

Peter Bender
Universität Paderborn EIM-Fak
bender@upb.de
08.03.2005

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

Dies ist eine ausführliche Version eines Vortrags gleichen Namens, den ich auf der Bundestagung der Gesellschaft für Didaktik der Mathematik (GDM) in Bielefeld am 01.03.2005 gehalten habe. Dieser Vortrag ist quasi die Fortsetzung einiger Veröffentlichungen aus den letzten beiden Jahren (s. Literatur-Verzeichnis). Er enthält deshalb viele mir wichtige Gesichtspunkte, die ich schon früher behandelt habe, *nicht*.

0. Vorbemerkungen

Fast durchweg geht es um den jeweils mathematischen Teil der internationale Vergleichs-Untersuchungen.

Man kann und muss TIMSS, PISA und IGLU ganz grundsätzlich erkenntnis- und wissenschaftstheoretisch kritisieren, wie dies insbesondere Meyerhöfer (z.B. 2005), Eva Jablonka oder Jahnke (2005) tun. Man kann und muss sich die einzelnen Aufgaben vornehmen und die vielen Mängel aufspießen wie z.B. Hagemester (1999) für die Naturwissenschaften, u.v.a. Auch bei Test-Design und -Durchführung gibt es zahlreiche Probleme. Besonders wichtig sind die gesellschafts-, wirtschafts- und bildungs-politischen Bedingungen, Implikationen usw., die in den Berichten und Analysen angesprochen werden, und diejenigen, die nicht angesprochen werden.

Ich habe alle diese Gesichtspunkte immer auch in den Blick genommen, einen Schwerpunkt aber darauf gelegt, hart an den PISA-Daten und -Konstrukten selbst zu argumentieren, weil ich meine, dass dabei niemand mehr mit dem Argument ausweichen kann, man ginge nicht wirklich auf PISA usw. ein. Dabei mache ich wohl oder übel eigentlich unzulässige Vergleiche von Punktzahlen u.ä. und Einordnungen in eigentlich ungeeignete bzw. schlecht begründete Kategorien-Systeme mit. Es zeigt sich, dass sich bereits da oft ganz andere Folgerungen ergeben und Interpretationen aufdrängen, als von interessierter Seite in Medien, Politik und Wissenschaft (!) verlaublich. Die größten Fehler und Missbräuche stammen weniger von den PISA- (TIMSS-, IGLU-) Leuten. Allerdings sind "diese" mit ihrer gesamten Öffentlichkeits-Politik nicht ganz unschuldig, und wenn es nur deswegen ist, dass sie den vielen unzulässigen Schlüssen nicht energisch genug (i.d.R. nämlich gar nicht) entgegengetreten.

Die PISA-Leute wiederum stellen keinen monolithischen Block dar; vielmehr sind an PISA unterschiedliche Institutionen und Gruppierungen mit teilweise divergenten Interessen beteiligt. Da ist zunächst die OECD als Auftrag-Geberin mit ihrer reduktionistischen ökonomistischen Denkweise, die PISA nicht zuletzt den dezidierten Länder-Vergleich aufgedrückt hat. Das internationale PISA-Konsortium hat zwar den Charakter der Untersuchung im Großen und im Kleinen mit geprägt, ist aber öffentlich weniger in Erscheinung getreten. Ganz anders das deutsche Konsortium, beim ersten Durchgang von Jürgen Baumert (der auch schon bei TIMSS-Deutschland federführend war) und beim zweiten Durchgang von Manfred Prenzel geleitet, mit seinen Bemühungen in den Berichten und in der Öffentlichkeits-Arbeit, den Ergebnissen und Interpretationen (!) den Anstrich von Objektivität und Bedeutungs-Schwere mitzugeben. Schließlich (für mich als Mathematik-Didaktiker relevant) die PISA-Mathematik-Deutschland-Gruppe (PMDG), die sich bemüht, innerhalb der Setzungen durch die drei übergeordneten Institutionen den mathematik-didaktischen Belangen zur Geltung zu verhelfen und sie mit den test-theoretischen und -praktischen Erfordernissen zu vereinbaren, und das meistens unter hohem Zeitdruck. Zugleich beziehen einzelne Aktive der PMDG Kritik an irgendeiner Facette des gesamten PISA-Unternehmens gern auf sich selbst und reagieren empfindlich (s. die treffende Beschreibung von Jahnke, 2005).

Ich war vorübergehend von der Publizität angetan, die Bildung, insbesondere mathematische Bildung, im Gefolge von TIMSS, PISA und IGLU in Deutschland und anderswo erfahren hat. Ich wer-

de mir aber in dieser positiven Einschätzung zunehmend unsicher, weil ich immer wieder erlebe, wie es bei den Schlüssen und Konsequenzen, die "die" Politik unter tätiger Mithilfe von (meistens durchaus wohlmeinenden) Kolleginnen und Kollegen aus Schule, Seminar, Hochschule, Behörden usw. zieht, nicht um Bildung, sondern bestenfalls um Leistung, oft um Wahlkampf-Parolen, Haushalts-Umschichtungen, Bedienung von Ideologien usw. geht.

1. Gesellschafts-, wirtschafts- und bildungs-politische Gesichtspunkte

1.1 Die Veröffentlichung der Statistiken von PISA 2003 am 07.12.2004

Natürlich erbrachte die Veröffentlichung der Statistik des zweiten PISA-Durchgangs am 07.12.04 zahlreiche mehr oder weniger neue Ergebnisse (P2, O2). Aber gegenüber dem ersten Durchgang war der Erkenntnis-Fortschritt nicht allzu stark ausgeprägt. Aus ausländischer Sicht reagierten Medien und Bildungs-Politik in Deutschland weniger hysterisch als in den drei Jahren bis dahin. Dafür gab es zwei gute Gründe: Zum einen schnitten die Deutschen in allen drei Bereichen (Mathematik, Lesen, Naturwissenschaften) deutlich besser ab; in Mathematik z.B. erzielten sie 503 Punkte (P2, 70) nach 490 (P1, 174). Zum anderen standen in allen Medien direkt gelassener Meinungen bereit, in der Süddeutschen Zeitung z.B. Meyerhöfer (07.12.04). Mit Recht begründeten die Kultus-Ministerien die nach wie vor mittelmäßigen Werte damit, dass sie in den 1 1/2 Jahren von der Veröffentlichung der Statistik von PISA 2000 bis zur Durchführung von PISA 2003 gar keine Gelegenheit für wirklich durchgreifende Veränderungen hatten. Dagegen wurden von der PMDG fach-didaktische Reaktionen auf TIMSS wie das SINUS-Programm, die ja schon seit einigen Jahren umgesetzt werden, als Erklärungs-Möglichkeit für die Steigerung von 13 Punkten angeboten. Vielleicht auch gewöhnen sich unsere Lehrkräfte und indirekt unsere Jugendlichen allmählich an standardisierte Tests, wie sie ja in anderen Ländern schon lange gang und gäbe sind.

