Die etwas andere Sicht auf den mathematischen Teil der internationalen Vergleichsuntersuchungen PISA sowie TIMSS und IGLU[1]

von Peter Bender

Der vorliegende Vortrag fand auf der Jahrestagung für Didaktik der Mathematik 2004 in Augsburg bemerkenswert großes Interesse. Der Hörsaal war völlig überfüllt. Die vorliegende Bearbeitung enthält die wesentlichen dabei geäußerten Aspekte. *(MT)*

Meine Lieblingsschlussfolgerung „aus PISA" stammt vom ehemaligen Ministerpräsidenten eines großen Bundeslandes, der behauptete, PISA habe gezeigt, dass die Schulzeit bis zum Abitur von 13 auf zwölf Jahre gesenkt werden müsse.

These 1: IGLU (PISA, TIMSS usw. entsprechend) misst nicht *die* Mathematik- (Lese-, naturwissenschaftlichen usw.) „Leistungen am Ende der 4. Jahrgangsstufe im internationalen Vergleich" (I1, Buchtitel), sondern die Leistungen bei *diesem* Test unter *diesen* Bedingungen, die, wie bei jedem Test, zu einem großen Teil darin bestehen, das von den Autoren Gemeinte zu entschlüsseln.

These 2: Es steht dahin, ob (i) dem im Angelsächsischen entwickelten Konstrukt der *mathematical literacy* in der dort festgelegten Ausprägung wirklich die fundamentale Rolle im Mathematik-Unterricht zukommen soll, (ii) die Test-Aufgaben für dieses Konstrukt valide sind, (iii) es angebracht ist, die Länder-Curriculum-Validität erklärtermaßen zu ignorieren und zugleich Länder-Punktzahl-Vergleiche anzustellen und zu veröffentlichen, (iv) bei noch so guten Übersetzungen nicht doch diejenigen Schüler im Vorteil sind, in deren Muttersprache die Aufgaben ursprünglich formuliert worden waren, nämlich auf Englisch.

These 3: Es steht dahin, ob die statistischen und organisatorischen Vorgaben wirklich weltweit hinreichend skrupulös eingehalten wurden.

These 4: (i) Für die deutsche Öffentlichkeit waren die schockierenden Ergebnisse heilsam, da die Relativierungen in den Thesen 1–3 nichts an den Grund-Tendenzen ändern. (ii) Für jemanden mit Einblick in die Schulrealität waren sie keine Überraschung: Wer wissen wollte, konnte auch schon vorher wissen. (iii) Hoffentlich wird der ganze Aufwand durch den Erfolg von in die Wege geleiteten und noch zu leitenden Maßnahmen gerechtfertigt.

These 5: (i) Viel mehr als über das Bildungssystem sagen die Statistiken jedoch etwas über gewisse Mangelerscheinungen unserer Gesellschaft aus, besonders die schlechte Integration der Kinder mit Migrationshintergrund (mindestens ein Elternteil nicht in Deutschland geboren), vor allem wenn dies für beide Elternteile zutrifft. Diese können nicht allein innerhalb des Bildungssystems repariert werden, wenn überhaupt mittelfristig. (ii) Die gesellschaftlich-kulturellen Bedingungen werden unterschätzt, z. B. der Einfluss der (aus unserer Sicht vielleicht abzulehnenden, aber eben testerfolgreichen) Leistungs- und Disziplinorientierung in vielen ostasiatischen Ländern oder deren Verlust in den früheren sozialistischen Ländern in Verbindung mit dem Verlust der autoritären Strukturen dort. Dagegen sind z. B. die aus der TIMSS-Video-Studie (deren Repräsentativität außerdem natürlich in Frage steht) extrahierten Unterschiede in den Unterrichts-Stilen in Japan, Deutschland und den USA eher belanglos. (iii) Die Vergleichsstudien geben weder etwas *für* noch *gegen* eine Bevorzugung der Einheitsschule her.

Das Idealbild eines Tests mathematikbezogener Kompetenzen und die Realität

Welche mathematikbezogenen Fähigkeiten, Fertigkeiten, Wissensbestände, Einstellungen usw. hält man für wichtig, so dass man mit dem Grad ihres Vorhandenseins die mathematische Leistungsfähigkeit eines Individuums oder einer ganzen Population bestimmt? Wie misst man diese Tugenden? – Üblicherweise lässt man übliche Aufgaben lösen, und zwar bei einem voluminösen Unternehmen wie PISA oder IGLU überwiegend solche, bei denen die Antwort entweder richtig oder falsch ist, also 1 oder 0 Punkte ergibt. Dabei unterstellt man Validität, d. h. dass tatsächlich die interessierenden Tugenden relevant sind. Allerdings gibt es dazu keine robusten mathematikdidaktischen Forschungsergebnisse.

Es ist klar, dass in einem solchen Test viele durchaus wichtige Tugenden gar nicht berücksichtigt werden:

1 Gekürzte und vor allem aufgrund der Diskussion auf der GDM-Bundestagung 2004 in Augsburg weiter entwickelte Version des Vortrags (Bender 2003). Unter der u. a. Internet-Adresse findet man vollständige Quellen-Angaben. Aus Platzgründen fasse ich die meisten mir wesentlichen Gesichtspunkte in Thesen zusammen und stelle lediglich einige mathematik-didaktische Aspekte etwas ausführlicher dar.

