

Peter BENDER, Paderborn (in: Beiträge zum MU 2005, Franzbecker)

## **Neue Anmerkungen zu alten und neuen PISA-Ergebnissen und -Interpretationen**

1. In der OECD (Organisation for Economic Co-Operation and Development) haben sich entwickelte, den USA verbundene Staaten und einige, die auf dem Weg dahin sind, zusammengeschlossen. Eine der Aufgaben ist die Erhebung und Publikation ökonomischer Daten zur Unterstützung der Entscheidungsträgerinnen & -träger in Politik, Wirtschaft usw. Aus dieser Perspektive ist auch der Bildungsstand der 15-Jährigen Teil des Kapitals eines Landes, und er wird in der PISA-Studie in den Jahren 2000, 2003 und 2006 in den OECD- und einigen sog. Partner-Ländern in den grundlegenden Bereichen "Lesen", "Mathematik" und "Naturwissenschaften" untersucht.

2. In jedem Land werden jedes Mal Stichproben von (meistens) etwa 5000 Jugendlichen gezogen. Diese werden in den 3 Bereichen getestet. Die Ergebnisse werden so geeicht, dass in jedem Bereich für die OECD-Jugendlichen der Mittelwert 500 und die Standardabweichung 100 beträgt. Die Zahlen sagen also nur etwas über den *relativen* Stand beim jeweiligen Test, und unmittelbare Punktzahlvergleiche zwischen verschiedenen Bereichen oder zwischen verschiedenen Durchgängen sind i.d.R. unzulässig bzw. mit größter Vorsicht zu führen. Ich lasse mich auf solche nur ein, um die Fragwürdigkeit vieler Schlüsse bereits PISA-endogen auf den Punkt zu bringen.

3. Außer vom Leistungsvermögen der Jugendlichen und der Anpasstheit des jeweiligen nationalen Curriculums an die inhaltlichen PISA-Vorstellungen sind die PISA-Zahlen von vielen weiteren Einflussfaktoren abhängig, z.B. vom tatsächlichen Alter der Jugendlichen.

4. Es wurden Länderdurchschnittszahlen ausgerechnet und in Rangfolgen gebracht. Deutschland schnitt nicht gut ab; z.B. erzielte es im Jahr 2000 in Mathematik 490 Punkte; weil nur etwa 44% der Jugendlichen in Deutschland den von PISA definierten sog. "Grundbildungsstandard" erreichten.

5. In Deutschland (und in anderen "schlechten" Ländern, z.B. USA) rückte dadurch Bildung in den Fokus von Medien und Politik. Leider wurden die PISA-Statistiken von Manchen mit z.T. haarsträubenden Argumenten zur Unterstützung ihrer (bildungs-) (politischen, pädagogischen) (Vor-) Urteile (Entscheidungen) missbraucht.

6. Der Mathematik-Test 2000 wurde in Deutschland um einen nationalen Test erweitert, mit Vergleichen u.a. zwischen den Bundesländern und zwischen den Schulformen. Dabei schnitt Bayern am besten und die integrierte Gesamtschule erheblich schlechter als die Realschule ab.

7. Außerdem werden, auf der Basis eines probabilistischen Testmodells mit sehr stark idealisierten Aufgabencharakteristiken, den Testaufgaben gemäß ihren Schwierigkeitsgraden (je weniger Jugendliche eine Aufgabe lösen, desto schwerer ist sie) Punktzahlen zugeordnet, so dass sie mit ihren Schwierigkeitswerten auf derselben Skala wie die Jugendlichen mit ihren Leistungswerten angesiedelt werden können.

8. Diese Skalen wurden willkürlich ("dividing ... into levels though useful, ... [is] essentially arbitrary", Technical Report, 197) in sog. Kompetenzstufen eingeteilt, deren inhaltliche Beschreibung in Mathematik den nationalen Gruppen überlassen wurde. Die deutsche PISA-Mathematik-Gruppe (DPMG) hat sich mit der inhaltlichen Füllung viel Mühe gegeben, mit dem Nahziel, bei jeder Aufgabe vorhersagen zu können, auf welcher Stufe sie landet, und mit dem Fernziel, mathematikdidaktische Testkonstruktion auf eine wissenschaftliche Basis zu stellen und vom Laientum zu befreien.

