

PETER BENDER

## **Die etwas andere Sicht auf die internationalen Vergleichs- Untersuchungen TIMSS, PISA und IGLU**

Herrn Prof. Dr. Heinrich Winter zum 75. Geburtstag gewidmet

### **0. Einstimmung**

Es lag nahe, dass das Paderborner Lehrerbildungszentrum (**PLAZ**) im Zuge seiner Vortragsreihe im Sommersemester 2003 zur IGLU-Studie mich als Paderborner "Eigengewächs" und zugleich Mitglied der IGLU-Arbeitsgruppe "Mathematik-Didaktik" zu einem Vortrag einlud.

Als Thema hatte ich "Die Ergebnisse der IGLU-Studie in Mathematik — Was sagen sie uns für die Arbeit (nicht nur) in der Grundschule?" vorgesehen. Während jedoch noch bei TIMSS (1997; T2, 68) die Daten interessierten Wissenschaftlerinnen & Wissenschaftlern zur Verfügung gestellt werden sollten, war bei PISA fünf Jahre später davon keine Rede mehr, und nun wurden sie sogar mir als Mitglied der genannten IGLU-Arbeitsgruppe vor-enthalten. Eigentlich hatte ich anhand konkreter Analysen konkreter Schülerinnen- & Schüler- (S&S-) Bearbeitungen konkreter Aufgaben Hinweise für die konkrete Arbeit im Mathematik-Unterricht (MU) geben wollen. Stattdessen musste ich meinen Vortrags-Schwerpunkt und damit den Titel erheblich verändern und mich hauptsächlich mit der Statistik auseinandersetzen, da mir ja nur diese zur Verfügung steht.

Natürlich kann mit den Daten Unfug getrieben werden und ist in der Debatte über die PISA-Statistiken bereits intensiv getrieben worden. Meine persönlichen "**Top Five**" lauten:

**5. Platz:** "Weil Finnland bei PISA so gut abgeschnitten hat und in Helsinki alle Schulen 'am Netz sind', müssen auch in Deutschland '**alle Schulen ans Netz**'."

Die Forderung hat durchaus etwas für sich; die Begründung taugt jedoch nichts.

**4. Platz:** "PISA hat gezeigt, dass das **dreigliedrige Schul-System von Nachteil** ist; denn in fast allen Ländern, die besser als Deutschland abgeschnitten haben, ist die Einheits-Schule die vorherrschende Schul-Form."

Ich bin zwar kein glühender Verfechter des dreigliedrigen Schul-Systems, muss aber darauf aufmerksam machen, dass auch so gut wie alle Länder, die **schlechter** als Deutschland abgeschnitten haben, über das System der **Einheits-Schule** verfügen und dass noch bei dem Bericht über **TIMSS betont** wurde, dass aus den Ergebnissen **kein Rückschluss auf** die Leistungs-Fähigkeit der **Schul-Systeme** als solche gezogen werden könne (T2, 18f). Womöglich hatte man die deutsche **Gesamtschule** ein wenig in Schutz nehmen wollen, die ja — zu meiner damaligen großen Verblüffung — insgesamt klar unter der Realschule rangiert.

**3. Platz:** "Zwar hat unter den Bundesländern **Bayern** in fast allen Belangen am besten abgeschnitten; aber trotzdem ist Bayern zu **kritisieren**: (a) Der **Abstand** zu den anderen Ländern ist gar nicht so groß, z.B. sind die **schleswig-holsteinischen Gymnasien in Naturwissenschaften** besser als die bayrischen; und international sind einige Länder immer noch viel besser als Bayern; (b) außerdem führen die **anderen Bundesländer** relativ viel **mehr Jugendliche zum Abitur** als Bayern."

Ich bin zwar kein Bayern-Fan (im doppelten Sinn), aber an dem (im deutschen Maßstab) **guten Abschneiden** dieses Bundeslands gibt es **nichts zu deuteln**. Der besondere deutsche Mangel, nämlich das katastrophale Niveau der schwächsten 25%, ist in Bayern geringer ausgeprägt. Und dass die anderen Länder mehr Jugendliche trotz schwächerer Leistungen zum Abitur führen, fordert eher zur Skepsis diesen gegenüber heraus.

**2. Platz:** In Frankfurt **missbrauchte** Ende April 2003 die Bildungs-Dezernentin die **PISA-Ergebnisse**, um **für die Schließung von Stadtteil-Büchereien** zu plädieren.

Selbstverständlich ist der Zusammenhang nur ein indirekter, indem die Konzentration von Finanzmitteln auf bestimmte Projekte und damit die Entfinanzierung anderer Projekte mit PISA legitimiert werden soll. — Aber genau das stört mich, nämlich dass **PISA** für alle möglichen, z.T. kontraproduktiven Maßnahmen **als Begründung herhalten** muss.

**1. Platz:** Den Vogel schoss wohl der ehemalige Ministerpräsident eines großen Bundeslandes und inzwischen Mitglied der Bundesregierung ab, als er

behauptete, **PISA** habe gezeigt, dass die **Schulzeit bis zum Abitur von 13 auf zwölf Jahre gesenkt** werden müsse.

Im Vortrag werde ich folgende Thesen entwickeln:

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

**2. These:** Es **steht dahin**, ob (a) dem im Angelsächsischen entwickelten Konstrukt der **Mathematical Literacy**, jedenfalls in der in den Studien realisierten Ausprägung, wirklich **die** fundamentale Rolle im MU zukommen soll, (b) die Test-Aufgaben für dieses Konstrukt **valide** sind, (c) der ab PISA erklärte **Verzicht** auf die Kontrolle von nationaler **Curriculums- und Unterrichts-Validität** die Studien nicht doch entwertet, (d) bei **noch so guten Übersetzungen** nicht doch diejenigen Kinder es leichter haben, in deren **Mutter-Sprache** die Aufgaben ursprünglich formuliert worden waren, nämlich in **Englisch**.

**3. These:** (a) Es **steht dahin**, ob die statistischen und organisatorischen **Vorgaben** wirklich weltweit hinreichend **skrupulös eingehalten** wurden. (b) **Darüber hinaus** gibt es **Variablen**, die die Ergebnisse von 9- bis 10-Jährigen beeinflussen, aber nichts über die Unterrichts-Qualität aussagen, z.B. Schul-Eintritt mit 5 oder mit 6 Jahren. (c) **Kleine Änderungen in den Punkt-Zahlen können große Änderungen in den Platz-Ziffern** nach sich ziehen. (d) Die **Fragwürdigkeit** der engen **Statistik-Gläubigkeit** zeigt sich daran, dass die völlig **unterschiedlichen Ergebnisse** einiger Länder bei TIMSS II und PISA, aber auch bei PISA und IGLU, **letztlich unerklärlich** sind.