Unauthenticated
Download Date | 3/18/19 11:36 PM

Die Fähigkeit, komplexe Probleme anzugehen, überhaupt Mathematisierbarkeit zu prüfen, ein Problem längerfristig und mit Ausdauer zu bearbeiten, 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.

Mehr oder weniger gute Aufgabenformulierungen

In der deutschen mathematik-didaktischen Arbeitsgruppe von IGLU, deren Mitglied ich war, hatten wir uns u. a. vorgenommen, die Aufgaben des internationalen Tests für deutsche Viertklässlerinnen und -klässler adäquat, d. h. insbesondere auch kindgemäß zu formulieren. Als Vorlage standen uns die Aufgaben für die Kinder aus Österreich von TIMSS I (1995) zur Verfügung. Abgesehen von österreich-typischen Wendungen wie „Buben" oder „Milchpackerl" strotzten die Formulierungen von irreführenden, ungenauen, kindfremden Wendungen, z. B.:

H6: Dies ist ein Rechteck mit einer Länge von 6 cm und einer Breite von 4 cm. Die Strecke rund um seine Form nennt man Umfang. *(Zeichnung)* Was gibt den Umfang des Rechtecks in Zentimetern an?

A	$6 + 4$	B	$6 \cdot 4$
C	$6 \cdot 4 \cdot 2$	D	$6 + 4 + 6 + 4$

H8: Dies ist ein Zahlenmuster:
100, 1, 99, 2, 98, $\boxed{}$, $\boxed{}$, $\boxed{}$.
Welche drei Zahlen passen in die Kästchen?

A	3, 97, 4	B	4, 97, 5
C	97, 3, 96	D	97, 4, 96

A3: Es sind 9 Schachteln mit Stiften gegeben. Jede Schachtel hat 125 Stifte. Wie lautet die Gesamtzahl der Stifte?

A	1025	B	1100	C	1125
D	1220	E	1225		

Man wundert sich, dass die Kinder aus Österreich damals mit 559 Punkten (T1, 24) recht gut abgeschnitten haben. Dies lässt sich jedenfalls nicht damit erklären, dass solche schlechten Formulierungen sich nicht auf die Leistungen auswirken würden; denn es ist bekannt, dass durch Variation von Sprachwendungen oder nur des Layouts starke Effekte erzielt werden können.

Leider konnten wir unsere Verbesserungsideen nicht in voller Konsequenz umsetzen, damit die Vergleichbarkeit mit TIMSS I einigermaßen erhalten blieb. Allerdings waren die originalen englischen Formulierungen oft auch nicht astrein. Aber die alte deutsche Übersetzung ist bei fast allen Aufgaben noch deutlich schlechter.

Ein weiteres Beispiel, zur Abwechslung auf SII-Niveau (T3.2, 93):

K14: Eine Schnur ist symmetrisch um einen zylindrischen Stab gewickelt. Die Schnur windet sich genau viermal um den Stab. Der Umfang des Stabes beträgt 4 cm und seine Länge 12 cm. *(Zeichnung)* Bestimmen Sie die Länge der Schnur. Schreiben Sie Ihre Arbeitsschritte auf.

Von Karl Kießwetter (2002) stammt eine überzeugende Kritik dieser und anderer Testaufgaben, die ich durchweg teile. Jetzt möchte ich nur die Verwendung des Worts „symmetrisch" aufspießen, die jedenfalls in der deutschen Fachsprache falsch ist (da auch die – den Autoren allerdings vermutlich nicht bekannte – Schraub-Symmetrie sich wie die Translations-Symmetrie nur auf unendlich lange Figuren beziehen kann) und für deutsche Schüler nicht zielführend ist. Angebracht wäre stattdessen die Rede von „gleichmäßig" bzw. „konstanter Ganghöhe".

Was testen die Aufgaben eigentlich wirklich?

Auch bei TIMSS, PISA und IGLU wird ganz wesentlich eine extrinsische Fähigkeit der Schüler abgeprüft, nämlich herausfinden zu können, was die Aufgabenautoren wohl gemeint haben. Diese Herausforderung ist bei innermathematischen Aufgaben naturgemäß geringer, aber bei den Aufgaben mit einem irgendwie gearteten außermathematischen Kontext beliebig schwierig (und wird durch misslungene Formulierungen bzw. schlechte Übersetzungen noch verschärft). Da sind natürlich wieder diejenigen im Vorteil, die an solche (standardisierten, von Fremden gestellten) Tests gewöhnt sind und in deren Sprache und Kultur die Aufgaben ursprünglich angesiedelt sind.

Wenn in Schweden die Lösungshäufigkeit der Zylinderwickel-Aufgabe mit 24 % sechsmal so hoch ist wie in Frankreich, so ist das für mich als Geometrie-Didaktiker zunächst einmal ein Indiz, dass im schwedischen Curriculum raum-geometrische Aufgaben, womöglich von diesem speziellen Typ, intensiver behandelt werden als im Geometriemutterland Frankreich, wo allerdings bekanntlich der Geometrieunterricht sowieso stark algebraisiert ist und sich hauptsächlich auf die affine Ebene beschränkt.