Fast alle Diskussions-Beteiligten, inklusive der PISA-Leute, sind dieser verführerischen Suggestion des Vergleichs der Zahlen zweier verschiedener Untersuchungen erlegen, obwohl *vor* echten Leistungs-Änderungen viele andere Faktoren im Zuge der Test-Durchführung Einfluss ausüben, wie im PISA-Bericht mit Recht betont wird (P2, 38ff) und wie sich immer wieder an starken Punkte-Schwankungen bei zahlreichen Ländern zeigt. — Die Zahlen sagen doch nur etwas über die relative Lage der einzelnen Länder im Vergleich zur gesamten Population aller teilnehmenden und berücksichtigten OECD-Länder bei jeweils *einer* Untersuchung aus. Da braucht nur ein leistungsschwaches bevölkerungs-starkes Land wie die Türkei (mit 423 Punkten) dazuzukommen und ein leistungs-starkes bevölkerungs-starkes Land wie Großbritannien (mit 529 Punkten beim ersten Mal) wegen ungesicherter Repräsentativität wegzufallen, und schon haben die anderen Länder einen höheren Wert, ohne dass sich an den Leistungen irgendetwas geändert hätte. Zwar kann man dies auffangen, indem man für einen der beiden Durchgänge die Eichung an die des anderen auf 500 als Mittelwert und 100 als Standard-Abweichung anpasst; aber dazu müssen wenigstens die Aufgaben übereinstimmen, und das ist ja in Mathematik zwischen 2000 und 2003 nicht gegeben.

1.2 Der Artikel "Nicht für die Schule, sondern für das Leben lernen" von Peter Struck

Jedenfalls meinte ich, nach 2004 nicht noch einmal auf einer GDM-Bundestagung zu diesem Thema vortragen zu müssen, bis mir in der Frankfurter Rundschau am 05.01.05 der Artikel "Nicht für die Schule, sondern für das Leben lernen" unterkam, in dem bekannte Fehl-Interpretationen, Missbrauchs-Muster und ideologische Vorurteile zu einer Argumentations-Kette zusammengebastelt wurden, bei der sinngemäß herauskam: "PISA 2003 hat gezeigt, dass das deutsche Schul-System, insbesondere seine Mehr-Gliedrigkeit, schlecht ist und u.a. durch die Einheits-Schule ersetzt werden muss."

Es war nämlich bei PISA 2003 neben Mathematik, Lesen und Naturwissenschaften ein vierter Bereich, das Problemlösen, untersucht worden. Dort hatte Deutschland 513 Punkte erzielt (P2, 157), also 10 mehr als in Mathematik. Dieser Umstand war bereits im PISA-Bericht dem deutschen Schul-System als gewisser Mangel angekreidet worden. Das Problemlösen ist nämlich keinem Schulfach zugeordnet, die Kompetenz dazu wird daher angeblich nicht in der Schule erworben, die deutsche Schule schöpft also das Potenzial der Jugendlichen ungenügend aus. Während der

PISA-Bericht noch einer gemäßigten Diktion frönt, geht der Rundschau-Autor in die Vollen: Da werden die Straßenkinder in Mittel-Amerika und Rumänien (nicht mehr jedoch die in den PISA-Ländern Brasilien und Mexiko) als Beleg für die Leistungsfähigkeit des realen Lebens als Lehrmeister gegenüber der Schule bemüht, da wird "Menschen in unserem Land", die "weniger ... von Kindern, Schulen und Lernen verstehen", unterstellt, dass sie "für ein ... Schulsystem ... mit angstmachenden Unterrichtsvollzugsbeamten ... plädieren", u.v.a.m. Da werden Entwicklungs-Linien schon ab TIMSS gezogen, was noch falscher ist als zwischen zwei PISA-Durchgängen. Darüber hinaus werden eine Menge Fakten verfälscht, von denen ich einige hier nenne und gleich richtig stelle:

- Das Einheitsschul-Musterland Schweden ist bei TIMSS nicht "auf Platz 1 gelandet", sondern in Mathematik auf dem 6. und in Naturwissenschaften auf dem 4. Platz (T2, 90, 98). Allerdings waren die damals teilnehmenden schwedischen Jugendlichen durchschnittlich ein halbes Jahr älter als die anderen, und rechnet man diese Bevorzugung à la TIMSS und PISA heraus, dann lag Schweden bereits bei TIMSS gerade mal noch auf einem guten Mittelplatz.
- Die Jungen können in der Schule nicht nur *doch* "mit den Mädchen mithalten" (P2, 213); in Mathematik sind sie überall auf der Welt außer in Island besser, meistens signifikant, insbesondere auch in Deutschland. In Naturwissenschaften haben die Jungen einen (wegen Biologie) zwar geringeren, aber ebenfalls weltweit noch signifikanten Vorsprung. Lediglich in Lesen werden sie deutlich von den Mädchen ausgestochen.
- Deutschland hat nicht "die größten Leistungsbandbreiten weltweit", sondern in Mathematik Belgien und die Türkei (P2, 70), in Lesen Uruguay, Brasilien und Belgien (P2, 99) sowie in Naturwissenschaften Frankreich und Deutschland (P2, 118). In allen drei Inhalts-Bereichen sind die Leistungs-Bandbreiten zahlreicher weiterer Länder etwa so groß wie die unseren.
- Dass "die deutlichsten Fortschritte in Sachen Schule in Finnland, in den drei baltischen Staaten, in Polen, in Bulgarien und in Tschechien zu verzeichnen sind", ist jedenfalls durch TIMSS und PISA nicht zu belegen. In Mathematik und Naturwissenschaften war es vielmehr so: Finnland "stagniert" seit TIMSS auf hohem Niveau. Für Estland gibt es überhaupt keine Zahlen. Für Litauen und Bulgarien existiert lediglich *ein* unsicherer Wert bei TIMSS. Tschechien hat von TIMSS 1994 bis PISA 2000 einen gewaltigen Rückschritt erlitten und sich von diesem bis PISA 2003 wieder leicht erholt. Lettland hat eine vergleichbare Bewegung auf erheblich niedrigerem Niveau durchgemacht. Bei Polen dürfte es ähnlich gewesen sein (war allerdings bei TIMSS nicht beteiligt). Bei Deutschland war es wie bei Tschechien und Lettland; es lag mit seinen Punktzahlen und deren Trend immer zwischen diesen beiden Ländern. Ich möchte nochmals darauf hinweisen, dass diese Vergleiche eigentlich sinnlos sind und ich sie nur mache, um das Tendenziöse an der Argumentation aufzuzeigen: Ein und dieselbe Entwicklung wird bei Tschechien, Polen sowie Lettland positiv und bei Deutschland negativ bewertet.

Mathematik	TIMSS	PISA 2000	PISA 2003
Tschechien	564	498	516
Deutschland	509	490	503
Polen		470	490
Lettland	493	463	483

Wir lesen hier nicht die Bild-Zeitung, sondern, wie gesagt, die Frankfurter Rundschau. Außerdem stammt dieses Machwerk nicht von einer Lain oder einem Laien, sondern von einem ausgewiesenen Experten für Pädagogik, dem Hamburger Erziehungswissenschaftler Peter Struck. Es geht mir hier nicht um die Person "Struck", von der ich schon durchaus Ansprechendes gelesen habe, sondern um diese Begründungs-Versatzstücke, mit denen aus den PISA-Zahlen allerlei absurde Folgerungen gezogen werden und die ich in dieser Geballtheit noch nie erlebt habe.

1.3 Die Spitzen-Region ist nicht Skandinavien, sondern Ost-Asien

Die skandinavischen sog. Muster-Länder (außer Finnland mit 544) befinden sich etwa in Augenhöhe mit Deutschland (Island 515, Dänemark 514, Schweden 509, Norwegen 495). Die wahre Spitzen-Region ist Ost-Asien mit Hong-Kong (550), Süd-Korea (542), Japan (534), Macau (527) und Singapur (Spitze bei TIMSS (T2, 90) und IGLU (I1, 210); bei PISA nicht dabei gewesen). Diese

Staaten haben zwar alle das Einheitsschul-System; aber sie würden selbstverständlich diese guten Platzierungen auch mit einem gegliederten Schul-System erreichen. Denn diese Leistungen beruhen auf weit wichtigeren Faktoren, nämlich der Leistungs-Orientiertheit der dortigen Gesellschaften sowie der Hochschätzung von Schul-Bildung, zumal in Mathematik. — Man müsste auch einmal intensiver prüfen, wie ausgeprägt jeweils das System von Privat-Schulen ist, mit dem auch dort ja doch wieder differenziert wird.

