergleich im Mittelfeld und muss sich wohl vorhalten lassen, dass viele Länder besser abgeschnitten haben.

Nun stellen findige Leute in diesem Land fest, dass in anderen Ländern gar nicht alle Probanden getestet wurden, so dass der durchschnittliche Erfassungsgrad bei PISA nur 89 Prozent beträgt.¹ Wenn aber „der Erfassungsgrad außer Acht gelassen wird, dürften die Leistungen der 15-jährigen Kohorte eines Bildungssystems überschätzt werden, und zwar umso mehr, je niedriger der Erfassungsgrad ist“, räumt die OECD ein.² Offensichtlich machen andere Länder von der Möglichkeit Gebrauch, „Schüler mit kognitiven oder körperlichen Behinderungen oder mangelnden Kenntnissen in der Testsprache (Schüler, die z.B. weniger als ein Jahr lang Unterricht in der Testsprache hatten)“ vom Test auszuschließen.³

Das geschah 2018 besonders in Schweden, wo laut OECD „die hohe Zahl der Schülerinnen und Schüler, die seit 2015 als Zuwanderer oder Flüchtlinge in das Land gekommen waren, zu einem deutlichen Anstieg der Ausschlussquote (um 5 Prozentpunkte) und einem starken Rückgang des Erfassungsindex 3 (um 8 Prozentpunkte) führte“.⁴ So wurden 2018 nur 86 Prozent aller Fünfzehnjährigen vom PISA-Test erfasst, und Schweden verbesserte sich gegenüber den vorhergehenden Erhebungen auf durchschnittlich 502 Punkte.

Auch unser Land hat seit 2015 viele Migranten aufgenommen, darunter Schülerinnen und Schüler, die auch nach einem Jahr Unterricht in der Testsprache diese verständlicherweise noch nicht wirklich beherrschten. Angesichts dessen meint man nun in unserem Land, ein Erfassungsgrad von 90 Prozent sei völlig ausreichend, und streicht die schlechtesten 10 von 100 Messwerten. Das Ergebnis ist folgendes:

Punkteskala	300	350	400	450	500	550	600	650	700	Gesamt
Messwerte		6	12	14	16	14	12	11	5	90
Punktzahl		2100	4800	6300	8000	7700	7200	7100	3500	46750
Mittelwert										519

Damit liegt dieses Land nun nicht nur vor Schweden, sondern auch vor Finnland und Kanada, die immer wieder als Vorbild hingestellt werden, aber bei PISA 2018 im Durchschnitt der drei Testbereiche nur auf 516 bzw. 517 Punkte kommen.

Doch das ist noch nicht alles, denn bei PISA genügt offiziell schon eine Mindestbeteiligungsquote von 80 Prozent. Die chinesischen Provinzen Peking, Shanghai, Jiangsu und Zhejiang, die 2018 den Spitzenplatz einnahmen, erreichten nur einen Erfassungsgrad von 81 Prozent.⁵ Und als Shanghai 2012 selbst diese Hürde riss, wurde es trotzdem von der OECD als Sieger und Vorbild für die Welt präsentiert.⁶ Außen vor blieben dort die Kinder von Millionen Wanderarbeitern, weil sie gar nicht die weiterführenden Schulen besuchen durften, an denen der PISA-Test stattfand.⁷ Warum also sollte unser Land nicht nach chinesischem Vorbild 20 Prozent der Schülerpopulation unberücksichtigt lassen? Schließt man gezielt weitere schwächere Schülerinnen und Schüler aus, sieht das Ergebnis so aus:

Punkteskala	300	350	400	450	500	550	600	650	700	Gesamt
Messwerte			8	14	16	14	12	11	5	80
Punktzahl			3200	6300	8000	7700	7200	7150	3500	43050
Mittelwert										538

Mit diesen 538 Punkten hätte unser Land bei PISA 2018 seinen Platz im Mittelfeld gegen eine Bronzemedaille getauscht, ohne dass auch nur ein Testergebnis besser geworden wäre. Mit einer Erfassungsquote von 99 Prozent, der höchsten aller teilnehmenden Länder, ist ein solcher Spitzenplatz jedoch nicht zu erreichen. Das ist der Nachteil deutscher Gründlichkeit. Denn tatsächlich geht es hier um Deutschland, das 2018 bei PISA auf durchschnittlich 500 Punkte kam.