Hier drängt sich die Frage nach der Unterrichts- und Curriculums-Validität auf. Diese ist in IGLU und TIMSS eher, in PISA weniger gut gegeben.

Unauthenticated
Download Date | 3/18/19 11:36 PM

Das „mathematical literacy"-Konzept auf Kosten der curricularen Validität

Während bei TIMSS die Validität bezüglich der Länder-Curricula noch ein erklärtes Ziel war, findet bei den PISA-Tests ein „Verzicht auf transnationale curriculare Validität" statt; stattdessen „führen sie ein didaktisches und bildungstheoretisches Konzept mit sich, das normativ ist" (P1, 19). Als hierfür „in mancher Hinsicht vorbildlich" (P1, 25) werden die NCTM-Standards für den Mathematikunterricht (NCTM 1989, überarbeitet 2000) zitiert. Es ist klar, dass Länder, die ihr geschriebenes und reales Curriculum stärker daran ausgerichtet haben, ob durch explizite Übernahme oder durch eigene Entwicklung, „bevorzugt" werden. Für Deutschland trifft das im Großen und Ganzen nicht zu, aber m. W. für viele angelsächsischen und skandinavischen Länder sowie die Niederlande.

Wie den meisten Kolleginnen und Kollegen aus der deutschsprachigen mathematik-didaktischen Kommunität sagt mir das Grundbildungskonzept[2] von PISA für den Mathematikunterricht durchaus zu, und ich teile die schon von den TIMSS-Leuten vertretene Auffassung, dass der deutsche Mathematikunterricht allzu sehr auf Faktenwissenserwerb und die Beherrschung von Verfahren zielt (T2, 31).

DEFINITION VON MATHEMATICAL LITERACY: „Die Rolle zu erkennen und zu verstehen, die die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens einer Person als konstruktivem, engagiertem und reflektierendem Bürger entspricht." (P1, 23)

Diese „Definition" passt durchaus zur Tradition der deutschen bildungstheoretischen Didaktik, wie sie z. B. vom alten Wolfgang Klafki (1958) oder speziell zum Mathematikunterricht von Heinrich Winter (1975) mit seinem Begriff der „Umwelterschließung" verkörpert wird. Allerdings enthält die Konzeption von PISA einen stärker pragmatischen Zug (P1, 19); ich persönlich vermisse dabei z. B. die Rolle der Mathematik als Kulturgut.

Der dominierende Bestandteil von mathematical literacy im Sinne von PISA ist die Kompetenz zum Modellieren (offenbar i. W. mit Problemlösen und sogar erklärtermaßen mit Aufgaben-Lösen zu identifizieren; I2, 118 f) in den Kompetenzklassen: (i) Verfahren ausführen; (ii) Zusammenhänge herstellen; (iii) Verallgemeinern (mit Betonung der Klassen (ii) und (iii) auf Kosten von (i)). Besonders wichtig unter den NCTM-Standards scheint der folgende zu sein: „Vorbereitung auf offene Aufgabenstellungen, da realistische Probleme und Aufgaben in der Regel nicht gut definiert sind" (P1, 25).

Wie weit kann mathematical literacy überhaupt getestet werden?

Insgesamt kann ein Test wie PISA oder IGLU, zumal mit überwiegend Multiple-Choice-Charakter, der o. a. Definition natürlich nicht gerecht werden (so schon Kießwetter). Kein einziger Aspekt kann sich in solchen Aufgaben *genuin* wiederfinden: Es ist nirgends nötig, eine vorgelegte Situation überhaupt auf Mathematisierbarkeit zu prüfen; denn es ist immer klar, dass zu mathematisieren ist. Es kann nirgends das Erkennen und Verstehen der Rolle der Mathematik in der Welt wirklich aufgezeigt werden; usw. Keine einzige dieser Häppchenaufgaben, schon wegen der notwendig fehlenden Komplexität, stellt ein authentisches Sachproblem dar, gar ein Problem der Schüler selbst. Natürlich ist keine Aufgabe wirklich offen; es ist lediglich immer wieder der Versuch erkennbar, ein direktes Anwenden von Faktenwissen und Fertigkeiten durch Einkleidung des mathematischen Gehalts in allerlei inner- und außermathematische Kontexte zu verhindern und so immerhin Modellieren zu erzwingen.

Das Wort „Aufgabe" sagt eigentlich alles: Unterricht und Tests sind Unternehmen, wo Erwachsene den Schüler etwas „aufgeben", um damit bei diesen etwas zu erreichen (z. B. Einstellungsänderung, Lernzuwachs, Demonstrieren von Wissen usw.), auch dann, wenn die Schüler die eine oder andere Fragestellung scheinbar selbst hervorbringen. *Alle* Schüler wissen um diesen Schonraum-Vermitteltheits-Pädagogik-Charakter von Schule, in den ja die Testsituation integriert ist. Wo die „offizielle" Rechenmethodik (und die Pädagogik überhaupt) bis in die 1960er Jahre diesen Charakter mehr oder weniger bewusst ausgelebt hat, merken deren Kritiker, die seitdem mit den unterschiedlichsten Alternativen auf den Plan getreten sind (darunter die Verfechter des mathematical literacy-Konzepts), nur nicht, dass sie – wohl etwas subtiler – aber dennoch: nichts anderes tun; und eigentlich auch nichts anderes tun können.