Der mit der Leistungs-Orientierung verbundene Leistungs-Druck entspricht natürlich nicht den idyllischen Vorstellungen westlicher Pädagogik, und deswegen wird der Vergleich nicht so gern mit den ost-asiatischen Tiger-Staaten gezogen. Hinzu kommt, dass dort sowie in Kanada, Australien und Neuseeland die Migrations-Struktur hohe PISA-Punktzahlen eher zulässt als in den älteren EU-Ländern oder in den USA: Entweder ist die Migrations-Quote fast 0, oder haben die Migrations-Familien im Durchschnitt ein relativ hohes soziales und kulturelles Niveau, so dass in einigen dieser Länder einige der Ergebnisse durch die Jugendlichen mit Migrations-Hintergrund (MH) z.T. sogar verbessert werden (z.B. 11, 296).

1.4 Die Rolle des Migrations-Hintergrunds

In den PISA-Berichten wird der Faktor "MH" zwar ausführlich diskutiert, aber nach meinem Dafürhalten systematisch unterschätzt, und zwar zugunsten des Faktors "sozialer Status". So wurde in einer Regressions-Analyse zur Abhängigkeit der Mathematik-Punkte von 8 verschiedenen Einfluss-Faktoren (die schrittweise nacheinander einbezogen wurden) deren Anteil an der aufgeklärten Varianz zu 67% für den sozialen Status, zu 12% für den MH und zu 21% für die anderen (in dieser Reihenfolge: Kulturelle Besitztümer, Kindergarten-Besuch, Bildungs-Abschluss, Familien-Umgangssprache, Vater-Erwerbstätigkeit, Familien-Struktur) bestimmt (P2, 274). Dass in Deutschland und in den älteren EU-Ländern alle diese Faktoren, zumal "sozialer Status", ihrerseits stark vom Faktor "MH" abhängen und ihre Aufklärungs-Anteile wieder partiell diesem zugeschlagen werden müssten, wird hierbei nicht deutlich.

Zur Stützung der These von der Überlegenheit der Gesamtschule wird Deutschland gern mit Schweden verglichen, u.a. wegen der ähnlichen Bevölkerungs-Struktur und der vergleichbaren Migrations-Quote. Einen deutlichen Unterschied gibt es aber doch. Während in Deutschland unter allen Jugendlichen mit MH 75% *doppelten* MH haben, beträgt in Schweden dieser Anteil nur gut die Hälfte (P2, 257); und diese sind es ja, die besondere Schwierigkeiten haben. Rechnet man einmal sämtliche Jugendlichen mit MH (also auch die leistungs-stärkeren) heraus, liegt Deutschland mit 527 Punkten auf einmal merklich *vor* Schweden mit 518 Punkten (P2, 257), und nicht mehr allzu weit hinter Finnland mit seinen 544 Punkten und einer Migrations-Quote bei 0. — Wer jetzt hier herauslesen möchte (was, wie gesagt, eigentlich unzulässig und missbräuchlich ist), dass das schwedische Schul-System besser für die Bewältigung der Migrations-Problematik ist, muss zugleich mitlesen, dass es die autochthone Bevölkerung benachteiligt.

Ich meine, dass diese Differenz von 24 Punkten in Deutschland, die auf der Welt sonst nirgends so hoch ist, unabhängig von dem Vergleich mit Schweden ein Indikator für die bei uns besonders ausgeprägte Migrations-Problematik ist. — Um nicht falsch verstanden zu werden: Wenn man sich überhaupt auf das Zahlenwerk von PISA einlässt, dann ist 503 genau die richtige Punktzahl; denn diese steht für die Leistungen, die die 15-jährigen Jugendlichen in Deutschland erbracht haben. Diese Zahl in Verbindung mit der großen Leistungs-Bandbreite stößt uns schmerzlich auf die Fehler und Versäumnisse unserer Migrations-Politik seit langem, die inzwischen in eine völlig unzulängliche Integration weiter Teile unter den Migrations-Familien gemündet ist. Hier wurde eine gewaltige Aufgabe für die ganze Gesellschaft, und nicht allein für das Bildungs-System, aufgetürmt, und ein Verdienst von PISA ist es, sie uns vor Augen geführt zu haben.

1.5 PISA zeigt nicht die Überlegenheit der Einheits-Schule

In den Berichten zu TIMSS und PISA wurde immer betont, dass sich aus diesen Studien keine Schlüsse auf die Überlegenheit eines Schul-Systems ziehen lassen (T2, 18f u.a.), nicht zuletzt wohl auch ein wenig zum Schutz der deutschen Gesamtschule, die ja zur großen Verblüffung von

mir und vielen anderen noch deutlich schlechter als die Realschule abgeschnitten hatte (P1.3, 273). Eine entscheidende Ursache für diese Schwäche ist, in der Tat unbestreitbar, die bloße Existenz des Gymnasiums in Deutschland. — Mit dieser Erklärung meint man, wenn man *für* die Gesamtschule plädieren möchte, dieses Ergebnis getrost ignorieren und sich nach passenderen Zahlen umschauen zu können: Im internationalen Vergleich gibt es zahlreiche Länder mit Einheits-Schule und mehr Punkten als Deutschland, und schon hat man scheinbar die Überlegenheit der Gesamtschule mit PISA-Zahlen belegt. — Dass auch sämtliche Länder mit weniger Punkten als Deutschland über die Einheits-Schule verfügen, wird da geflissentlich übersehen. Mehr noch: Alle Länder mit früh gegliedertem Schul-System, nämlich Belgien, Österreich, Schweiz, Slowakei, Tschechien (P1.1, 425, P2, 293), befinden sich in der oberen Hälfte bei PISA-2003-Mathematik (Deutschland ist unter diesen quasi das Schlusslicht), so dass sogar die PISA-Länder-Rangliste eher noch zugunsten des gegliederten Schul-Systems spricht.

1.6 Die OECD-Wirtschaftsdaten sagen nichts über Bildungs-Qualität aus

Die Diskussion um die Einheits-Schule könnte je nach der politischen Entwicklung in Schleswig-Holstein demnächst Auftrieb erhalten. Ein eifriger Kämpfer sitzt an einflussreicher Stelle bei der OECD, der (wohl zufällig deutsche) Koordinator für Bildungsfragen, speziell auch PISA, Andreas Schleicher. Dieser hat vor kurzem sehr publikums-wirksam eine OECD-Studie zu den nationalen Bildungs-Ausgaben vorgelegt, bei der — natürlich — Deutschland nicht gut abgeschnitten hat (Frankfurter Rundschau 15.09.04). Beim Anteil der Bildungs-Ausgaben an den öffentlichen Gesamt-Ausgaben liegt Mexiko mit einsamen 24,3% an der Spitze, gefolgt von Süd-Korea mit 17,7%. Der OECD-Durchschnitt beträgt 12,7% und die deutsche Quote 9,7%. Die Daten stammen freilich von 2001, wo die deutsche Bildungs-Politik ja noch nicht auf die PISA-Ergebnisse hatte reagieren können. Ob die deutsche Quote seitdem gestiegen ist, muss allerdings bezweifelt werden. Die öffentlichen Haushalte geben eine solche Steigerung jedenfalls nicht her; der Bildungs-Abbau am Gymnasium durch die Verkürzung der Schulzeit in vielen Bundesländern, die Einführung von Bachelor-Studiengängen, die Erhebung von Studien-Gebühren usw. werden eher zu einer weiteren Reduzierung der öffentlichen Bildungs-Ausgaben führen.