9. Bei allen solchen Stufungen (auch bei TIMSS und IGLU) sind die Beschreibungen der stufenbezogenen Kompetenzen problematisch. Oft sind sie trivial und/oder nicht operational. Verbindet man die Eigenschaften einer Stufe mit "und", werden viele Aufgaben nicht erfasst, verbindet man sie mit "oder", verliert die Stufe ihre Identität. Es treten immer wieder "widerpenstige" Aufgaben auf, die auf einer Stufe landen, wo sie wirklich nicht hingehören. Sicher ist nur die Tautologie: Je mehr Punkte, desto schwerer.

10. Zwecks Beseitigung dieser Probleme hat die DPMG u.a. das Konstrukt "Typen des mathematischen Arbeitens" mit 3 Ausprägungen und 3 eigenen Kompetenzstufungen, wenn auch mit gleichen Schwellenwerten, eingeführt, wonach jede Aufgabe einem der 3 Typen zuzuordnen ist. Dieses Konstrukt wird allerdings von PISA-2003-international ignoriert. In der Tat ist es weich mit unklarer Begrifflichkeit, fließenden Übergängen, weiten Überschneidungs-Bereichen, großen Lücken. Außerdem müssten oft Teile von Aufgaben unterschiedlichen "Typen" zugeordnet werden.

11. Eine hochrelevante Differenzierung wurde von Meyerhöfer identifiziert: Bei vielen Aufgaben erfordern unterschiedliche Lösungswege unterschiedliche Kompetenzen und führen zu unterschiedlichen Schwierigkeitswerten und damit oft zur Einordnung in unterschiedliche Stufen (unabhängig davon, ob diese Lösungswege tatsächlich gegangen werden). Statt Aufgaben müssten Paare von Aufgaben und Lösungswegen ins Auge gefasst werden. Diese Option ist natürlich nicht praktikabel. Sie soll hier lediglich auf den Punkt bringen, dass eine eindimensionale Skala, auf der jeder Aufgabe (jedem Item) genau ein Schwierigkeitswert zugeordnet ist, aus dem Rückschlüsse auf die zur Lösung erforderlichen Kompetenzen usw. gezogen werden könnten, eine Schimäre ist. Diese nötige Differenzierung wird

von der DPMG mit den tautologischen "Begründungen" verworfen: "sie passt nicht ins Paradigma" bzw. "bei jeder Aufgabe muss *ein* Wert stehen."

12. Hinweis auf die gründliche erkenntnis- und wissenschaftstheoretische Kritik mit zahlreichen Detailanalysen von Meyerhöfer (05): Tests im Test.

13. In 2003 wurde das (sog. analytische) Problemlösen (PL) als eigener "Kompetenzbereich" getestet. Was PL sein soll, wird i.W. tautologisch erklärt (148f). Nennen wir also "PL", was der PISA-PL-Test misst. Bei den oft notgedrungen textlastigen Aufgaben handelt es sich zumeist um "Logeleien" vom Typ des klassischen Stundenplanproblems, die sich, mit etwas anderen Einkleidungen, oft als Knobelaufgaben in Zeitungswochenendbeilagen finden. Es besteht naturgemäß eine besonders starke Korrelation zwischen der Problemlöse- und der mathematischen Kompetenz. Deutschland ist mit 513 Punkten im PL besser als in Mathematik; in welchem Fach sich also "eine mangelnde Ausschöpfung des kognitiven Potentials zum analytischen Problemlösen" zeigt (170f). Im Teilbereich ("Big Idea") "Quantität" ist mit 514 Punkten dieses Potenzial jedoch wiederum sogar "überkompensiert". — Der *unmittelbare* Vergleich der Punktzahlen ist jedoch eigentlich unzulässig (wie die PISA-Leute selbst immer wieder betonen). — Außerdem könnte Problemlösen sehr wohl ein Ergebnis schulischen Unterrichts, zumal in Mathematik sein. — Zweifelhaft ist weiterhin, ob das Lösen dieser Knobelaufgaben überhaupt eine eigene, hervorhebenswerte Kompetenz anzeigt. Müssten nicht noch ganz andere Aufgaben herangezogen werden? Welche? Und noch weiter gehend: so wenig wie Mathematical Literacy lässt sich Problemlösen in PISA prüfen. Denn den Probandinnen & Probanden (P&P) ist ja verwehrt, an einem Problem längerfristig zu arbeiten, Ansätze zu verwerfen oder nicht, es einmal eine Zeit lang liegen zu lassen, Anderen es darzustellen, von Gesprächen mit Anderen zu profitieren, Medien inklusive Internet zu nutzen, usw.