Um es noch einmal deutlich auszusprechen: In Schule und Tests sind Begrifflichkeiten wie „authentische" oder „offene Aufgabe" Widersprüche in sich, und entsprechend etikettierte Problemstellungen unterscheiden sich nur graduell von den verpönten so genannten „eingekleideten" oder „geschlossenen Aufgaben". Natürlich liegt dies nicht an dem deutschen Wort „Auf-

2 Als Übertragung des anglo-amerikanischen Begriffs „mathematical literacy"; Neubrand (2003) bevorzugt allerdings inzwischen die Rede von „mathematischer Literalität".

Unauthenticated
Download Date | 3/18/19 11:36 PM

gabe“, sondern es gilt auch für die englische Rede von „problems“, insofern es sich ja um „posed problems“ handelt.

Meine Kritik richtet sich nicht gegen diese Art von Tests. Man kann 900.000 Schüler (bei PISA, P1, 34 ff) oder 150.000 Schüler (bei IGLU) weltweit nicht anders testen. Aber vom Testen der mathematical literacy ist man viel weiter entfernt und klassischem Wissens- und Könnens-Abfragen steht man viel näher, als man glaubt oder glauben machen möchte. Vielleicht sind solides Wissen und Können (natürlich inklusive des Lernstoffs „Modellbildung“) viel bedeutsamer für die mathematical literacy, als die PISA-Leute annehmen. Ich habe jedenfalls schon einige Untersuchungen kennen gelernt, die dieses nahe legen und die mich deswegen so überzeugt haben, weil ihre Autoren ersichtlich eigentlich gerne das Gegenteil herausbekommen hatten wollen.

Ein Beispiel (das mit seiner Realitätswidrigkeit nicht besonders schlecht, sondern durchaus typisch ist) aus TIMSS II und TIMSS III (T3.1, 164), dort unter „social utility“ (!) firmierend:

> Diese beiden Anzeigen sind in einer Zeitung in einem Land erschienen, in dem die Währungseinheit *zeds* ist:
>
> GEBÄUDE A: Büroräume zu vermieten: 85–95 qm 475 *zeds* pro Monat; 100–120 qm 800 *zeds* pro Monat;
>
> GEBÄUDE B: Büroräume zu vermieten: 35–260 qm 90 *zeds* pro Quadratmeter pro Jahr.
>
> Eine Firma ist daran interessiert, ein 110 qm großes Büro in diesem Land für ein Jahr zu mieten. In welchem Bürogebäude, A oder B, sollte sie das Büro mieten, um den niedrigeren Preis zu bekommen? Wie rechnen Sie?

Vermutlich ist an folgende Lösung gedacht: Bei A muss man $12 \cdot 800 = 9600$ *zeds*, bei B $110 \cdot 90 = 9900$ *zeds* zahlen. – Die Schüler haben natürlich, schon aus Zeitgründen, jedes Nachdenken über den Realgehalt dieser Situation auszuschalten, und das wissen sie auch. – Problematisieren müssten sie eigentlich: Ist überhaupt ein genau 110 qm großes Büro vorhanden? Vor allem aber: Wo gibt es das, dass man 475 *zeds* pro Monat für 95 qm und 800 *zeds* pro Monat (fast das Doppelte) für 100 qm (fast die gleiche Größe) zahlen muss? Da wäre doch eine Firma mit dem Klammersack gepudert, wenn sie sich für das eine Jahr nicht mit 95 qm bescheiden würde (und zur Not noch 35 qm im anderen Gebäude hinzu mieten würde mit einer Gesamtsumme dann von $475 \cdot 12 + 35 \cdot 90 = 8850$ *zeds* für 130 qm(!)). – Solche Überlegungen wären Teil der mathematical literacy. Genau diese werden hier nicht erwartet und können nicht erwartet werden.

Bei IGLU spielt das Konzept der mathematical literacy noch nicht die Rolle wie bei PISA (I1, 14), vor allem, weil die IGLU-Mathematik ja letztlich TIMSS I ist. – Allerdings kann auch hier das Bemühen um Überwindung der simplen Wissens- und Verfahrensabfrage zu Verunsicherungen führen; ein Beispiel aus TIMSS I (Österreich):

> B7: In Marks Garten gibt es 84 Reihen mit Krautköpfen. In jeder Reihe sind 57 Krautköpfe.
> Welche der folgenden Gleichungen bietet die *beste* Möglichkeit, die Gesamtzahl der Krautköpfe abzuschätzen?
>
> A $100 \cdot 50 = 5000$ B $90 \cdot 60 = 5400$
> C $80 \cdot 60 = 4800$ D $80 \cdot 50 = 4000$

Aus meiner Sicht sind A und C beide gut, und welche besser ist, hängt vom Gütekriterium ab. Für A spricht, dass die Schätzzahlen besonders einfach sind und das Ergebnis gleich auf volle Tausender gerundet ist, während man bei C dafür einen zweiten Rundungsschritt bräuchte. Wenn man das genaue Ergebnis $84 \cdot 57 = 4788$ errechnet, sieht man natürlich, dass C näher dran liegt. – Ganz bestimmt nicht waren bei dieser Aufgabe solche Überlegungen erwartet worden; aber gerade sie wären der Ausfluss von mathematical literacy gewesen. – Diesem Dilemma entgeht man auch nicht dadurch, dass man beide Antworten als richtig akzeptiert.