Das Beispiel "Mexiko" (mit seinem letzten Platz unter den OECD-Ländern bei PISA 2003 mit 385 Punkten in Mathematik) zeigt aber drastisch die zweifelhafte Aussage-Kraft dieses Zahlen-Materials. Was sind denn öffentliche Ausgaben in Mexiko und in Deutschland, und wie hoch sind sie? Was sind Bildungs-Ausgaben? Bedeutet die Rang-Liste, dass in Mexiko die öffentlichen Gesamt-Ausgaben besonders niedrig und in Deutschland besonders hoch sind? — Jedenfalls wurde wieder einmal das "dem Ausland kaum vermittelbare" drei-gliedrige deutsche Schul-System als eine Wurzel des Übels hingestellt. — Welchen Übels eigentlich? — Z.B., dass das Wachstum des BIP in Deutschland in den letzten zehn Jahren erheblich geringer als etwa in Spanien (mit allerdings nur 485 PISA-Mathematik-Punkten) ausgefallen ist. — Typisch für solche ökonomistische Betrachtungs-Weisen ist die Konzentration auf die Änderungs-Rate und die Vernachlässigung des absoluten Bestands. Weiterhin sind auch hier, wie so oft, die entscheidenden (ökonomischen) Einfluss-Faktoren ausgeblendet wie Eingliederung der DDR ab 1990, hohe Migrations-Quote in den 1990-er Jahren, Ablösung der D-Mark durch den Euro, die Rollen als Netto-Zahler und Netto-Empfänger in der EU, usw. usf. — Als dann noch die Kritik nachgeschoben wurde, dass die deutschen Lehrerinnen & Lehrer (L&L) zu alt seien, war der Bogen aber anscheinend überspannt und die Diskussion verlief sich. Das Dilemma ist doch offensichtlich: Man muss gewärtigen, dass das Pensions-Alter auf ≥ 67 Jahre heraufgesetzt wird, und das passt nicht zusammen mit dem von der OECD mit vertretenen kapitalistischen Grundsatz, dass man ab 45, 50 Jahren zum alten Eisen gehört.

In diesem Zusammenhang erweist sich die von der deutschen Bildungs-Politik geplante Reduktion der Bildungs-Zeit i.V.m. Stress-Erhöhung für die Heranwachsenden wie für die Lehrenden (frühe Einschulung, Verkürzung der Schulzeit bis zum Abitur, Verkürzung des Studiums) als besonders kontraproduktiv. Wenn alle diese Maßnahmen wirksam werden, wird z.B. der akademische Arbeitsmarkt bei sowieso sinkendem Volumen um ca. 12% zusätzlich belastet. Und eine gediegene Bildung wäre gerade dann vonnöten, wenn immer mehr Freizeit sinnvoll gestaltet werden muss.

Man muss einer Organisation wie der OECD dankbar sein, dass sie so etwas wie PISA ermöglicht. Man darf aber deren Interessen als Vertretung von entwickelten Ländern, die mit den USA politisch verbunden sind, bzw. solcher, die auf dem Weg dahin sind, nicht aus den Augen verlieren. Diese Interessen müssen sich nicht mit denen einzelner Länder und schon gar nicht mit denen der Bevölkerung decken: Bildung ist da eben ein Wirtschafts-Faktor unter vielen. Diese Sichtweise hat wohl auch ihre Daseins-Berechtigung, für mich aber erst unter "ferner liefern".

2. Inhaltliche Gesichtspunkte

2.1 Fächer-übergreifendes Problemlösen im zweiten PISA-Durchgang

Fraglich ist für mich immer, was ein Problem im Sinne des Problemlöse-Paradigmas zu einem Problem macht, bzw. was ein Problem von einem Nicht-Problem unterscheidet. Im angelsächsischen Sprachraum z.B., wo das Konstrukt "problem solving" besonders gepflegt wird, werden dieselben Wörter auf das Bearbeiten irgendwelcher (z.B. Mathematik-) Aufgaben gemünzt. Die PISA-Definition, nach der analytisches Problemlösen hauptsächlich "in der Analyse gegebener oder erschließbarer Informationen und dem Entwickeln einer Lösung" einer sich aus einer "verbal, oft auch unter Nutzung von Graphiken, beschriebenen Ausgangslage ... ergebenden Problemstellung" (P2, 148f) besteht, hilft nicht weiter, weil der Begriff "Problemlösen" da i.W. durch sich selbst erklärt wird. Der Bedingung, "dass der Lösungsweg nicht unmittelbar erkennbar ist" (P2, 148), schließlich unterliegt doch jede Test-Aufgabe (englisch: "Problem") und nicht nur solche in einem abgetrennten Teil-Bereich "Problemlösen". — Ich weiß also nach wie vor nicht, was Problemlösen ist (was aber an der Sache, und nicht am PISA-Bericht liegt), und verwende im Folgenden in klassischer Weise die Arbeits-Definition "Problemlösen ist, was der PISA-Problemlöse-Test misst".

Das *dynamische* Problemlösen (das nur Teil des deutschen Ergänzungs-Tests war; P2, 162) ist schnell abgehandelt. Unter Zeit-Begrenzung müssen die Jugendlichen, vereinfacht gesagt, auf einem Computer-Bildschirm mit Reglern Grafiken verändern und mit anderen vorgegebenen Grafiken zur Deckung bringen. Die so zu testenden Kompetenzen machen nach meinem Dafürhalten keine nennenswerte Facette von Bildung aus. Der Erfolg dürfte vor allem eine Sache von Erfahrung sein, wobei 15-Jährige durchaus besser abschneiden können als vielleicht Mathematik-L&L, wie auch beim Versenden von MMS, Umgang mit Computer-Spielen, Inline-Skaten usw.

Beim *analytischen* Problemlösen sind die oft notgedrungen text-lastigen Aufgaben zumeist "Logeleien", grob gesprochen, vom Typ des klassischen (in voller Allgemeinheit übrigens nicht lösbaren) Stundenplan-Problems, z.B. "Kinobesuch" (P2, 152): Drei 15-Jährige mit bestimmten Zeit-Wünschen und -Restriktionen und weiteren Bedingungen wollen in den Ferien gemeinsam ins Kino gehen. Wann klappt es? Oder: "Bewässerung" (P2, 401), wo statt eines Stromkreises mit Schaltern ein System von Wasser-Kanälen mit Schleusen zu analysieren ist. Viele der vorgestellten Aufgaben könnten sich, mit denselben oder anderen Situations-Kontexten, als Knobel-Aufgaben in Zeitungs-Wochenendbeilagen finden.

Insgesamt erfolgte eine Auswertung wie in den drei Haupt-Fächern, hier mit einem Schwierigkeits-Leistungs-Stufen-Modell (in PISA und den anderen Studien als Kompetenzstufen-Modell bezeichnet) mit zwei begrenzten und je einer halb-offenen Stufe nach oben und unten (P2, 150f), also 2+2 Stufen.

Es besteht, wen wundert's, eine besonders starke Korrelation zwischen der Problemlöse- und der mathematischen Kompetenz (P2, 167), die Länder-Rangfolgen sind sehr ähnlich, insbesondere sind auch beim Problemlösen Süd-Korea, Hong-Kong, Finnland und Japan ganz und Macau mit vorne. Deutschland ist mit 513 Punkten deutlich besser als in Mathematik, die Niederlande dagegen z.B. ist 18 Punkte schlechter als in Mathematik (P2, 157f). "An deutschen Schulen" zeigt sich "im Hinblick auf die Entwicklung mathematischer Kompetenz eine mangelnde Ausschöpfung des kognitiven Potentials zum analytischen Problemlösen", dagegen wird "in den Niederlanden ... dieses Potential ... optimal genutzt beziehungsweise sogar überkompensiert" (P2, 170f). — Was sagt uns das?