**4. These:** (a) Für die **deutsche Öffentlichkeit** waren die **schockierenden Ergebnisse heilsam**, da die Relativierungen 1.–3. nichts an den Grund-Tendenzen ändern. (b) Für Jemanden mit Einblick in die Schul-Realität waren sie **keine Überraschung**: Wer wissen wollte, konnte auch schon vorher wissen. (c) **Hoffentlich** wird der ganze **Aufwand** durch den Erfolg von in die Wege geleiteten und noch zu leitenden Maßnahmen **gerechtfertigt** (Gegen-Beispiel der o.a. 2. Platz).

**5. These:** Viel mehr als über das Bildungs-System **sagen die Statistiken** jedoch etwas über gewisse Mangel-Erscheinungen unserer **Gesellschaft** aus, besonders die **schlechte Integration der Familien mit Migrations-Hintergrund** (MH; mindestens ein Elternteil nicht in Deutschland geboren), die nicht allein innerhalb des Bildungs-Systems repariert werden können, wenn überhaupt.

Aus Zeitmangel (im Vortrag) und Platzmangel kann ich die Thesen nicht alle mit der gebotenen Gründlichkeit belegen. Hierzu ist eine ausführlichere Ausarbeitung geplant. Hinweisen möchte ich an dieser Stelle auf einen Kurzbericht von Schmidt & Fertig (2003) über eine Re-Analyse der individuellen Daten-Sätze der deutschen Teilnehmerinnen & Teilnehmer, die tiefere (und z.T. andere) Einsichten ermöglicht als die PISA-übliche Gründung der Erkenntnisse auf Durchschnittswerte (vor allem von Ländern).

## 1. Wie ist IGLU konzipiert?

### 1.1 Grundlegendes

Die erste internationale Vergleichs-Studie zu Schul-Leistungen, die bei uns Beachtung gefunden hat, war TIMSS (**Third International Mathematics and Science Study**), bei der **drei Populationen untersucht** wurden, grob gesagt, die 3./4., die 7./8. und die 12. Jahrgangsstufe (T2, 44, T3.1, 31ff). **Deutschland** machte nur in den **Sekundarstufen** mit. Die Leistungen der **deutschen 13–14-Jährigen**, die 1994/95 erhoben worden waren, führten bei ihrer Publikation 1997 (T2) zu einem ersten **Schock**; die der 17–18-Jährigen (2000, T3.1, T3.2) waren etwas besser.

Inzwischen wurde **PISA** begonnen, in dem weltweit die **Reading, Mathematical und Scientific Literacy** (Publikationen bis jetzt P1, P2, P3, 2001ff) bei den i.W. **15-Jährigen** (P1, 34f) in den Jahren 2000, 2003 und 2006 (P1, 11) untersucht werden. Als die Ergebnisse von PISA 2000 Anfang Dezember 2001 der deutschen Öffentlichkeit vorgestellt wurden, saß der **Schock noch viel tiefer**, weil sie medial besser aufbereitet und noch schlechter waren als die von TIMSS.

Die **Internationale Grundschul-Lese-Untersuchung (IGLU)** untersucht die Lese-Fähigkeit der Viertklässlerinnen & -klässler (V&V). In Deutschland wurde sie u.a. um Mathematik und Naturwissenschaften erweitert (**IGLU-E**). In diesen beiden Disziplinen wurde durch Aufnahme entsprechender Aufgaben die IGLU-Studie an **TIMSS angebunden** (IG, 2). Um weitere Forschungen zu ermöglichen, wurden **zahlreiche zusätzliche Aufgaben** gestellt. Lei-

der haben sich aus finanziellen Gründen die Länder Brandenburg, Mecklenburg-Vorpommern, Niedersachsen und Sachsen-Anhalt aus IGLU-E ausgeklinkt (IG, 195ff).

Weltweit nahmen knapp 150.000, in Deutschland 7.633 Kinder teil, davon knapp 6000 bei IGLU-E (IG, 7ff). Außerdem gab es **Fragebögen**, mit denen **zahlreiche Merkmale** erhoben wurden wie MH, sozialer Status, Bildungsnähe des Elternhauses (Anzahl der Bücher zu Hause; — IG, 50) usw., insgesamt das sog. "**soziale und kulturelle Kapital**" (IG 270, P1, 326).

## 1.2 Skalierung der Schülerinnen- & Schüler-Leistungen und der Aufgaben-Schwierigkeitsgrade

Für jedes Kind wurde die Zahl der richtigen Antworten im Mathematik-Test festgestellt. Für den **gesamten Datensatz**, bestehend aus den Daten aller Kinder, wurden das arithmetische Mittel  $\bar{a}$  und die Standard-Abweichung (SA)  $s$  errechnet, und der Datensatz wurde so transformiert, dass das **arithmetische Mittel dann 500 und die SA 100** betragen. Hat ein Kind also  $p$  Punkte erreicht, so beträgt seine **transformierte Punkt-Zahl**  $100 \cdot (p - \bar{a}) / s + 500$  (P1, 48ff). Man kann nun sämtliche 150.000 Kinder in eine Rangfolge bringen, und man kann für jede interessierende Teil-Population (insbesondere für jedes Land) deren Mittelwert und SA berechnen.

Über die **absolute Leistungsfähigkeit** eines Kindes oder eines Landes sagen diese Zahlen jedoch gar **nichts**. **Je mehr Schwache** insgesamt mitmachen, umso niedriger wird der Mittelwert und **umso eher bzw. höher liegt man** mit seiner Punktzahl über diesem, d.h. dann über 500. Dies könnte z.T. erklären, warum Deutschland bei PISA mit 490 (P1, 173) noch schlechter abgeschnitten hat als bei TIMSS II mit 509 fünf Jahre zuvor (T2, 90): Es fehlten nämlich z.B. Südafrika, Kolumbien, Kuwait, Iran, und nur Brasilien und Mexiko kamen dazu.

Mit Hilfe der Punkt-Zahlen für die S&S wurden zugleich **Schwierigkeits-Grade für die Aufgaben** konstruiert (P1, 517ff). **Je weniger S&S** eine bestimmte Aufgabe lösen, **desto mehr Punkte** erhält diese, genauer: Für jede Punkt-Zahl  $p$  wird die Schicht  $S_p$  aus allen S&S mit dieser Punkt-Zahl gebildet. Jeder **Aufgabe** wird dann **diejenige Punkt-Zahl  $p$**  zugewiesen, wo 62% der S&S der Schicht  $S_p$  sie lösen. (Dabei wird — keineswegs denknotwendig — unterstellt, dass die S&S einer Schicht  $S_q$  mit  $q > p$  bei **jeder** Aufgabe einen höheren Lösungs-Anteil als die in  $S_p$  erzielt haben.) Auch die Aufgaben erhalten somit Punkt-Zahlen um 500.