Zu konzedieren ist, dass aus mathematik-didaktischer Sicht die Aufgaben in PISA insgesamt von höherer Qualität zu sein scheinen als die in TIMSS.

Das fragwürdige Konstrukt der Kompetenzstufen

Die Kompetenz-*Klassen* sind bei TIMSS, PISA und IGLU noch nicht das letzte Wort. *Nach* den Tests wurde die Punkteskala in Abschnitte eingeteilt, und jeder Abschnitt wurde zu einer Kompetenz-*Stufe* erklärt. Zugleich liefert diese Kompetenzstufeneinteilung für die Schüler eine Schwierigkeitsstufeneinteilung für die Aufgaben.

Bei PISA und IGLU hat man es damit genauer genommen als bei TIMSS und wollte mehr daraus machen. Bei PISA waren die Grenzen vom internationalen Konsortium festgelegt worden. Man hat einen Bereich von 329 bis 696 angenommen, und diesen in vier gleichlange Intervalle zerlegt (P1, 187), also mit der Länge $\frac{696-329}{4} = 91{,}75$. Wie dieser Bereich zustande kam, konnte ich nicht eruieren. Auf eine inhaltliche Beschreibung der Stufen wurde verzichtet; diese wurde erst für den deutschen Bericht nachgeholt:

Unauthenticated
Download Date | 3/18/19 11:36 PM

V	≥696	Komplexe Modellierung und innermathematisches Argumentieren
IV	604–695	Umfangreiche Modellierungen auf der Basis anspruchsvoller Begriffe
III	512–603	Modellieren und begriffliches Verknüpfen auf dem Niveau der Sekundarstufe I
II	421–511	Elementare Modellierungen
I	329–420	Rechnen auf Grundschulniveau
0	≤328	

Offensichtlich ist die Stufenabfolge 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. Dass da jemand Rechenfehler macht und komplex modellieren kann, oder innermathematisch argumentieren, aber nur elementar modellieren kann (o. ä.) und die Einordnung dieser Person etwa in Stufe III diesem Kompetenzprofil nicht gerecht wird, ist hier nicht vorgesehen. Stattdessen wird in den Kurzerläuterungen bei Stufe IV und V von „offenen Aufgaben“ geredet, die ja eigentlich im Test gar nicht vorkommen (können). – Die Fragwürdigkeit dieser Etikettierung zeigt sich schön an der Pyramiden-Aufgabe (P1, 151 ff):

> Die Grundfläche einer Pyramide ist ein Quadrat. Jede Kante der skizzierten Pyramide misst 12 cm. *(Zeichnung)*
> Bestimme den Flächeninhalt einer der dreieckigen Seitenflächen. Erkläre, wie du deine Antwort gefunden hast.

Mit 810 Punkten hat sich diese Aufgabe als sehr schwer erwiesen. Ich hätte sie eigentlich leichter eingeschätzt: man muss ja nur den Flächeninhalt eines gleichseitigen Dreiecks bei bekannter Seitenlänge a = 12 cm ermitteln, und der beträgt $\frac{\sqrt{3}}{4} \cdot a^2 = 62\,\text{cm}^2$. So gesehen, ist das eine reine Wissensaufgabe, und eine „komplexe Modellierung und innermathematisches Argumentieren“ vermag ich nicht zu erkennen.

Eine Schwierigkeit der Aufgabe besteht darin, dass die Dreiecke in einer dreidimensionalen Situation gegeben sind und weltweit zu wenig Raum-Geometrie getrieben wird, so dass bei den meisten Schüler schon deswegen die Klappe fällt, obwohl es sich um ganz gewöhnliche (notwendig ebene) gleichseitige Dreiecke handelt. Wenn man in der Schule schon einmal solche dreidimensionalen Situationen behandelt hat, fällt einem diese Sichtweise leichter; und wenn nicht, dann eben nicht.

Die zweite Schwierigkeit liegt darin, dass i. d. R. keine Formeln auswendig gewusst werden. In früheren Jahren war das weltweit anders, und für manche kaufmännische und gewerbliche Berufe war es ausgesprochen nützlich, über einen Vorrat an Formeln zu verfügen, so dass man nicht jedes Mal bei Bedarf irgendwo nachschauen musste. Mit Recht ist zwar das Auswendiglernen von Formeln heutzutage verpönt, und man muss sie sich nur noch herleiten können. Wenn man in der Schule schon einmal die Flächeninhaltsformel für das gleichseitige Dreieck hergeleitet und/oder man sich die Herleitung oder die Formel eingeprägt hat und sich daran erinnert, fällt einem die Bearbeitung der PISA-Aufgabe natürlich viel leichter, als wenn nicht.