Als Fach-Didaktiker möchte ich zunächst einmal auf die Werte der einzelnen mathematischen Inhalts-Bereiche (international: "Big Ideas"; "übergreifende Ideen") hinweisen: Quantität 514, Veränderung und Beziehungen (V&B) 507, Raum und Form 500 sowie Unsicherheit 493 (P2, 75). Beim Bereich "Quantität" hat die deutsche Schule, in der PISA-Diktion, das Potenzial also doch ausgeschöpft, sogar "überkompensiert". Bei den anderen drei Bereichen, zumal bei Geometrie und Stochastik, besteht in der Tat Intensivierungs-Bedarf, was wir allerdings schon lange wissen. — Mit der Mehr-Gliedrigkeit des Schul-Systems hat das jedenfalls nichts zu tun.

Der *unmittelbare* Vergleich der Punktzahlen in Problemlösen und in Mathematik ist natürlich unzulässig (wie die PISA-Leute selbst immer wieder betonen). Es könnte doch sein, dass, wenn man die Leistungen in den beiden Bereichen wirklich irgendwie in Bezug zueinander setzen könnte, die Deutschen in Problemlösen sogar schlechter als in Mathematik sind. Die Zahlen sagen nur etwas im jeweiligen Bereich über die Relation zu den anderen Ländern aus, mehr nicht. — Außerdem könnte dieses Problemlösen doch sehr wohl ein Ergebnis schulischen Unterrichts sein, insbesondere in Mathematik, mit seinem Anspruch einer allgemeinen Denk-Förderung. Ich will nun nicht behaupten, dass man dabei erfolgreich ist; aber ebenso wenig ist bewiesen, dass man keinen Erfolg hat. Mit derselben Apodiktik wie Struck könnte man die Ergebnisse im Problemlösen gerade als Beleg für die Qualität der deutschen Schule nehmen.

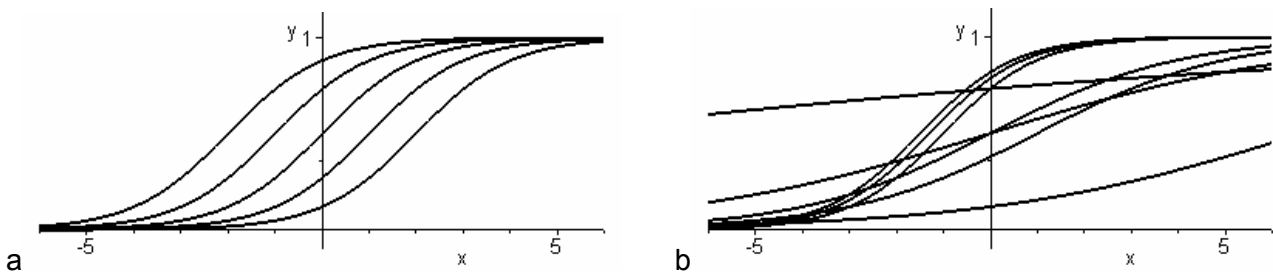
Zweifelhaft ist natürlich, ob das Lösen dieser Knobel-Aufgaben überhaupt eine eigene, hervorhebenswerte Kompetenz anzeigt. Müssten nicht noch ganz andere Aufgaben dazu herangezogen werden? Und noch weiter gehend: so wenig wie Mathematical Literacy lässt sich m.E. Problemlösen letztlich in einem Test wie PISA prüfen. Denn den Probandinnen & Probanden (P&P) ist ja verwehrt, an einem Problem längerfristig zu arbeiten, Ansätze zu verwerfen oder weiter zu verfolgen, das Problem einmal eine Zeit lang liegen zu lassen, Anderen es verständlich darzustellen, von Gesprächen mit Anderen zu profitieren, Medien inklusive Internet zu nutzen usw. — Man muss auch immer wieder darauf hinweisen, dass die Rede von Problemlöse-Potenzial sich auf den speziellen PISA-Aufgaben-Pool bezieht. Zwar wissen die Test-Leute das, und sie befeißigen sich ja lediglich einer ökonomischen Sprache; aber die beim Publikum leicht entstehenden Missverständnisse tragen zumindest dazu bei, den Eindruck von Bedeutsamkeit zu verstärken, und da fällt es schon schwer, dem auch noch selbst aktiv entgegenzuwirken. — Darüber hinaus muss ich nochmals grundsätzlich in Frage stellen, ob Problemlösen als Test-Gegenstand überhaupt sinnvoll ist. Es ist doch bei jeglicher geistigen Arbeit allgegenwärtig wie das Atmen beim Leben!

2.2 Das den Belangen der Mathematik-Didaktik fern stehende, über-idealisierte PISA-Testmodell

Die Test-Leute fechten solche inhaltliche, gar filosofische Bedenken wenig an. Die Tiefe der Kluft zwischen fach-didaktischen und test-orientierten Interessen ist mir erst im letzten Jahr so richtig bewusst geworden, als ich mich zusätzlich zu den vielen Berichten durch den Technical Report und das Database-Manual von PISA 2000 u.a. kämpfte, wobei mir am meisten die Erläuterungen von Detlef Lind in dem JMD-Artikel geholfen haben. Ich bilde mir zwar ein, jetzt grob zu wissen, wie aus einem bestimmten Antwortdaten-Satz die entsprechende PISA-Punktzahl zu errechnen ist. Aber genauer wüsste ich es, wenn es mir einmal exemplarisch vorgeführt worden wäre. Dass so etwas auf den vielen tausend Seiten nicht zu finden ist, sehe ich als didaktisches Versäumnis.

Bei aller Kompliziertheit ist das Test-Modell, das die PISA-Test-Fachleute ansetzen (müssen, damit sie überhaupt rechnen können), stark idealisiert auf Kosten von fach-didaktisch differenzierenden Gesichtspunkten. Besonders deutlich wird dies bei den Aufgaben-Charakteristiken (Lösungs-W'keit für diese Aufgabe in Abhängigkeit von der Leistungs-Stärke der P&P, gemessen in PISA-Punkten). Da wird nämlich unterstellt, dass sie alle den Funktions-Term $\frac{1}{1+\exp(c-x)}$ (Knoche u.a.

2002, 174), also alle denselben Graf (bis auf waagrechte Verschiebung) haben (s. Abb. a). Das Schaubild eines Bruchrechen-Tests von Herrn Lind (1994, 317, 282) zeigt, wie eine solche Kollektion realistischerweise auch aussehen kann (s. Abb. b). Natürlich kann man seine Test-Aufgaben so auswählen und gestalten, dass man dem idealen Bild näher kommt. Wie weit das ohne Verlust von fach-didaktischer Substanz möglich ist, steht jedoch dahin.



(Die beiden Abbildungen sind nicht Kopien der Originale, sondern diesen nur nachempfunden.)

Damit man aber die Aufgaben-Schwierigkeit auf derselben Skala wie die P&P-Leistungsfähigkeit ansiedeln kann, kommt man um diese Idealisierung nicht herum, und man muss sie für diesen Zweck, dem ich durchaus etwas abgewinnen kann, akzeptieren. Etwas irritiert hat mich an der Kalibrierung der Aufgaben-Schwierigkeit allerdings die Tatsache, dass nicht i.W. sämtliche P&P-Punktzahlen herangezogen wurden (wie ich mir das ursprünglich vorgestellt hatte), sondern aus jedem OECD-Land gleichgewichtig je 500 P&P (d.h. etwa ein Zehntel) (P1.1, 51, 520f, O1.2, 105). Hierbei haben also die Jugendlichen aus Island dasselbe Gewicht wie die aus den USA, obwohl diese für etwa 900-mal so viele stehen (O1.2, 135f), und z.B. die aus dem Nicht-OECD-Staat Brasilien haben das Gewicht 0. Ich nehme an, dass sich die resultierenden Über- und Unterschätzungen, jedenfalls innerhalb der OECD, ungefähr ausgleichen. Vermutlich trifft diese Ausgleichs-Annahme auch auf die Verfälschung der "wahren" Aufgaben-Charakteristiken durch den einheitlichen logistischen Ansatz zu.

