

Bildungsstatistik, die wir nicht brauchen

Die IQB-Studie

Den ausführlichen Bericht des IQB „Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich 2015“ (Stanat u. a. 2016), den man im Netz herunterladen kann, habe ich mir angesehen. Leider finde ich hier alle meine sachlichen und methodologischen Vorbehalte bestätigt. Diese Studie gibt nichts her, was solche Meldungen wie „dramatische Veränderungen“ oder „Baden-Württemberg hat der Strukturreform zu viel Aufmerksamkeit geschenkt und die Qualität des Unterrichts vernachlässigt“ belegt. Die Studie ist handwerklich miserabel gemacht. Bildungsforscher von diesem Kaliber sollten wissen, dass man Daten nicht bis zur Unkenntlichkeit „transformiert“ und dass man alternativen Erklärungsmöglichkeiten nachgehen muss, die allen Experten bekannt sind.

Die Art und Weise, wie das IQB die Ergebnisse seiner Untersuchung berichtet (siehe unten), sollte uns misstrauisch machen: Keine Angabe von Testwerten, keine Angabe der Zahl der gelösten Aufgaben, keine Vergleich dieser Angaben zwischen den Bundesländern und zwischen den Erhebungszeiträumen. Solche Angaben gehören in den USA zum Standard guter Berichterstattung der Bildungsforschung. Ohne solche Angaben ist es der Öffentlichkeit nicht möglich, die Schlussfolgerungen, die interessierte Kreise sofort ziehen, zu überprüfen. Ohne solche Angaben sind alle Folgerungen für die Lehrerbildung und die Bildungspolitik unbegründet. Sollte das IQB all das in seinem Bericht an die Ministerien angegeben haben, hätte das Kultusministerium mit der Auslassung dieser Informationen der Bildung einen Bärendienst erwiesen.

Stattdessen werden uns Rangplätze berichtet, wie in Sportberichten, die im Bildungsbereich völlig fehl am Platz sind, weil sie rein gar nichts über die wahren Verhältnisse aussagen. Stattdessen werden „statistisch signifikante“ Unterschiede berichtet, die etwas über die Größe der Untersuchungsgruppe aussagen (bei so großen Gruppen wird statistisch fast alles „signifikant“), aber nicht über die wahren Leistungsunterschiede. Stattdessen werden die Unterschiede statistisch in Schuljahre umgerechnet, die zu absurden Befunden führen, aber nicht die tatsächliche Lernleistung widerspiegeln.

Zudem werden wichtige Faktoren, die für die (vermutlich sehr geringen) Unterschiede verantwortlich sein können, offenbar nicht erhoben oder im offiziellen

Bericht unterschlagen: Faktoren wie die Zahl der Immigranten mit Sprachdefiziten. Baden-Württemberg und Bayern sind die Hauptzuwanderungsländer. Kein Wunder, dass die Deutschleistungen hier (vermutlich leicht) abgefallen sind. Gewonnen haben vor allem die Länder mit geringer Ausländerakzeptanz. Wie viele Schüler werden dort in Sonderschulen abgeschoben? Ausnahme: Schleswig-Holstein, das große Anstrengungen unternommen hat, schwache Schüler besser zu integrieren. Nicht berichtet wird auch die Zahl der vor dem 9. Schuljahr sitzengebliebenen Schüler. Aus den USA ist bekannt, dass Schulen schwache Schüler vor dem Schuljahr, in denen getestet wird, verstärkt wiederholen lassen, um die Testwerte zu schönen. Bei uns wird das Thema „Test-Schönung“ noch totgeschwiegen.

Solange die Bildungsstatistik nicht Minimalanforderungen an eine umsichtige und transparente Berichterstattung genügt, sollten wir sie ignorieren (und nicht weiter mit Aufträgen füttern). Bei diesen Studien wird wissenschaftliche Arbeit vorgetäuscht. Aber das IQB ist nur ein An-Institut der Humboldt Universität Berlin und gibt zudem die Aufträge der KMK meist weiter an kommerzielle Umfrageinstitute, die mit diesen Ergebnissen Geld verdienen müssen und keine Zeit haben, über die Zahlen, die sie erheben, lange nachzudenken. Das hat für die Kultusminister den fragwürdigen Vorteil, dass sie mehr „Durchgriff“ haben als bei regulären Forschern, die einen Ruf zu verlieren haben. Sie nehmen damit aber eine Qualitätsminderung in Kauf, die unserem Bildungssystem nicht gut bekommt.

Die Medien benötigen unbedingt Empirie-Beauftragte ähnlich den Beauftragten für juristische Fragen, die solche Berichte kommentieren können, wenn sie auf Sendung gehen.

Der PISA-Schwachsinn hört nicht auf

PISA-Vertreter vermarkten die einträgliche Testerei mit starkem Tobak: „PISA-Schock“, „Rakete hat nicht gezündet“, „Deutschland im Mittelfeld“ etc. etc. Journalisten tun oft nicht mehr, als die Werbematerialien PISA-Industrie umzuschreiben, und Bildungspolitiker, kaufen das ab und machen es zur Grundlage ihrer Entscheidungen. Von Starksinn ist wenig zu sehen und zu hören, in diesem Heft von „Lehren & Lernen“ finden sich einige Stimmen.

Ein tiefer Griff in die Werbekiste stellt auch der immer wieder herausgestellte „Praxisbezug“ der PISA-Aufgaben dar. Der Praxisbezug besteht aber nicht darin, dass die Schüler wirklich praktische Aufgaben bekommen. Der angebliche Praxisbezug wird nur durch Text herzustellen versucht, den die Testteilnehmer lesen müssen.