## 2. Was gibt die Statistik von IGLU (und PISA) her?

### 2.1 Deutschland bei IGLU

Deutschland lag bei **IGLU in Mathematik bei 545** Punkten (IG, 209) und in den **Naturwissenschaften sogar bei 560** (IG, 166), also deutlich über 500 Punkten. Außerdem lag die **SA mit 81** (IG, 209) bei den geringeren, während sie bei PISA mit 103 (P1, 173) zu den höheren gehört hatte.

In der Öffentlichkeit war man mit diesem Ergebnis **einigermaßen zufrieden**, aber: (a) Im **Vergleich zu Österreich 1995 mit seinen 559** (IG, 209) Punkten hatten es unsere S&S mit einigen kindgemäßerer Formulierungen doch leichter. (b) Man kann durchaus unterstellen, dass so **manches Land im Jahr 2001 besser** abgeschnitten hätte, als es im Jahr 1995 abgeschnitten hatte, warum auch immer. (c) (und hauptsächlich:) Man darf den **deutschen Wert** von 545 nicht zum Mittelwert 500 in Bezug setzen, da in diesem ja auch die Leistungen der damaligen Drittklässlerinnen & -klässler enthalten sind. Da in Deutschland 2001 **nur die V&V** untersucht wurden, muss man sie mit der V&V-Teilpopulation von 1995 vergleichen. Deren Wert betrug **529** (IG, 213), und entsprechend käme man bei TIMSS I auch bei den anderen Ländern zu deutlich höheren Punktzahlen.

### 2.2 Noch einmal ein Blick auf TIMSS und PISA

Nicht nur die **neuen deutschen Bundesländer**, auch **ost-europäische Staaten wie Tschechien, Ungarn und Russland** haben von 1995 bis 2000 erheblich verloren. Der gesellschaftliche Wandel dort hat in den 1990-er Jahren die Schulen voll erfasst und, pauschal gesprochen, zu einem **Straffheits-Abbau** an Schul-Organisation und im Unterricht verbunden mit einer **Lässigkeits-Zunahme** bei allen Beteiligten geführt.

Interessanter Weise **erklären die PISA-Leute** diesen Trend gerade **umgekehrt** mit der Perpetuierung der "traditionell dort vorherrschenden Methode eines stark lehrergesteuerten, auf Kenntnis mathematischer Fakten ausgerichteten Unterrichts" (P1, 177). Hier möchte ich nur **Franz E. Weinert** (1996) zitieren, der als einer der weltweit ganz Wenigen über **breite Feld-Untersuchungen** verfügt und zu dem Ergebnis gekommen ist: "Während viele 'neue Lerntheorien' die Bedeutung der intrinsischen Motivation und die aktive, konstruktive und selbständige Rolle des Lernenden betonen und dem Lehrenden nur noch eine anregende, beratende und moderierende Funktion zuschreiben, zeigen alle verfügbaren Unterrichtsstudien die **Wichtigkeit einer lehrergesteuerten, aufgabenorientierten und effektiven Instruktion** ... gu-

te Lehrer machen jedoch durch ihren Unterricht nicht passiv, sondern sorgen dafür, dass möglichst alle Lernenden in einer je geeigneten Weise sich mit der Lernaufgabe aktiv auseinandersetzen."

Gerade **in den ost-asiatischen Ländern** scheint mir dieses realisiert zu sein, wie sich in der TIMSS-Video-Studie in Japan, Deutschland und den USA andeutet (T2, 215ff). Allerdings erkläre ich das Leistungs-Vermögen der S&S in Ländern wie Japan, Korea und Singapur vor allem mit der starken **Leistungs-Orientierung** der Gesellschaft insgesamt, die ja schon zu einem enormen Leistungs-Druck auf die Heranwachsenden bis hinab zu Kindergarten-Kindern führt. Man muss wissen, ob man einen solchen Leistungs-Druck haben möchte (abgesehen davon, dass er in den europäischen und europa-affinen Gesellschaften wohl nicht durchzusetzen wäre), und **wenn nicht**, muss man sich **mit geringeren Leistungen zufrieden** geben.

Festzuhalten ist darüber hinaus, dass in **Japan und Korea** der **Anteil von S&S mit MH gering** ist und dass z.B. in Singapur die IGLU-Lese-Leistung der V&V, deren beide Eltern im Ausland geboren sind, höher ist als die Leistung der V&V ohne MH (IG, 296) (ebenso z.B. in Neuseeland, Hong Kong, Kuwait).

### **2.3 Kinder mit niedrigem Sozialstatus und mit Migrations-Hintergrund**

Die Daten von TIMSS, PISA und IGLU liefern uns über den Schul-Bezug hinaus wichtige Einblicke in den **Aufbau unserer Gesellschaft**: Insgesamt haben knapp **ein Viertel** (22%, in West-Deutschland über 25%; IG, 269) aller V&V in Deutschland **MH**, in west-deutschen Großstädten (>300.000) im Durchschnitt 36%, in der alten DDR nur knapp 4% (P2, 190).

Kinder mit MH erbringen **im Mittel erheblich schlechtere Leistungen**, z.B. bei den V&V in Lesen 55 Punkte weniger als die anderen (IG, 296), ein Unterschied, mit dem Deutschland in der Spitzengruppe liegt. Der wichtigste Einfluss-Faktor ist die **Verkehrs-Sprache in der Familie und in den Peer-Groups** und damit Beteiligung an der Kultur des aufnehmenden Landes. Die **Familien mit MH** werden in **Deutschland zu schlecht integriert**. Viele europäische und überseeische Länder stehen vor ähnlichen Herausforderungen. Allerdings haben es die **alten Kolonial-Mächte leichter**, weil die Menschen aus deren ehemaligen Kolonien schon vor der Zuwanderung eine Affinität zu deren Sprache und Kultur haben. Überhaupt geht es mit dem Englischen als Welt-Sprache leichter. Und in PISA-Musterländern wie Japan und Finnland gibt es relativ wenige Familien mit MH. Wenn man aus den deutschen Er-

gebissen einmal die **Familien mit MH herausrechnet**, wird der Leistungs-**Durchschnitt deutlich besser** ( 513 gegenüber 490 in PISA-Mathematik; P3, 245). Aber bei den anderen Ländern muss man dann zum Zwecke des Vergleichs ähnliche Korrekturen vornehmen.