2.3 Das fragwürdige Stufen-Modell (bei PISA, TIMSS und IGLU, bei "Lesen", "Mathematik", "Naturwissenschaften", "Problemlösen" sowie Teil-Bereichen davon)

Allerdings zieht die Einheitlichkeit der Aufgaben-Charakteristiken die gleiche Breite aller (begrenzten) Schwierigkeits-Leistungs-Stufen nach sich, und das kommt mir verdächtig vor und veranlasst mich, diese Idealisierung als zu stark in Frage zu stellen. Aus äußerlichen Handhabbarkeits-Gründen hat nämlich das internationale PISA-Konsortium die Schwierigkeits-Leistungs-Skala in Stufen eingeteilt ("dividing ... into levels though useful, ... [is] essentially arbitrary", O1.2, 197). Da wurde ein bestimmter Punkte-Wert als Untergrenze von Stufe I ausgewählt, und von da aus nach oben vier Stufen folgendermaßen festgelegt: die Obergrenze einer Stufe ist dadurch bestimmt, dass die P&P, die auf der Untergrenze dieser Stufe angesiedelt sind, 50% und alle P&P auf dieser Stufe im Durchschnitt 62% aller Aufgabe auf dieser Stufe lösen. Diese Obergrenze ist zugleich Untergrenze der nächsten Stufe. Außer den vier begrenzten gibt es noch je eine nach oben und nach unten offene Stufe (letztere von mir "Stufe 0" genannt), insgesamt also 4+2 mit den Grenzen 329, 421, 512, 604, 696. Weil PISA-2000-international nur 31 Mathematik-Aufgaben enthält, hat man auf eine inhaltliche Fassung dieser Stufen-Einteilung verzichtet (P1.1, 159) und eine solche den nationalen Gruppen überlassen. Es wäre einmal interessant zu prüfen, ob die 30 bzw. 40 anderen Länder eigentlich zum selben Modell wie die Deutschen gekommen sind.

Solche Schwierigkeits-Leistungs-Stufungen wurden für sämtliche TIMSS-, PISA- und IGLU-Studien aufgestellt, und es wurde versucht, sie inhaltlich zu füllen. Aber sobald mehr ausgesagt wird, als dass die Aufgaben mit zunehmender Punktzahl immer schwerer werden, tauchen immer "widerspenstige" Aufgaben auf, die mit ihrer tatsächlichen Schwierigkeits-Punktzahl in einer offensichtlich falschen Stufe landen. Oft bleiben die inhaltlichen Beschreibungen trivial. Z.B. ist in der deutsch-sprachigen internationalen Mathematik-Skala bei PISA 2000 die Stufen-Abfolge durch zunehmende Leistungen beim Modellieren (inkonsistent von "nicht" über "elementar", "SI-Niveau", "umfangreich" bis "komplex") mit nicht-konsistenten und nichtssagenden Zusätzen ("begriffliches Verknüpfen", "anspruchsvolle Begriffe", "innermathematisches Argumentieren") geprägt (P1.1, 168). Diese von mir wörtlich zitierten Schlagwörter stellen Kurzfassungen aus dem PISA-Bericht selbst dar. Die ausführlicheren Beschreibungen (P1.1, 160) wiederum sind nicht genügend operational und führen bei der Einordnung von konkreten Aufgaben erst recht zu Inkonsistenzen. Da ist viel die Rede von "Modellbildung", und genau diese wird in unserer Vergleichsuntersuchungen-Kommunität insgesamt so inflationär gebraucht, dass sie banal wird. Wie vom PISA-Ableger IGLU explizit eingeräumt (I2, 118f), wird sie nämlich mit jeglichem Bearbeiten von Aufgaben identifiziert.

Die PMDG hat die gesamte Problematik früh erkannt. Für sie ist sie deswegen besonders relevant, weil sie mit dem Schwierigkeits-Leistungs-Modell ehrgeizigere Pläne hat: Es soll ermöglichen, dass man für jede Aufgabe aufgrund einer fach-didaktischen Analyse der erforderlichen Kompetenzen und weiterer Faktoren voraussagen kann, auf welcher Stufe sie landet. Damit sollen die Basis für wissenschaftliche, klare Vorgaben für die Konstruktion von schul-mathematischen Tests gelegt werden, dieser Zweig der Mathematik-Didaktik vom Ruch der Beliebigkeit und des Laientums befreit werden und nicht zuletzt Aufgaben-Kollektionen entstehen, die von vorneherein vielleicht sogar dem vorhin beschriebenen Ideal-Bild mit breit gestreuten einheitlichen Aufgaben-Charakteristiken besser entsprechen.

2.3.1 Die fragwürdige zusätzliche Kategorie "Typen mathematischen Arbeitens"

So hat die PMDG als zusätzliche Kategorie die "Art" bzw. den "Typ mathematischen Arbeitens" mit "den" drei Ausprägungen "technische Aufgaben", "rechnerische" und "begriffliche Modellierungs- und Problemlöse-Aufgaben" (anscheinend unterschwellig mit wachsendem Anspruchs-Grad) eingeführt (z.B. Neubrand 2004, 88ff). Im Gegensatz zur Kategorie "PISA-Punkte-Stufen" ist diese nun weich mit unklarer Begrifflichkeit, fließenden Übergängen, weiten Überschneidungs-Bereichen und vermutlich großen Lücken. Versucht man dennoch, Aufgaben einzusortieren, so wird man oft feststellen, dass verschiedene Aufgaben-Teile zu verschiedenen Ausprägungen gehören.

Ein Beispiel für die Problematik ist die "31-Pfennig-Aufgabe" (797 Punkte "schwierig"): "Wie kannst du einen Geldbetrag von genau 31 Pfennig hinlegen, wenn du nur 10-Pfennig-, 5-Pfennig- und 2-Pfennig-Münzen zur Verfügung hast? Gib *alle* Möglichkeiten an." — Sie wird unter "begriffliches Modellieren und Problemlösen" eingeordnet (Neubrand 2004, 89). Ich kann diese Einordnung nicht nachvollziehen. Es handelt sich doch um eine begrifflich völlig anspruchlose Abzähl-Aufgabe an vorgestellten oder aufgezeichneten konkreten Objekten, und die Schwierigkeit liegt in der Erfassung aller Fälle. Es stellt sich die Frage, ob z.B. die erwähnte Kategorie "Übergreifende Ideen" nicht viel relevanter ist und die Schwierigkeit so mancher Aufgabe daher rührt, dass es sich um ungewohnte Kombinatorik, unbewältigte Raum-Geometrie oder unverstandene Stochastik handelt.

Unter den kognitions-bezogenen Ansätzen erscheint mir der von Cohors-Fresenborg, Sjuts & Sommer (Neubrand 2004, 109ff) noch am fundiertesten. Wie weit er zum Schwierigkeits-Leistungs-Stufen-Modell passt, müsste sich noch erweisen; einen direkten Niederschlag konnte ich kaum ausmachen (in P2 nur bei einem Beispiel: 61).