Der PISA-Mathe-Test korreliert daher so hoch mit der PISA-Lesefähigkeit, dass man hier von einem Lesetest sprechen muss. Bei 150 Sekunden Lösungszeit für einzelne Aufgabe fällt die Zeit zum Lesen der Aufgaben stark ins Gewicht. Schüler mit Deutsch als Zweitsprache und deutsche Schüler mit Leseschwäche sind bei PISA außerordentlich benachteiligt, und damit auch die Bundesländer, in denen viele davon die öffentlichen Regelschulen besuchen.

Selbst progressiven Bildungsjournalisten und -politiker suchen und finden immer noch etwas Gutes bei PISA: Solche Tests würden die Bildungsungleichheit bei uns aufdecken. Aber dazu brauchen wir diese Tests nicht! Ein Blick in die öffentlich zugänglichen Zahlen des Bundesamts für Statistik tut das auch und noch dazu viel klarer und verständlicher. Man muss sich nur die Übergangsquoten ins Gymnasium und die Universität anschauen und wie diese von der sozialen Herkunft abhängen.

Der Rest ist Kaffeesatz-Leserei. Kein Bildungsforscher, der von PISA & Co unabhängig ist, bricht in Jubel aus, wenn die Medien sich wie heute wieder über das Thema hermachen. Liebe Journalisten und Politiker, schaut euch doch einmal die Fakten an:

Es ist nicht auszuschließen, dass die Qualität eines Bildungssystems sich irgendwie in den PISA-Befunden niederschlägt. Bei den Ländern, die abgeschlagen auf den Plätzen 69 bis 71 liegen, ist die Situation sicherlich nicht gut. Aber was sagen die Unterschiede zwischen den Plätzen 1 bis 27, die gerade einmal 61 von schätzungswise 1.000 möglichen Punkten auseinander liegen? Bei PISA wird durch Zahlenakrobatik eine höhere Genauigkeit vorgetäuscht als gegeben ist.

Verdächtig ist, dass die Streuung der Werte um den künstlichen Mittelwert von 500 recht klein ist. Sie sieht sehr nach einer Verteilung von Messfehlern aus und nicht von substantiellen Kompetenzunterschieden.

Solche Differenzen sind natürlich „statistisch signifikant“, da die Untersuchungsgruppen dank der Größigkeit der Kultus- und Finanzminister riesig sind. Da wird jede noch so unbedeutende Differenz „signifikant“. Das ist arglistige Täuschung, die nicht von PISA erfunden wurde, aber gern genutzt wird.

Im Detail: Deutschland liegt ca. 50 Punkte hinter Singapur. Bereiten die Schulen dieses Landes besser auf die Erfordernisse des Lebens vor als deutsche Schulen? Ob PISA etwas mit dem Leben zu tun hat, wird immer wieder behauptet, gezeigt hat es noch keiner.

Wuttke (2006) und andere zeigen, wie wenig PISA-Punkte wirklich etwas über die Qualität des Schulsystems aussagen, sondern über viele Faktoren, die bildungspolitisch völlig irrelevant sind:

50 Punkte Unterschied waren bei PISA I gerade einmal 2,5 Aufgaben-Lösungen. Bricht die Welt zusammen, wenn unsere Schüler 2,5 Mathe-Aufgaben von insgesamt 26 weniger gelöst haben als Schüler in Singapur? Haben sie vielleicht sogar gleich gute oder bessere Lösungen gehabt als die Testkonstrukteure und sind dafür mit Punkte-Abzug bestraft worden?

In Deutschland war bisher die Teilnehmerrate an PISA sehr hoch. Wie hoch sie aktuell war, wird in den Medien nicht berichtet. Aus früheren PISA-Studien weiß man, dass die Teilnehmerrate zwischen den Ländern stark schwankt und diese Unterschiede einen so großen Einfluss auf die Testergebnisse haben, dass sich eine inhaltliche Interpretation der Differenzen verbietet.

Auch sprachliche Unterschiede der Tests haben einen Einfluss auf die Testergebnisse, wie viele Studien zeigen. Ist der Test in Deutsch geschrieben, dann ist er länger als im Englischen, aber die Zeit für die Lösung ist gleich kurz. Das kann schon eine Differenz von 10 Test-Punkten ausmachen.

Diese Liste ließe sich fortsetzen, muss aber nicht, da schon durch die obigen Beispiele klar wird, dass alle gut daran täten, Pressemeldungen der PISA-Vertreter nicht für bare Münze zu nehmen, sondern sie von unabhängigen Fachleuten begutachten und diskutieren zu lassen.

Auch wäre es wünschenswert, dass Journalisten und Politiker sich in das Gebiet der Bildungsforschung soweit einarbeiten würden, dass sie sich ein eigenes Urteil bilden können. Sie würden damit Schülern und Lehrern einen großen Gefallen tun, die PISA-Meldungen am Ende ausbaden müssen, und auch der Wirtschaft, die nichts mit Absolventen anfangen kann, die nur test-schlau, aber nicht wirklich kompetent sind.

Literatur

- Brügelmann, H.: „Erster, Zweiter, Dritter – Letzter“. Was sagen uns IQB-Bildungstrend, TIMSS, PISA und andere Ländervergleiche? In diesem Heft.
- Stanat, P./Böhme, K./Schipolowski, S./Haag, N. (Hrsg.): IQB-Bildungstrend 2015. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich. Münster 2016.
- Wuttke, J.: Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung. In: Jahnke, T./Meyerhöfer, W. (Hrsg.): PISA & Co. Kritik eines Programms. Hildesheim 2006, S. 101-154.