Die PISA-Leute kommen schließlich zu der Feststellung, dass man erst ab der Kompetenzstufe III von einem ausreichenden Standard für mathematical literacy sprechen sollte (P1, 161), und zu Bezeichnungen wie Risiko- und Spitzengruppe. In diesem Sinn verfügt Deutschland über eine Risikogruppe von 25 % der Heranwachsenden, und nur 44 % befinden sich mindestens auf Stufe III. Die Schlagzeile, dass deutlich weniger als die Hälfte unserer 15-Jährigen den PISA-Grundbildungsstandard erfüllt, beruht jedoch zunächst einmal lediglich auf dieser willkürlichen Klasseneinteilung.

In der *Diskussion* zu meinem Vortrag auf der GDM-Bundestagung 2004 wurde hier sinngemäß eingewandt, dass man nun endlich einmal ein kognitionswissenschaftliches Instrumentarium zur Verfügung habe, mit dem man (potenzielle oder reale) Testaufgaben und ihre (erwarteten oder beobachteten) Lösung(sschwierigkeit)en konstruieren bzw. analysieren könne. – Das Bemühen um inhaltsunabhängige Kriterien ist schon uralt; und schon immer hat man solche für die Analyse von Aufgaben ad hoc herangezogen. – Auch gibt es einige kognitionswissenschaftliche Theorien dazu, die aber alle daran kranken, dass sie für die Anwendung auf konkrete, stofflich breit angelegte Tests wie TIMSS, PISA oder IGLU ersichtlich *nicht brauchbar* sind. Folgerichtig konnte ich keine solche Theorie in Erfahrung bringen, die bei der Entwicklung und Auswertung der genannten Tests eingesetzt worden wäre. Wenn nun für die Zukunft solche Theorien für die Analysen angekündigt werden, so ist zum einen angesichts der mageren Erfolge bis Anfang 2004 Skepsis angebracht, zum anderen hätten sie natürlich *vor* jedem der entscheidenden Schritte „Durchführung“, „Auswertung“ und „Verkündung der Statistik und der Folgerungen“ zugrunde gelegt werden müssen.

In der erwähnten Diskussion zu meinem und mehr noch zu Meyerhöfers (2004) Vortrag rügten Kompetenzstufen-Apologetinnen und -Apologeten (die Verwendung des Plurals hier ist eigentlich ein grammatikalisch bedingter Euphemismus) (i) das

Unauthenticated
Download Date | 3/18/19 11:36 PM

Heranziehen überhaupt von Testaufgaben zur Überprüfung des Kompetenzstufenmodells, (ii) die „negative“ Auswahl der Aufgaben und (iii) das Ignorieren späterer Verbesserungsbemühungen.

Meyerhöfer hat die Aufgabenanalyse weiter getrieben als ich und nachgewiesen, dass für fast jede der zu PISA veröffentlichten Aufgaben gilt: je nach dem eingeschlagenen Weg zu ihrer Lösung ist sie auf mehreren verschiedenen Kompetenzstufen (gemäß deren inhaltlichen Charakterisierung durch die PISA-Leute selbst) anzusiedeln. Ich wüsste gar kein passenderes Instrument als die Analyse der Testaufgaben, um die Qualität eines Tests bzw. die Eignung eines so zentralen und folgenschweren Konstrukts wie des Kompetenzstufenmodells zu prüfen. Die Auswahl der Aufgaben wiederum ist durch die Geheimhaltungspolitik besonders bei PISA und IGLU vorgegeben, und man muss vernünftigerweise unterstellen, dass die veröffentlichten Aufgaben musterhaft bzw. typisch sind. Allerdings hat Meyerhöfer ein Übriges getan, die geheimen Aufgaben des PISA-Tests ebenfalls untersucht und einen vergleichbaren Befund erhalten. So viel zu den Rügen (i) und (ii). Nun zu (iii): Dass einige Zeit nach der Publizierung der PISA-Ergebnisse das Kompetenzstufenmodell überarbeitet wurde, ist zwar löblich, kommt aber für die Auseinandersetzung mit PISA-2000 insofern zu spät, als die dort vorgenommenen Einschätzungen, weitreichenden Folgerungen und medialen Ausschlachtungen ja auf den originalen Veröffentlichungen beruhen.

In der Tat macht die Fortentwicklung, wie sie etwa in (Knoche u. a. 2002, 182ff) dargestellt ist, einen mathematik-didaktisch professionelleren Eindruck, wobei die enge Ausrichtung an bewährten stoffbezogenen Kriterien positiv auffällt. Die Stufenmerkmale werden zusätzlich „ je nach dem angesprochenen Typ mathematischen Arbeitens“ den drei Kategorien „technische“, „rechnerische“ und „begriffliche Aufgaben“ zugeordnet (S. 182). Man muss also bei einer Aufgabe (und im Meyerhöferschen Sinn: bei jedem der möglichen Lösungswege) zunächst einmal festlegen, zu welchem „Typ mathematischen Arbeitens“ sie gehört, und dann nur diejenigen Stufenmerkmale in Betracht ziehen, die dem entsprechenden Typ zugeordnet sind.