2.3.2 Die zusätzliche Kategorie "Übergreifende Ideen" (\approx Teil-Bereiche der Mathematik)

In PISA 2003 war ja Mathematik der Schwerpunkt und mit 84 (statt 31) Aufgaben vertreten (P2, 51). Deswegen wurde dieses Fach noch einmal zerlegt, allerdings nicht etwa in die Typen mathematischen Arbeitens (die tauchen im internationalen Bericht gar nicht auf), sondern in die vier "Big Ideas" mit vier eigenen Länder-Rangfolgen, wenn auch (ohne inhaltliche Begründung) mit gemeinsamer Stufung (erst zwei sind veröffentlicht). Es gibt jetzt eine Stufe mehr, also 5+2 statt 4+2, was natürlich die Willkür dieser Stufung unterstreicht und die Bemühungen um eine Kanonisierung auch PISA-endogen konterkariert.

So relevant mir diese Unterteilung in "Big Ideas" erscheint, so unscharf ist auch sie, und in der Mathematik-Didaktik weiß man dies schon lange. "Raum und Form" spielt fast überall eine Rolle, nämlich bereits sobald es um Anfertigen oder Interpretieren eines Funktions-Grafs geht. Vergleichbares gilt sowohl für "Quantität", als auch für "V&B". — "Unsicherheit" dagegen ist nicht so allgegenwärtig. Allerdings geht es bei den gern gestellten Aufgaben aus der Beschreibenden Statistik oft gar nicht um Unsicherheit, und aus diesem Grunde wurde ja (in P2, 49) die Überschrift "Daten und Zufall" vorgeschlagen (aber später im Bericht doch nicht realisiert). Diese beiden Ideen wiederum werden zwar aus mathematik-systematischen Gründen üblicherweise gemeinsam behandelt; der epistemologische und psychologische Umgang mit ihnen ist jedoch völlig unterschiedlich, und das Thema "Daten" gehört m.E. in die Bereiche "Quantität" bzw. "V&B". Die PISA-Leute scheinen das auch zu ahnen, und sie haben bei den beiden völlig struktur-gleichen Statistik-Aufgaben "Größer werden 2" und "Raubüberfälle" den Kompromiss geschlossen, die erste bei "V&B"

und die zweite bei "Unsicherheit" einzusortieren. Bei beiden sind Daten in Grafen repräsentiert, bei der ersten "muss eine Graphik interpretiert werden", bei der zweiten "muss eine Graphik verständlich interpretiert werden" (P2, 54f).

2.3.3 Die dringend nötige zusätzliche Kategorie "Lösungswege"

Mit Händen und Füßen wehrt man sich aber gegen eine wirklich erforderliche Differenzierung, wie sie Meyerhöfer (2004) ins Spiel gebracht hat: Je nach Lösungsweg kann eine Aufgabe unterschiedliche Kompetenzen erfordern und damit unterschiedliche Schwierigkeits-Grade besitzen, und zwar unabhängig davon, ob und in welchem Umfang bei einem bestimmten Test-Durchgang die verschiedenen Lösungswege tatsächlich benutzt wurden. Meyerhöfer hat dies an einigen der veröffentlichten Aufgaben aus PISA 2000 nachgewiesen und, nach einer persönlichen Mitteilung, auch an allen geheim gehaltenen. Dies stellt die Zuordnung einer jeden Aufgabe zu einer bestimmten Punktzahl und damit das Paradigma einer eindimensionalen Skala, die es erlauben würde, mehr als die Schwierigkeit abzulesen, absolut in Frage. Um das Paradigma zu retten, müsste man statt Aufgaben: Paare von Aufgaben und Lösungswegen betrachten. Dieser Ansatz müsste sich zwar nicht in den Test-Heften für die P&P niederschlagen, aber er würde Vorbereitung und Auswertung erheblich aufblähen und erschweren. Er wird von niemandem ernsthaft vorgeschlagen, sondern soll lediglich die Problematik des PISA-Kompetenzstufen-Modells auf den Punkt bringen.

Als ich auf der Tagung des Arbeitskreises "Vergleichs-Untersuchungen" in der GDM am 26.11.04 Vertreter der PMDG nach einer Begründung für die Ablehnung dieses Ansatzes fragte, erhielt ich, zugegeben: in der Hitze der Diskussion, die tautologischen Antworten "passt nicht ins Paradigma" und "bei jeder Aufgabe muss *eine* Zahl stehen".

2.3.4 Abschließende "Würdigung" des simultanen Schwierigkeits-Leistungs-Stufen-Modells

Die Gesichtspunkte, die von der PMDG im Zuge ihres Schwierigkeits-Leistungs-Stufen-Modells berücksichtigt werden, sind alle interessant und haben eine Rolle bei jeglicher Test-Konstruktion zu spielen. Ich habe den Eindruck, dass von der PMDG der ganze PISA-Komplex gründlicher durchdacht ist als von anderen Gruppierungen und dass sie von manchen Anfangs-Setzungen und fortwährenden Restriktionen gehemmt wird. Aber ihr Schwierigkeits-Leistungs-Stufen-Modell hat trotzdem nicht die Aussage-Kraft, die sie ihm zuspricht, da diesem die willkürliche Stufen-Setzung von Beginn an anhaftet und eine stringente wissenschaftliche Begründung nicht erkennbar ist. Eine der zentralen Parolen des PISA-2000-Berichts, nämlich dass nur 44% der 15-Jährigen in Deutschland über den mathematischen Grundbildungs-Standard verfügen (P1.1, 161), bedeutet nicht mehr und nicht weniger, als dass diese 44% bei jenem Test mit seinem Auswertungs- und Berechnungs-Verfahren 512 oder mehr Punkte erreicht haben.

3. Zum Schluss noch eine neue mangelhafte Aufgabe

Vom deutschen PISA-2003-Koordinator Manfred Prenzel wurde in der "Zeit" (09.12.04) folgende Aufgabe aus dem Naturwissenschaften-Test als beispielhaft vorgestellt (P2, 394; 591 Punkte) (Hinweis von Anselm Lambert):

"Tageslicht 1: Welche Aussage erklärt, warum es auf der Erde Tageslicht und Dunkelheit gibt?

- A. Die Erde rotiert um ihre Achse.
- B. Die Sonne rotiert um ihre Achse.
- C. Die Erdachse ist geneigt.
- D. Die Erde dreht sich um die Sonne."

Alle Antworten sind falsch, insbesondere auch A. Die Erklärung lautet vielmehr: "Die Erde rotiert mit einer anderen Winkelgeschwindigkeit um die eigene Achse, als sie sich um die Sonne dreht." Wären die beiden Winkelgeschwindigkeiten gleich, würde die Erde der Sonne immer dieselbe Seite zuwenden, und es gäbe keinen Tag-Nacht-Wechsel. Bei der Drehung des Mondes um die Erde z.B. besteht genau dieser Zustand.

Die Frage ist außerdem für das Gemeinte schlampig gestellt. Die genaue Antwort auf sie lautet nämlich: "Weil das Sonnenlicht nur aus einer Richtung kommt, liegt immer eine Hälfte der Erde im Tageslicht und die andere in der Dunkelheit." Man hätte die Frage deshalb vielleicht so formulieren sollen: "..., warum an jedem Ort der Erde sich Tageslicht und Dunkelheit regelmäßig abwechseln."