Die **Probleme** entstehen vor allem bei Familien, wo **beide Elternteile im Ausland** geboren sind. Da wo **nur ein Elternteil im Ausland** geboren ist, **fallen die Leistungen gar nicht so stark ab** (P1, 378). — Dann gibt es **Bundesländer**, wo die Leistungen der S&S mit MH besser als in anderen sind. Man muss hier **Bayern** nennen, wenn man alle drei Domänen (Lesen, Mathematik, Naturwissenschaften) zusammen betrachtet (P2, 202). Mit seinem Anteil von etwa 22% liegt es etwas unter dem der westdeutschen Flächenländer und hat es daher leichter. Hinzu kommt das höhere Niveau der Hauptschule dort.

Im Gegensatz dazu gibt es in **Ländern wie Nordrhein-Westfalen** und **Bremen** mit Quoten von 32% bzw. 40% viele Schulklassen mit einem so hohen Anteil von S&S mit MH, dass sowohl die mit, als auch die ohne MH schlecht gefördert werden.

Die zweite wesentliche Ursache für die schwachen Leistungen der **S&S mit MH** ist der **soziale Status** dieser Familien mit sozial "niedrigeren Berufen" bzw. Sozialhilfe-Bezug (P1, 344). Diese Ursache ist natürlich für **S&S ohne MH genau so wesentlich**. Aber sie tritt in Deutschland bei denen mit MH naturgemäß besonders häufig auf.

Hier hat die **deutsche Gesellschaft**, und keineswegs nur die deutsche Schule, bis jetzt i.W. **versagt**. Allerdings sehe ich gerade zurzeit und in nächster Zukunft in Anbetracht der galoppierenden Arbeitslosigkeit und des unübersehbaren Qualitäts-Verlusts unseres Sozial-Systems eher zunehmende Schwierigkeiten.

Wenigstens in Sachen "**Erwerb der deutschen Sprache**" werden seit einiger Zeit **Programme** gestartet, und dies ist nicht zuletzt der PISA-Studie zu verdanken. Mit Recht setzt man hier schon in der **Vorschul-Zeit** an, und die Kritikerinnen & Kritiker sollten sich jetzt nicht wieder für die Freiheit, das Deutsch-Lernen zu verweigern, stark machen. Auch die **Ganztags-Schule** ist, jedenfalls unter dem Aspekt der Integration, ein Schritt in die richtige Richtung. Allerdings fallen da eine Menge Kosten "für Ausbau, Einrichtung und Renovierung" an, während "**zusätzliche Lehrerstellen nicht gefördert werden**" (Westfälisches Volksblatt vom 12.06.2003 über die "erste 'Offene Ganztagschule' in Paderborn").



## 2.4 Struktur-Veränderungen im Schul-System?

Als Beispiel für eine **besser gelungene Integration** und überhaupt geringere Leistungs-Streuung bei höherem Niveau, bei gesellschaftlichen Verhältnissen, die den unseren sehr ähnlich sind, wird gerne auf **Schweden** verwiesen (P1, 350) und als Begründung die dort realisierte **Einheits-Schule sowie die Schul-Autonomie** bis hin zu den Lehrplänen herangezogen (noch intensiver bei dem noch besseren **Finnland**).

Ich stelle dahin, ob wirklich diese **schul-strukturellen** Unterschiede etwas bewirken, denn: aus Finnland wird berichtet, dass man dort für das Lehramts-Studium die **Auswahl aus den allerbesten** Abiturientinnen und Abiturienten hat. Oder: Für ein schwedisches Grundschul-Kind wird jährlich ein 50% **höherer Betrag des Brutto-Inlandprodukts pro Kopf** aufgewendet als für ein deutsches. Und: während **in Deutschland** die Quote der V&V, die schon einmal **eine zusätzliche Lehrkraft** im Unterricht erlebt haben, bei einsam niedrigen 38% liegt, beträgt diese Quote bei Ländern wie England, Kanada, Schottland, Schweden, USA um die 90% (IG, 34).

Allerdings ist dort auch das **Gehalts-Niveau** im Schul-Dienst deutlich **niedriger** als bei uns. Wenn wir das unsere erhalten wollen (und ich lehne ein **Opfer** der deutschen Lehrerinnen & Lehrer — L&L — strikt ab), dann müssen wir uns damit abfinden, dass wir ein System wie in Schweden oder Finnland nicht finanzieren können.

Überraschender Weise haben viele **gute Länder im Mittel erheblich größere Klassen** als wir, schon bei PISA Japan 39, Korea 38 gegenüber Deutschland 24 (P1, 422), und bei den V&V England 29, Neuseeland 28, Niederlande, Kanada, Schottland 26 gegenüber Deutschland 22,5 (IG, 34). Zwar ist mit kleineren **Klassen** bessere Förderung möglich, aber andere Länder schaffen das auch mit größeren Klassen.

Man kann auch prüfen, ob man, wie z.B. in England, Schottland, Australien und Neuseeland, die **Schule schon ein Jahr früher beginnen** lässt (T1, 13, 24, 28). Ich persönlich bin **gegen** diese frühe Einschulung: Das **Arbeitsplätze-Volumen** hierzulande ist so **gering** (und wird noch geringer), dass überhaupt **kein Anlass** besteht, durch frühe Einschulung, Vorverlegung des Abiturs, Verkürzung des Studiums unsere Jugend früh auf den Arbeitsmarkt und damit verstärkt in die Arbeitslosigkeit zu entlassen. **Lediglich der Abbau sprachlicher Defizite**, die übrigens zunehmend auch bei Kindern ohne MH zu beobachten sind, sollte **als absichtsvoller Unterricht (natürlich in adäquater Form) schon im Alter von unter 6 Jahren** stattfinden.

Ich bin durchaus ein **Anhänger von Leistung** und erblicke **die Haupt-Ursache** für das ausgeprägt schwache Abschneiden des schwachen Viertels unserer S&S gerade in einer **Distanz zur Leistung in relevanten Gruppen** unserer Gesellschaft. Zum einen wirkt die 68er-Studentinnen- & -Studenten-Bewegung (von der ich übrigens aktives Mitglied war und deren Ideale ich heute noch hochhalte) mit ihrem **anti-autoritären Prinzip** (gegenüber Bildungsinstitutionen, deren Vertreterinnen & Vertretern sowie den Fächern) in Teilen der **Pädagogik**, kurz gesagt, in einer leistungs-fernen Grund-Atmosphäre nach. Zum zweiten haben die **Medien** im Großen und Ganzen nicht gerade einen den Leistungs-Gedanken fördernden Einfluss auf unsere Heranwachsenden. Zum dritten sehen unsere Jugendlichen, besonders die mit schlechten Schul-Leistungen, für sich **wenig Zukunfts-Perspektiven**. Diese Stimmung kann man nicht mit Schulleistungs-Tests erfassen, aber z.B. mit den **Schul-Schwänzerinnen- & -Schwänzer-Zahlen**: Die Bertelsmann-Stiftung geht hier von einer halben Million "Schulmüder" vor allem in Haupt- und Sonderschulen aus, die wöchentlich mehrere Stunden unentschuldigt fehlen (Frankfurter Rundschau vom 17.05.2003).