Die Klasseneinteilung der Aufgaben in die fünf Kompetenzstufen (ohne Stufe 0) wird also auf das Dreifache verfeinert. Entsprechend vervielfacht sich die Gefahr der Inkonsistenz, wie ich sie oben im Anschluss an die Vorstellung der Kompetenzstufen angedeutet habe und Meyerhöfer sie mit seiner Betrachtung unterschiedlicher Lösungswege bei ein- und derselben Aufgabe überzeugend dargelegt hat. Direkt ins Auge springt z. B. das Problem, dass eine Aufgabe mehr als einen der drei Typen mathematischen Arbeitens beinhalten kann und dann bei verschiedenen Typen auf verschiedenen Stufen landet. Da könnte man einfach vorschreiben, dass zur Charakterisierung der Aufgabe (des Lösungswegs) immer die höchste der bei den drei Typen erreichten Stufe zu wählen sei. Aber da käme es dann doch auf das Gewicht an, mit dem die Typen mathematischen Arbeitens jeweils das Wesen einer Aufgabe (eines Lösungswegs) ausmachen; und dieser Einwand zeigt, dass dieses einfache Vorgehen letztlich auch inkonsistent ist. Offenbar ist die scharfe Stufung, verbunden mit dem Anspruch auf eindeutige Einordnung der Aufgaben, *nicht haltbar*, erst recht nicht, wenn man noch (unvermeidlich) nach Lösungswegen differenziert.

Die Diskussion kulminierte in dem das Thema verfehlenden Vorwurf an Meyerhöfer, er verstünde nichts von Testkonstruktion. Ähnliche Elemente finden sich auch in der Replik von Reiss & Törner (2003) auf Kießwetter.

Es ging überhaupt nicht um die Testkonstruktion bei TIMSS, PISA und IGLU, sondern um Test-*Interpretation.* Und da geben diese Tests einfach nicht mehr her als die triviale klassische Feststellung (mit allen ihren problematischen Begleiterscheinungen): je mehr Punkte eine Aufgabe erhält, umso „schwerer“ ist sie. Selbstverständlich lassen sich bei den Merkmalen des Kompetenzstufenmodells Tendenzen herauslesen; aber mehr ist nicht drin (was man bei TIMSS übrigens noch wusste). Man kann den PISA-Leuten nur empfehlen, auf die starken Aussagen zu verzichten, die sie (manche) aus dem Kompetenzstufenmodell gezogen haben oder demnächst wieder ziehen wollen, auch wenn dadurch einiges an Medienwirksamkeit verloren geht.

Die Aufgaben kann man durchaus mit dem wohlbekannten (und selbstverständlich ständig weiter zu entwickelnden) Instrumentarium aus Mathematik-Didaktik und Bezugswissenschaften analysieren, und man kann lokale und globale Schwierigkeiten identifizieren sowie Verallgemeinerungsansätze entwickeln. Allerdings müsste dies schon *vor* den Tests durchgeführt worden sein und insbesondere auch die Auswahl der Aufgaben geleitet haben. Vielleicht ist diese Forderung ja bei PISA-2003 erfüllt, wodurch dann immerhin einer der Mängel von PISA-2000 vermieden wäre. – Es wäre übrigens interessant zu erfahren, was in anderen Ländern an Kompetenzstufentheorie aus der vorgegebenen Intervalleinteilung abgeleitet wurde. Bei dem universalistischen Anspruch von PISA müsste da ja überall i. W. dieselbe wie in Deutschland entstanden sein, oder?

Die Kompetenzstufen bei IGLU

Bei IGLU ist man mit den Kompetenzstufen pragmatischer umgegangen (I1, 200 ff):

Unauthenticated
Download Date | 3/18/19 11:36 PM

V	≥651	Problemlösen bei Aufgaben mit innermathematischem oder außermathematischem Kontext;
IV	521–650	Beherrschung der Grundrechenarten, Bewältigung von Aufgaben der räumlichen Geometrie und begriffliche Modellentwicklung;
III	411–520	Verfügbarkeit von Grundrechenarten und Arbeit mit einfachen Modellen;
II	290–410	Grundfertigkeiten zum Zehnersystem, zur ebenen Geometrie und zu Größenvergleichen;
I	≤289	Rudimentäres schulisches Anfangswissen.

Natürlich ist jegliches Bearbeiten mathematischer Aufgaben ein Problemlösen; Problemlösen ist also kein Merkmal einer besonders hohen Kompetenzstufe. Auch inner- und außermathematische Kontexte kommen auf jeder Kompetenzstufe vor, z. B. auf Stufe I: „Jan hat 4 Bonbons, Marie hat 3 Bonbons, wie viele haben sie zusammen?“ – Insofern liefert die Charakterisierung von Stufe V z. B. noch keinen Hinweis auf hohe Kompetenz.

Insgesamt schätze ich die Rolle des Modellbildens als im schulischen Mathematikunterricht zu erwerbende Kompetenz nicht als so überwältigend ein wie die IGLU- oder gar die PISA-Leute. Inzwischen musste ich allerdings zur Kenntnis nehmen, dass von diesen die Rede vom Modellbilden auf *jegliches* Lösen mathematischer Aufgaben angewendet wird (I2, 118 f), womit sie so universell und zugleich so nichts sagend wie die vom Problemlösen geworden ist.

Auch im Bereich der Naturwissenschaften in der Grundschule (I1, 156 ff) kommt mir die Kompetenzstufenbildung nicht gut gelungen vor:

V	≥638	Naturwissenschaftliches Verständnis und Lösungsstrategien;
IV	523–637	Beginnendes naturwissenschaftliches Verständnis;
III	469–522	Anwenden naturwissenschaftsnaher Begriffe;
II	401–468	Anwenden alltagsnaher Begriffe;
I	323–400	Einfache Wissensreproduktion;
0	≤322	Vorschulisches Alltagswissen.