Der hier modellhaft aufgezeigte Zusammenhang zwischen Erfassungsgrad und PISA-Punktzahl ist der OECD durchaus bekannt, doch geht sie darauf nur im „Kleingedruckten“ ein und spielt das Problem herunter. So findet sich in den Berichtsbänden – nicht aber in den Pressemitteilungen – seit 2003 an unterschiedlichen Stellen folgender Satz: „Selbst wenn man z.B. unterstellt, dass die ausgeschlossenen Schülerinnen und Schüler durchgehend schlechtere Ergebnisse erzielt hätten als die teilnehmenden und dass dieser Zusammenhang mittelstark ausgeprägt ist, würde eine Ausschlussquote in einer Größenordnung von 5% wahrscheinlich nur zu einer Überzeichnung der nationalen Durchschnittsergebnisse um weniger als 5 Punkte auf der PISA-Skala führen.“⁸ Unsere Beispielrechnung zeigt, dass dieser Effekt fast doppelt so stark sein kann.

Das Problem unterschiedlicher Teilnahmequoten betrifft auch andere internationale Vergleichsstudien wie TIMSS (Trends in Mathematics and Science Study), deren letzte Erhebung 2019 durchgeführt wurde. TIMSS geht aber offener damit um. Hier werden Abweichungen von den vorher definierten Stichprobenumfängen nachvollzogen und dokumentiert, „damit bei vertiefenden Vergleichen von Staaten nur solche einander gegenübergestellt werden, die tatsächlich aufgrund gleicher Teilnahmebedingungen vergleichbar sind“.⁹ So findet sich unter 49 Tabellen mit Darstellungen der Ergebnisse bei gut 20 Teilnehmerstaaten folgende Fußnote: „Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.“ In den meisten Fällen liegt die Ausschlussquote zwischen fünf und zehn Prozent, nur im Falle des Spitzenreiters Singapur deutlich darüber (12,8).¹⁰ Zudem werden in zwei Tabellen die Mittelwerte für Deutschland dargestellt, die sich ergeben würden, wenn leistungsschwächere Schülerinnen und Schüler bis zum Erreichen einer fiktiv höheren Ausschlussquote ausgeschlossen würden.¹¹

Die OECD dagegen hat 2015 Vietnam in ihre Vergleichstabellen aufgenommen, ohne an gleicher Stelle darauf hinzuweisen, dass dort nur knapp 49 Prozent der Fünfzehnjährigen vom PISA-Test erfasst wurden. Erst nach Kritik an diesem Vorgehen¹² findet sich im Bericht 2018 folgender verklausulierter Hinweis: „Die Daten zu Vietnam wurden in den meisten Tabellen in Anhang B berücksichtigt, nicht aber in Tabellen, Abbildungen und Texten zu Leistungsvergleichen mit anderen Ländern und Volkswirtschaften oder zu Leistungsvergleichen im Zeitverlauf, da eine umfassende internationale Vergleichbarkeit der Ergebnisse zum Zeitpunkt der Veröffentlichung des Berichts nicht gewährleistet werden konnte.“¹³ Bei einem Erfassungsgrad von mittlerweile knapp 70 Prozent kann in der Tat von Vergleichbarkeit auch weiterhin nicht die Rede sein.

(Februar 2021)

-
- 1 <http://www.oecd.org/berlin/themen/pisa-studie> (22.11.2019).
 - 2 OECD: PISA 2018 Ergebnisse (Band I), Was Schülerinnen und Schüler wissen und können, S. 57.
 - 3 <http://www.oecd.org/berlin/themen/pisa-studie/hacufig-gestellte-fragen.htm> (10.01.2021).
 - 4 OECD: PISA 2018, S. 58.
 - 5 OECD: PISA 2018, S. 183.
 - 6 OECD, PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Band I (überarbeitete Ausgabe, Februar 2014): Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften, 2014, S. 20.
 - 7 Tom Loveless, Attention OECD-PISA: Your-Silence-on-China-is-Wrong, 12.12.2013 (<http://educationnext.org/attention-oecd-pisa-your-silence-on-china-is-wrong>); Valerie Strauss, Did Shanghai cheat on PISA? 15.12.2013 (<http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/12/15/did-shanghai-cheat-on-pisa>)
 - 8 OECD: PISA 2018, S. 179.
 - 9 Knut Schwippert u.a.: TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich, Münster 2020, S. 39.
 - 10 Ebd., S. 354.
 - 11 Ebd., S. 40-42.
 - 12 Vgl. Rainer Bölling, Zur Fragwürdigkeit des PISA-Rankings, 2017 (<https://bildung-wissen.eu/fachbeitraege/zur-fragwuerdigkeit-des-pisa-rankings.html>).
 - 13 OECD: PISA 2018; S. 23.