Die L&L sind damit **überfordert**; vielmehr bedürfte es eines **gewaltigen Rucks der ganzen Gesellschaft**; aber dafür erkenne ich, ehrlich gesagt, keine Anzeichen. Vielleicht tut man sich mit Tugenden wie Leistung, Disziplin usw. **nirgends so schwer wie in Deutschland**. Hier hängt uns unsere Vergangenheit nach, wo die maßlose Verzerrung solcher Tugenden die Voraussetzung für **ungeheure Verbrechen** legten.

## 2.5 Insbesondere Beseitigung des mehrgliedrigen Schul-Systems?

Wie schon gesagt, lässt sich aus den Ergebnissen von **TIMSS und PISA keine Überlegenheit der Einheits-Schule** ableiten; denn nicht nur viele Länder, die besser, sondern auch viele, die schlechter als Deutschland abgeschnitten haben, verfügen über dieses System, und umgekehrt haben Länder wie **Österreich, Schweiz, Belgien** mit mehrgliedrigem Schul-System (P1, 425) durchaus passable Leistungen gezeigt. Im **innerdeutschen Vergleich der Schulformen** schneidet die integrierte Gesamtschule mit 455 Punkten in PISA-Mathematik (gegenüber Gymnasium 574, Realschule 501, Hauptschule 439; P3, 273) überraschend schlecht ab.

Man muss allerdings **hinter die Statistik schauen**: Es ist klar, dass, so lange das Gymnasium existiert, die **Gesamtschule keine wirkliche Chance** hat, ihre Leistungsfähigkeit unter Beweis zu stellen. Allerdings ist genau so klar, dass da, wo die Gesamtschule besteht, wiederum die **Hauptschule keine Chance** hat. Auf jeden Fall muss die **Misere der Hauptschule beendet** wer-

den, sei es durch eine **Gesundungs-Rosskur**, sei es durch ihre **Abschaffung**. Ob die **Gesellschaft** zu einer dieser beiden gewaltigen Anstrengungen **in der Lage** ist, sei **dahin gestellt**.

Was in **Finnland geklappt** hat, muss **bei uns noch lange nicht** klappen. Dieses kleine Land, mit einer deutlich homogeneren Gesellschaft, hat in einem Kraftakt in den 1970-er Jahren sein Schul-System von Mehr- auf Eingliedrigkeit umgestellt. Wir dagegen haben **16 eigenwillige Bundesländer**, im Durchschnitt jedes so groß wie Finnland, eine viel **heterogenere Gesellschaft**, und wir sind **30 Jahre weiter** von früheren idyllischen Zuständen entfernt als Finnland damals. — Vor allem aber: **mit Struktur-Änderungen allein**, auch noch **gegen den Willen vieler Beteiligten**, hat es noch **nie Fortschritt** gegeben.

### 3. Was sollte mit welchen Aufgaben getestet werden?

#### 3.1 Das Idealbild eines Tests mathematik-bezogener Kompetenzen und die Realität

Welche mathematik-bezogenen **Fähigkeiten, Fertigkeiten, Wissens-Bestände, Einstellungen** usw. hält man für wichtig, so dass man mit dem Grad ihres Vorhandenseins mathematische **Leistungs-Fähigkeit** eines Individuums oder einer ganzen Population bestimmt? **Wie misst** man diese **Tugenden**? — Üblicher Weise 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 mathematik-didaktischen Forschungs-Ergebnisse**.

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

Das Haupt-Interesse der IGLU-Arbeitsgruppe "Mathematik-Didaktik" richtete sich — unabhängig von internationalen oder nationalen Vergleichen — auf den Entwurf und die Erprobung **eigener Aufgaben**, bei denen es uns **weniger um direkte Curriculum-Validität**, sondern um das **Ausprobieren** ge-

wisser anspruchsvollerer Fragestellungen u.a. als Anregung für die **Weiter-Entwicklung** der Grundschul-Lehrpläne, um den **Vergleich mit älteren deutschen Tests**, z.B. mit meinen alten Sachrechen-Tests (wie die gleich noch zu zeigende Kakao-Aufgabe; s. Bender 1980), usw.

### 3.2 Mehr oder weniger gute Aufgaben-Formulierungen

Außerdem hatten wir uns vorgenommen, die Aufgaben des **internationalen** Tests für deutsche V&V adäquat, d.h. kindgemäß zu **formulieren**. Als Vorlage hatten wir die Aufgaben für die österreichischen Kinder von 1995. Abgesehen von österreich-typischen Wendungen wie "Buben" oder "Milchpackerl" strotzten die Formulierungen von irreführenden, ungenauen, kindfremden Wendungen, so dass ich bei weit über der Hälfte der 102 Aufgaben Änderungen anzubringen hatte, Beispiele:

"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, , , . 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 österreichischen Kinder 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. In meinen eigenen Untersuchungen z.B. habe ich durch Variation von Sprach-Wendungen oder nur des Layouts deutliche Effekte erzielt.

Leider konnten wir die Formulierungen nicht in voller Konsequenz verbessern, damit die Vergleichbarkeit einigermaßen erhalten blieb, aber bei uns lauteten die o.a. Fragen (immer noch nicht ideal) so:

"H6: Hier ist ein Rechteck mit den Maßen 6 cm und 4 cm. Die Länge rundherum nennt man seinen Umfang. (*Zeichnung*) Wie groß ist der Umfang des Rechtecks in Zentimetern?"

Vergleichbar mit H8: "Hier wurde zum Aufschreiben der ersten sechs Zahlen eine bestimmte Regel benutzt. Benutze diese Regel, und schreibe die nächsten zwei Zahlen in die freien Felder.

5,      7,      9,      11,      13,      15,      \_\_\_\_,      \_\_\_\_"

"A3: Hier sind 9 Schachteln mit Stiften. In jeder Schachtel sind 125 Stifte. Wie viele Stifte sind es insgesamt? ..."

Im englischen Original hatte diese Aufgabe so gelautet: "A3: There are 9 boxes of pencils. Each box has 125 pencils. What is the total number of pencils? ..."

So ganz **astrein** waren die **originalen** Formulierungen von H6 und H8 auch **nicht** gewesen. Aber die alte deutsche Übersetzung war bei fast allen Aufgaben noch schlechter. — Bei allen Studien wurde übrigens betont, wie **sorgfältig** man das **Übersetzen** betrieben habe.