Man muss überrascht zur Kenntnis nehmen, dass die Frage „Welches Tier säugt seine Jungen?“ mit den Antwortmöglichkeiten „Huhn, Frosch, Affe, Schlange“ den Schwierigkeitsgrad 474 hat und in Stufe III, Anwenden nicht mehr alltags-, sondern nun naturwissenschaftsnaher Begriffe, gerät. Diese Kategorisierung wird dieser Frage nicht gerecht. Da geht es vielmehr um Allgemeinbildung auf niedrigem Niveau, um elementares Wissen, das entweder vorhanden ist oder nicht.

Für die Stromkreis-Aufgabe (546 Punkte) gilt erst recht, dass man sie bei noch so viel naturwissenschaftlichem Verständnis nicht lösen kann, wenn man nicht massiv den Stoff und die übliche, auf Konventionenen beruhende Art der grafischen Darstellung behandelt hat, abgesehen von der Überlagerung des naturwissenschaftlichen Problems durch ein kombinatorisches bei den vorgegebenen Antwort-Möglichkeiten.

These 6: (i) Aus mathematik-didaktischer Sicht ist das Instrument der Kompetenzstufen, wie es insbesondere für PISA entwickelt wurde, fragwürdig. Die Einschätzung von Schülerleistungen oder der Vergleich von Populationen beruht nämlich letztlich auf dem alten Prinzip: „Je mehr Punkte, desto besser.“ Die Einteilung in Stufen ist willkürlich und ersichtlich nicht das Ergebnis einer mathematikdidaktisch-kognitionswissenschaftlichen Theorie.

(ii) Der Versuch, sich von der ausschlaggebenden Bedeutung der konkreten mathematischen Inhalte (und der nationalen Curricula) zu lösen, muss als gescheitert angesehen werden. Die zentralen Konstrukte „mathematical literacy“, „Modellbildung“, „Problemlösen“, „Anwenden“ sind, wenn man sie nicht ganz nichtssagend versteht, nicht mit solchen Tests abprüfbar.

Schlussbemerkung

Gestört hat mich vor allem bei PISA das Aufschwingen zu einer obersten bildungswissenschaftlichen Instanz, das bestimmt auch zu dem immer wieder zu beobachtenden Missbrauch durch Politik und Medien beigetragen hat, sowie die seit TIMSS zunehmende Geheimniskrämerei um die Ergebnisse zu den konkreten Aufgaben.

Trotzdem: Es war überfällig, dass auch Deutschland sich an diesen internationalen Vergleichsuntersuchungen beteiligt. Auch wenn man meint, dass die Durchschnittswerte und Platzziffern eigentlich (in welchem Sinn auch immer) ein bisschen günstiger auszusehen hätten, als sie sich tatsächlich ergeben haben, muss man zur Kenntnis nehmen, dass die Leistungen unserer Schüler international Mittelmaß (für die SII s. T3.1, 140) und lediglich in der Grundschule etwas besser sind. – Ganz wichtig ist insbesondere die Aufmerksamkeit, die Bildung, speziell auch mathematische Bildung, in Politik und Öffentlichkeit gewonnen hat.

Vor allem aber hat uns PISA die Versäumnisse beim schwachen Viertel unserer Heranwachsenden vor Augen geführt. Diese Versäumnisse sind jedoch nicht nur schulischer, sondern gesamtgesellschaftlicher Natur,

Unauthenticated
Download Date | 3/18/19 11:36 PM

und nicht ausschließlich, aber zu einem großen Teil bestehen sie aus mangelnder Integration der Familien mit Migrationshintergrund.

Literatur

(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, PISA) 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

(P2, PISA) 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

(P3, PISA) 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

(I1, IGLU) Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther, 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

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

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 2002, Heft 4, 49–58

Knoche, Norbert, Detlef Lind, Werner Blum, Elmar Cohors-Fresenborg, Lothar Flade, Wolfgang Löding, Gerd Möller, Michael Neubrand, Alexander Wynands (Deutsche PISA-Expertengruppe Mathematik, PISA-2000) (2002): Die PISA-2000-Studie, einige Ergebnisse und Analysen. In: Journal für Mathematik-Didaktik 23, 159–202

Meyerhöfer, Wolfram (2004): Das Kompetenzstufenmodell von PISA – eine empirische Dekonstruktion. Erscheint in: Beiträge zum Mathematikunterricht 2004. Hildesheim: Franzbecker

Neubrand, Michael (2003): „Mathematical literacy" / „Mathematische Grundbildung": Der Weg in die Leistungstests, die mathematikdidaktische Bedeutung, die Rolle als Interpretationshintergrund für den PISA-Test. In: Zeitschrift für Erziehungswissenschaft 6, 338–356

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

Adresse des Autors

Prof. Dr. Peter Bender
Fachgruppe Mathematikdidaktik
Universität Paderborn
Warburger Straße 100
33098 Paderborn
`bender@upb.de`

Unauthenticated
Download Date | 3/18/19 11:36 PM