14. Die o.a. "mangelhafte Ausschöpfung des ... Potentials zum Problemlösen" im Mathematikunterricht wurde vom Hamburger Erziehungswissenschaftler Peter Struck zu einer Philippika gegen das deutsche (dreigliedrige) Schulsystem unter Verwendung bekannter Fehlinterpretationen, Missbrauchsmustern und verfälschten Daten ausgebaut, die mir in der Geballtheit vorher noch nicht begegnet sind (Frankfurter Rundschau 05.01.05).

15. Für den OECD-Bildungsstudien-Koordinator Andreas Schleicher ist die Dreigliedrigkeit eine wichtige Ursache für das langsame Wachstum des deutschen BIP in den letzten Jahren. Schneller wuchs das BIP z.B. in Spanien (mit allerdings nur 485 PISA-Mathematik-Punkten). Ein anderes "positives" Beispiel ist Mexiko mit einem einsam hohen Anteil von 24,3% Bil-

dungsausgaben an allen öffentlichen Ausgaben (mit allerdings sogar nur 385 PISA-Mathematik-Punkten) gegenüber 9,7% in Deutschland.

16. Mit den skandinavischen sog. Musterländern (außer Finnland 544) liegt Deutschland auf Augenhöhe: Island 515, Dänemark 514, Schweden 509, Norwegen 495. Besonders mit Schweden wegen angeblich ähnlicher gesellschaftlicher Struktur inklusive Migrations-Anteil wird gern der Vergleich gezogen (tatsächlich ist die Struktur in Schweden günstiger). Rechnet man die Jugendlichen mit Migrationshintergrund (MH) heraus, so ist Deutschland mit 527 allerdings sogar deutlich besser als Schweden mit 518. Auch das sagt nichts über die Leistungsfähigkeit der Schulsysteme.

17. Alle Länder mit früh gegliedertem Schulsystem (Belgien, Österreich, Schweiz, Slowakei, Tschechien) befinden sich in der oberen Hälfte der Länder-Rangliste. Alle Länder, die schlechter als Deutschland abgeschnitten haben, verfügen über das System der Einheitsschule.

18. Die PISA-Spitzenregion ist Ostasien mit Hongkong 550, Südkorea 542, Japan 534, Macau 527 und Singapur (nur bei TIMSS und IGLU, dort aber Spitze) mit leistungsorientierten Gesellschaften, aber auch Migrationsquoten von fast 0% (wie auch Finnland) oder MH-Jugendlichen, die den PISA-Schnitt z.T. sogar heben (so auch Kanada, Australien, Neuseeland).

19. Der Unterschied der PISA-Mathematik-Zahlen bei der Auswertung mit und ohne MH-Jugendliche ist nirgends so hoch wie in Deutschland mit 24. Er ist ein Indikator für die mangelhafte Integration vieler MH-Familien und verweist auf eine gewaltige gesellschaftliche Aufgabe, die nicht allein vom Bildungssystem geschultert werden kann. — Ähnlich sieht es in vielen älteren EU-Ländern und in den USA aus. — PISA vernachlässigt den Einfluss des Faktors "MH" systematisch zugunsten des Faktors "sozialer Status".

20. Die Einteilung der Mathematikaufgaben in 4 Teilbereiche ("Big Ideas") mit separater Auswertung ist weniger eindeutig als wohl unterstellt. "Raum & Form" spielt fast überall eine Rolle, z.B. schon beim Umgang mit Grafen. Auch "Quantität" und "Veränderung & Beziehungen" (V&B) sind bei vielen Aufgaben dabei. — Unter "Unsicherheit" wird systematisch auch "Beschreibende Statistik" subsumiert, obwohl diese ( $\approx$  "Daten") epistemo- und psychologisch zu "V&B" gehört. Kompromiss von PISA: Von den beiden strukturgleichen Aufgaben (bei beiden "muss eine Graphik interpretiert werden") "Größer werden 2" und "Raubüberfälle" wird eine (zutreffend) bei "V&B", die andere (fragwürdig) bei "Unsicherheit" eingeordnet (54f).

Dieser Beitrag ist eine stark gekürzte Fassung des Vortrags, der sich, unter Berücksichtigung von 1. bis 9., auf 10. bis 20. konzentrierte. Eine ausführlichere Version des Vortrags unter: [math.www.uni-paderborn.de/~bender/](http://math.www.uni-paderborn.de/~bender/)