Vielleicht ist ja das Übersetzungs-Niveau bei PISA und IGLU gegenüber TIMSS gestiegen. Aber **Treffer Sicherheit und Durchsichtigkeit** von Formulierungen, die für die S&S einer Sprache und Kultur optimiert sind, bleiben in aller Regel bei noch so guter Übersetzung in eine andere Sprache **nicht voll erhalten**. Dies gilt natürlich in besonderem Maße für die Texte der Lese-Untersuchungen. Bei der Formulierung sparsamer mathematischer Texte von der Forschungs-Front bis zur Primarstufe scheint das lapidare Englisch gegenüber anderen Sprachen allerdings besonders gut geeignet zu sein.

Was ich bei unseren IGLU-Aufgaben leider zu spät gemerkt habe, ist folgende durchgängig **falsche Schreibweise von Brüchen** (*ohne Leertasten vor und nach den Klammern*): "10 Liter Kakao werden in kleine Flaschen von je einem Viertel( $\frac{1}{4}$ )Liter Inhalt gefüllt. Wie viele Flaschen werden voll?" — Die vorhin gezeigte Schreibweise der Multiplikation mit dem fetten Punkt kann übrigens bei manchen Kinder dazu führen, dass sie die Operation nicht erkennen (was sie hier allerdings gerade vor Fehl-Ankreuzungen bewahrt).

Ein treffendes Beispiel, **zur Abwechslung** einmal 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 PISA-Aufgaben, die ich durchweg teile. Jetzt möchte ich nur das hier

bedeutungs-widrig verwendete Wort "**symmetrisch**" kritisieren, das zumindest für deutsch-sprachige S&S nicht zielführend ist.

### 3.3 Was testen die Aufgaben eigentlich wirklich ab?

Wie bei allen Tests wird auch bei TIMSS, PISA und IGLU ganz wesentlich eine extrinsische Fähigkeit der S&S abgeprüft, nämlich **herausfinden** zu können, **was die Aufgaben-A&A wohl gemeint haben**. Diese Herausforderung ist bei inner-mathematischen Aufgaben naturgemäß geringer, aber bei den Aufgaben mit einem irgendwie gearteten außer-mathematischen Kontext beliebig schwierig (und wird **durch misslungene Formulierungen bzw. Übersetzungen noch verschärft**); denn die aufgeworfenen Probleme sind ja nie die der S&S. Da sind natürlich wieder **Diejenigen im Vorteil**, die an solche (standardisierten, insbesondere von Fremden gestellten) Tests gewöhnt sind und in deren Sprache und Kultur die Aufgaben ursprünglich angesiedelt sind.

Wenn in Schweden die Lösungs-Häufigkeit der Zylinderwickel-Aufgabe mit 24% sechsmal so hoch ist wie in Frankreich, so ist das für mich als Geometrie-Didaktiker mit einem besonderen Interesse an der Raum-Geometrie zunächst einmal ein Indiz, dass im schwedischen Curriculum **raum-geometrische Aufgaben**, womöglich von diesem speziellen Typ, **eher behandelt werden als** im Geometrie-Mutterland Frankreich, wo allerdings bekanntlich der Geometrie-Unterricht sowieso stark algebraisiert ist und sich hauptsächlich auf die affine Ebene beschränkt.

Hier drängt sich die Frage nach der **Unterrichts- und Curriculums-Validität** auf. In der Tat habe ich in allen Tests keine Aufgabe gefunden, die nicht irgendwie auf das deutsche Mathematik-Curriculum der jeweiligen Stufe bezogen und unterrichtlich behandelbar wäre. Das gilt auch für die **Zylinderwickel-Aufgabe in der deutschen Hauptschule**, allerdings erst im 10. Schuljahr: Der Pythagoras-Satz ist dort Thema, und wenn man die Fragestellung versteht und die Idee mit dem vierfachen Abwickeln hat, kann man die Aufgabe lösen.

Allerdings ist von entscheidender Bedeutung, ob die curriculare **Validität in den verschiedenen Ländern tatsächlich etwa in gleicher Weise** erfüllt ist. Das ist in IGLU eher, in PISA weniger gegeben, wie **zahlreiche Expertinnen & Experten** aus den 16 Bundesländern und den verschiedenen Schulformen bei Prüfung des Aufgaben-Materials (vermischt mit anderen Aufgaben, so dass sie nicht wussten, welche für PISA vorgesehen waren) festgestellt haben (P1, 43f).

### 3.4 Das "Mathematical-Literacy"-Konzept auf Kosten der curricularen Validität

Während bei TIMSS die Validität in Deutschland noch weitgehend vorhanden war, wird bei den **PISA-Tests ein "Verzicht auf transnationale curriculare Validität"** postuliert; 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 MU (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, **von PISA "bevorzugt"** werden. Für Deutschland trifft das im Großen und Ganzen nicht zu, aber m.W. z.B. für die angelsächsischen und skandinavischen Länder sowie die Niederlande.

Wie den meisten Kolleginnen & Kollegen aus der deutsch-sprachigen mathematik-didaktischen Kommunität **sagt mir das Grundbildungskonzept** (als Übertragung des anglo-amerikanischen Begriffs "**Mathematical Literacy**", ML) von PISA für den MU **durchaus zu**, und ich teile die schon von den TIMSS-Leuten vertretene Auffassung, dass der **deutsche MU allzu sehr auf Faktenwissens-Erwerb und die Beherrschung von Verfahren zielt** (T2, 31).

**Definition von ML:** "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 MU von **Heinrich Winter (1975)** mit seinem Begriff der "Umwelt-Erschließung" 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 **Kultur-Gut**, wonach auch einmal ein Wissens-Stoff um seiner selbst willen zu behandeln wäre, z.B. mehrere Beweise des Pythagoras-Satzes.

Die mathematische **Grundbildung** gemäß PISA schlägt sich in **drei Dimensionen** nieder, die entsprechend in den Tests abzuprüfen wären: Mathematische **Inhalte**, mathematische **Kompetenzen (Modellieren oder Problemlösen** in den **Kompetenz-Klassen:** (i) Verfahren ausführen, d.i. *Repro-*

*duction*; (ii) Zusammenhänge herstellen, d.i. *Connection*; (iii) Verallgemeinern, d.i. *Reflection*) und **Anwendung** von Mathematik (P1, 23). Das **Herzstück** sind die **drei Kompetenz-Klassen** (i), (ii), (iii), mit denen dezidiert keinerlei Schwierigkeits-Stufung festgelegt sein soll (s. Knoche u.a. 2002, 161).