Allerdings kommt hier sogar die Neigung der Erdachse ins Spiel: Ist sie nämlich nicht geneigt, dann hält sich die Sonne am Nord- und am Südpol immer am Horizont auf, und an diesen beiden Orten findet nie ein Wechsel zwischen Tageslicht und Dunkelheit statt, während an allen anderen Orten Tageslicht und Dunkelheit immer genau 12 Stunden lang sind. Diesen Zustand gibt es auf der Erde übrigens tatsächlich jährlich zweimal, nämlich am Frühlings- und am Herbst-Anfang. Man müsste also die von mir vorgeschlagene Fragestellung noch modifizieren, etwa: " ..., warum *in unseren Breiten* sich Tageslicht und Dunkelheit regelmäßig abwechseln." Dadurch käme allerdings die erschwerende Rede von "in unseren Breiten" ins Spiel, die ja von vielen nicht verstanden wird.

Zusatz: Wie in der Aufgabe ist auch bei meiner Analyse natürlich das übliche einfache geometrische Modell des Systems "Sonne-Erde" mit zunächst einmal konstanten Winkelgeschwindigkeiten zugrunde gelegt.

— Man hat's nicht leicht als PISA-Mensch.

Literatur

Berichte

(T1; TIMSS I) Ina V.S. Mullis, Michael O. Martin, Albert E. Beaton, Eugenio J. Gonzales, Dana L. Kelly & Teresa A. Smith (1997): Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study. TIMSS International Study Center. Boston College. Chestnut Hill, MA

(T2, TIMSS II) Jürgen Baumert, Rainer Lehmann u.a. (1997): TIMSS — Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske + Budrich

(T3.1, TIMSS III) Jürgen Baumert, Wilfried Bos & Rainer Lehmann (Hrsg.) (2000): TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen: Leske + Budrich

(T3.2, TIMSS III) Jürgen Baumert, Wilfried Bos & Rainer Lehmann (Hrsg.) (2000): TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen: Leske + Budrich

(P1.1, PISA 2000) Jürgen Baumert, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Petra Stanat, Klaus-Jürgen Tillmann & Manfred Weiß (= Deutsches PISA-Konsortium) (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich

(P1.2, PISA 2000) Jürgen Baumert, Cordula Artelt, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Klaus-Jürgen Tillmann & Manfred Weiß (= Deutsches PISA-Konsortium) (Hrsg.) (2002): PISA 2000 — Die Länder der Bundesrepublik Deutschland im Vergleich. Opladen: Leske + Budrich

(P1.3, PISA 2000) Jürgen Baumert, Cordula Artelt, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Klaus-Jürgen Tillmann & Manfred Weiß (= Deutsches PISA-Konsortium) (Hrsg.) (2003): PISA 2000 — Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland im Vergleich. Opladen: Leske + Budrich

(P2, PISA 2003) Manfred Prenzel, Jürgen Baumert, Werner Blum, Rainer Lehmann, Detlev Leutner, Michael Neubrand, Reinhard Pekrun, Hans-Günter Rolff, Jürgen Rost & Ulrich Schiefele (PISA-Konsortium Deutschland) (Hrsg.) (2004): PISA 2003 — der Bildungsstand der Jugendlichen in Deutschland — Ergebnisse des zweiten internationalen Vergleichs. Münster u.a.: Waxmann

(O1.1, PISA 2000) OECD (Hrsg.) (2002): Manual fort he PISA 2000 Database. Paris: OECD

(O1.2, PISA 2000) Ray Adams & Margaret Wu (Hrsg.) (2002): PISA 2000 Technical Report. Paris: OECD

(O1.3, PISA 2000) OECD (2004): Messages from PISA 2000. Paris: OECD

(O2, PISA 2003) OECD (2005): Lernen für die Welt von morgen — erste Ergebnisse von PISA 2003. Erscheint in Heidelberg u.a.: Spektrum

(I1, IGLU) Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walter & Renate Valtin (Hrsg.) (2003): Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Münster u.a.: Waxmann

(I2, IGLU) Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin & Gerd Walther (Hrsg.) (2004): IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich. Münster u.a.: Waxmann

Weitere Literatur

Bender, Peter (2003): Die etwas andere Sicht auf die internationalen Vergleichs-Untersuchungen TIMSS, PISA und IGLU. In: Paderborner Universitätsreden 89, 35–59
math-www.uni-paderborn.de/~bender/

Bender, Peter (2004): Die etwas andere Sicht auf den mathematischen Teil der internationalen Vergleichs-Untersuchungen PISA sowie TIMSS und IGLU. In: Mitteilungen der DMV 12, Heft 2/2004, 101–108, zugleich Mitteilungen der GDM 78

Bender, Peter (2004): Die etwas andere Sicht auf PISA sowie TIMSS und IGLU. In: Beiträge zum Mathematikunterricht 2004. Hildesheim: Franzbecker, 81–84

Bender, Peter (2005): Neue Anmerkungen zu alten und neuen PISA-Ergebnissen und -Interpretationen. Erscheint in: Beiträge zum Mathematikunterricht 2005. Hildesheim: Franzbecker, (auf CD)

Blum, Werner (2003): PISA-2000 Mathematik: Konzeption, Befunde, Interpretationen und Konsequenzen. In: PLAZ (Hrsg.): PISA-Studie 2000 — Impulse für Schule und Lehrerbildung aus zwei Blickwinkeln. Universität Paderborn: PLAZ, 9–21

Hagemeister, Volker (1999): Was wurde bei TIMSS erhoben? Über die empirische Basis einer aufregenden Studie. In: Die Deutsche Schule 91, 160–177

Jahnke, Thomas (2005): Ideologiekritisches und Versöhnliches zu PISA & Co. Erscheint in: Beiträge zum Mathematikunterricht 2005. Hildesheim: Franzbecker

Kaiser, Gabriele (2000): Internationale Vergleichsuntersuchungen im Mathematikunterricht — eine Auseinandersetzung mit ihren Möglichkeiten und Grenzen. In: Journal für Mathematik-Didaktik 21, 171–192

Kießwetter, Karl (2002): Unzulänglich vermessen und vermessen unzulänglich: PISA & Co. In: Mitteilungen der Deutschen Mathematiker-Vereinigung 10, Heft 4/2002, 49–58

Knoche, Norbert, Detlef Lind, Werner Blum, Elmar Cohors-Fresenborg, Lothar Flade, Wolfgang Löding, Gerd Möller, Michael Neubrand, Alexander Wynands (Deutsche PISA-Expertengruppe Ma-

thematik, PISA-2000) (2002): Die PISA-2000-Studie, einige Ergebnisse und Analysen. In: Journal für Mathematik-Didaktik 23, 159–202

Lind, Detlef (1994): Probabilistische Testmodelle. Mannheim u.a.: BI Wissenschaftsverlag

Meyerhöfer, Wolfram (2003): Testfähigkeit: Was ist das? In: Beiträge zum Mathematikunterricht 2003. Hildesheim: Franzbecker, 441–444

Meyerhöfer, Wolfram (2004): Zum Kompetenzstufenmodell von PISA. In: Journal für Mathematik-Didaktik 25, 294–305

Meyerhöfer, Wolfram (2005): Tests im Test: Das Beispiel PISA. Leverkusen: Barbara Budrich

NCTM (Hrsg.) (1989, 2000): Principles and standards for school mathematics. Reston, Va.: NCTM 1989, Überarbeitung 2000

Neubrand, Michael (Hrsg.) (2004): Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland — Vertiefende Analysen im Rahmen von PISA 2000. Wiesbaden: VS Verlag für Sozialwissenschaften

Reiss, Kristina & Günter Törner (2003): PISA 2000: Eine Klärung von Missverständnissen. In: Mitteilungen der Deutschen Mathematiker-Vereinigung 11, Heft 1/2003, 46–48

Internet-Adressen

math-www.uni-paderborn.de/~bender

www.math.nyu.edu/mfdd/braams

www.nychold.com

www.math.uni-potsdam.de/prof/o_didaktik/mita/me/Veroe

www.lisum.de