Eine andere verwendete Kategorie ist die des **mathematischen Arbeitens**: da gibt es die drei Ausprägungen "**Technische Fertigkeiten**", "**Rechnerisches Modellieren**", "**Begriffliches Modellieren**" (P1, 164), wovon in den Test-Aufgaben erstere fast nie und die beiden anderen etwa je zur Hälfte gefordert werden.

Der dominierende Bestandteil von ML im Sinne von PISA ist die Kompetenz zum **Modellieren** ((ii) und (iii) auf Kosten von (i)), und 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).

### 3.5 Wie weit kann Mathematical Literacy überhaupt getestet werden?

Insgesamt kann ein Test wie **PISA** oder **IGLU**, zumal mit überwiegend Multiple-Choice-Aufgaben, der o.a. **Definition von ML natürlich nicht gerecht** werden. Kein einziger Aspekt kann sich in solchen Aufgaben wieder finden: 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 **Häppchen-Aufgaben**, sei sie noch so komplex aufgebaut, stellt ein **authentisches** Sach-Problem dar, gar ein **Problem der S&S selbst**. Natürlich ist keine Aufgabe **wirklich offen**; es ist lediglich immer wieder der Versuch erkennbar, eine direkte Anwendung von Faktenwissen und Fertigkeiten durch **Einkleidung** des mathematischen Gehalts in allerlei inner- und außer-mathematische Kontexte zu verhindern und so immerhin Modellbildung zu erzwingen. Durchweg gilt aber: ist eine Aufgabe den S&S **schon einmal** so oder in ähnlicher Form **begegnet** und erkennen sie, dass sie Gelerntes einsetzen können, **steigt die Lösungs-Quote**. — Dies alles liegt am **Schonraum-Vermitteltheits-Pädagogik-Charakter von Schule**, worin die Test-Situation ja voll integriert ist; — und dies ist auch gut so.

Meine **Kritik** richtet sich **nicht gegen diese Art von Tests**. Man kann 900.000 S&S (bei PISA, P1, 34ff) oder 150.000 S&S (bei IGLU) weltweit nicht anders testen. Aber **vom Testen der ML ist man viel weiter entfernt**,



als man glaubt oder glauben machen möchte. **Vielleicht** sind **solides Wissen und Können** (natürlich inklusive der Lernstoffe "Modellbildung" und "Problemlösen") **viel bedeutsamer** für die ML, als die PISA-Leute annehmen. Ich habe jedenfalls schon **mehrere Untersuchungen** kennen gelernt, die dieses nahe legen und die mich deswegen so überzeugt haben, weil ihre A&A ersichtlich eigentlich gerne das Gegenteil herausbekommen hatten wollen.

Ein **Beispiel** (das mit seiner Realitäts-Widrigkeit nicht besonders schlecht, sondern typisch ist) aus TIMSS II und (!) TIMSS III (T3.1, 164):

"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 wird erwartet:** Bei A muss man  $12 \cdot 800 = 9600$  *zeds*, bei B  $110 \cdot 90 = 9900$  *zeds* zahlen. Die S&S haben natürlich, schon aus Zeitgründen, **jedes Nachdenken über den Real-Gehalt dieser Situation auszuschalten**, und das wissen sie auch. — Unter mehreren **realitäts-widrigen Situations-Merkmalen** möchte ich das folgende hervorheben: 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 Gesamt-Summe dann von  $475 \cdot 12 + 35 \cdot 90 = 8850$  *zeds*). — Solche Überlegungen wären Teil der **ML**. **Genau diese werden hier nicht erwartet und können nicht erwartet werden.**

**Bei IGLU** spielt das Konzept der **ML noch nicht die Rolle** wie bei PISA (IG, 14), vor allem, weil die IGLU-Mathematik ja letztlich die von TIMSS ist. — Allerdings kann auch hier das Bemühen um Überwindung der simplen Wissens- und Verfahrens-Abfrage zu Verunsicherungen der S&S führen; Beispiel aus TIMSS (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 Güte-Kriterium ab. Für A. spricht, dass die Schätz-Zahlen besonders einfach sind und das Ergebnis gleich auf volle Tausender gerundet ist, während ich bei C. dafür einen zweiten Rundungs-Schritt bräuchte. Wenn man das genaue Ergebnis  $84 \cdot 57 = 4788$  errechnet, sieht man natürlich, dass C. näher dran liegt. Aber selbst ich, der ich wirklich viel Erfahrung mit einem solchen Umgang mit Zahlen habe, war zunächst überrascht, dass der Abstand von A. zum genauen Ergebnis so viel größer ist als der von C. — Ganz bestimmt **nicht** waren bei dieser Aufgabe **solche Überlegungen erwartet** worden; aber genau sie wären der Ausfluss von ML gewesen. Diesem Dilemma entgeht man auch nicht dadurch, dass man beide Antworten als richtig akzeptiert.

### 3.6 Das fragwürdige Konstrukt der Kompetenz-Stufen

Die Kompetenz-Klassen sind aber bei TIMSS, PISA und IGLU noch **nicht das letzte Wort. Nach den Tests** wurden die Punkte-Skalen jeweils in fünf Abschnitte eingeteilt und jeder Abschnitt zu einer Kompetenz-Stufe erklärt. Zugleich liefert diese **Kompetenzstufen-Einteilung für die S&S eine Schwierigkeitsstufen-Einteilung 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 gleich lange Intervalle zerlegt (P1, 187), also mit der Länge  $(696-329)/4 = 91,75$ . 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:

0	$\leq 328$	
I	329–420	Rechnen auf Grundschulniveau,
II	421–511	Elementare Modellierungen,
III	512–603	Modellieren und begriffliches Verknüpfen auf dem Niveau der Sekundarstufe I,
IV	604–695	Umfangreiche Modellierungen auf der Basis anspruchsvoller Begriffe,

V  $\geq 696$  Komplexe Modellierung und innermathematisches Argumentieren.

Offensichtlich ist 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. Dass da Jemand **Rechen-Fehler macht und komplex modellieren** kann, oder **inner-mathematisch argumentieren, aber nur elementar modellieren** kann (o.ä.) und die Einordnung dieser Person etwa in Stufe III diesem **Kompetenz-Profil nicht gerecht** wird, ist hier nicht vorgesehen. Statt dessen wird in den Kurz-Erläuterungen bei Stufe IV und V von **"offenen Aufgaben"** geredet, die ja bekanntlich im Test gar nicht vorkommen (können). — Die **Fragwürdigkeit** dieser Etikettierungs-Bemühungen zeigt sich schön an der **Pyramiden-Aufgabe** (P1, 151ff):

"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  $\sqrt{3}/4 \cdot a^2 = 62 \text{ cm}^2$ . So gesehen, ist das eine **reine Wissens-Aufgabe**, 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 S&S 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 leicht; 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 das Auswendig-Lernen von Formeln **heutzutage verpönt**, und man muss sie sich nur noch **herleiten**

können. Wenn man **in der Schule schon einmal** die Flächeninhalts-Formel für das gleichseitige Dreieck mit Hilfe des Pythagoras-Satzes hergeleitet hat, fällt einem das im PISA-Test natürlich viel leichter, als wenn man noch nichts davon gehört hat.

PISA kommt schließlich zu der Feststellung, dass man **erst ab der Kompetenz-Stufe III** von einem **ausreichenden Standard für ML** sprechen sollte (P1, 161), und zu Bezeichnungen wie Risiko- und Spitzen-Gruppe. Da verfügt **Deutschland über eine Risiko-Gruppe von 25%** (gegenüber z.B. USA 23%, Schweden 17%, Japan 7%), eine **Spitzen-Gruppe von nur 1,3%** (gegenüber z.B. Großbritannien 2,9%, Schweiz 3,3%, Japan 3,9%), und **nur 44% befinden sich mindestens auf Stufe III**, erfüllen also den Grundbildungs-**Standard** (gegenüber z.B. USA 44%, Frankreich 53%, Japan 72%) (P1, 170ff). Die in den Medien gepushte **Schlagzeile**, dass deutlich weniger als die Hälfte unserer 15-Jährigen den PISA-Grundbildungs-**Standard** erfüllt, **beruht auf dieser willkürlichen Klassen-Einteilung**.

### 3.7 Die Kompetenz-Stufen bei IGLU

Bei IGLU ist man mit den **Kompetenz-Stufen pragmatischer** umgegangen und hat sie vermutlich an stoffliche Kriterien angepasst (IG, 202):

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

Natürlich ist **jegliches Bearbeiten** mathematischer Aufgaben ein **Problemlösen**; Problemlösen ist also kein Merkmal einer besonders hohen Kompetenz-Stufe. Auch inner- und außer-mathematische Kontexte kommen auf jeder Kompetenz-Stufe vor, z.B. auch: "Jan hat 4 Bonbons, Maja hat 3 Bonbons, usw." — Insofern liefert die Charakterisierung von Stufe V z.B. noch **keinen Hinweis auf hohe Kompetenz**.

Sowohl bei den veröffentlichten Muster-Aufgaben, als auch bei sämtlichen mir ja bekannten unveröffentlichten IGLU-Aufgaben schätze ich die **Rolle**

**des Modellbildens als entscheidende Kompetenz geringer ein als die IGLU- oder gar die PISA-Leute.**

Auch die Kompetenzstufen-Bildung im Bereich der **Naturwissenschaften** in der Grundschule (IG, 156ff) kommt mir nicht gut gelungen vor:

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

Man muss **überrascht** zur Kenntnis nehmen, dass die Frage "Welches Tier säugt seine Jungen?" mit den Antwort-Möglichkeiten "Huhn, Frosch, Affe, Schlange" immerhin den Schwierigkeits-Grad 474 hat. Sie gerät deshalb in **Stufe III**, d.i. Anwenden **nicht mehr alltags-, sondern nun naturwissenschaftsnaher** Begriffe. Ich meine, dass diese Kategorisierung dieser Frage nicht gerecht wird. Da geht es vielmehr um Allgemeinbildung auf niedrigem Niveau, um **elementares Wissen**, das entweder vorhanden ist oder nicht.

**Stromkreis-Aufgabe** (546 Punkte): "Die folgenden Bilder zeigen eine Taschenlampenbatterie und eine Glühbirne, die durch Drähte mit verschiedenen Substanzen verbunden sind." Tatsächlich sind vier Stromkreise (1, 2, 3, 4) mit einer Batterie als Quelle und einer Birne (entsprechend 1, 2, 3, 4) als Verbraucher gezeichnet, wobei in jedem Stromkreis der Draht unterbrochen und die Lücke durch einen "Gegenstand", bestehend aus einem bestimmten Stoff, überbrückt ist: 1 Luft, 2 Nagel, 3 Münze, 4 Radiergummi. "Welche der Glühbirnen werden leuchten? A. Glühbirnen 1 und 2, B. Glühbirnen 1, 2 und 3, C. Glühbirnen 2, 3 und 4, D. Glühbirnen 2 und 3, E. Glühbirnen 3 und 4."

Abgesehen von der Überlagerung des naturwissenschaftlichen Problems durch ein kombinatorisches, können S&S die Aufgabe bei noch so viel naturwissenschaftlichem Verständnis **nicht lösen, wenn mit ihnen nicht vorher massiv der Stoff und die Art der grafischen Darstellung behandelt wurde**.

**Samenwachstums-Aufgabe** (673 Punkte): "Um herauszufinden, ob Samen im Licht oder im Dunkeln besser wachsen, kannst du einige Samen auf ein feuchtes Papier legen und A. sie an einen warmen dunklen Ort stellen; B. einen Teil an einen hellen Ort und einen Teil an einen dunklen Ort stellen; C. sie an einen warmen, hellen Ort stellen; D. sie an einen hellen oder dunklen,

aber kühlen Ort stellen." — Mit dieser Aufgabe wird zwar in der Tat naturwissenschaftliches Verständnis abgefragt. Ein Teil der hohen Schwierigkeit ist nach meinem Dafürhalten aber durch die **verführerischen sprachlichen Wendungen** zu erklären, die zu Fehlern verleiten können. Z.B. ist ja **auch Strategie D. zielführend**. Die dabei auftretende Angabe zur Temperatur ist eine Zutat wie die insgesamt vorgegebene Feuchtigkeit, von der man ja auch nicht weiß, ob, und wenn ja, wieso sie wichtig ist.

Schon im mathematischen, und noch mehr im naturwissenschaftlichen Test ist **Lese-Kompetenz eine ganz wichtige Voraussetzung** (P1, 183). Unabhängig davon besteht eine sehr **hohe Korrelation zwischen Lesen und Mathematik, Lesen und Naturwissenschaften** sowie zwischen Mathematik und Naturwissenschaften. Kinder, die in einer dieser Domänen sehr hohe und in einer anderen sehr niedrige Kompetenz haben, sind eher die Ausnahme. Nach allen Erfahrungen sieht dies bei Primarstufenlehramts-Studierenden nicht anders aus.

#### 4. Schluss-Bemerkung

Bei aller Kritik muss man die **großartige Leistung** der internationalen Vergleichs-Untersuchungen TIMSS, PISA und IGLU würdigen. **Gestört** hat mich lediglich das Aufschwingen zu einer **obersten bildungs-wissenschaftlichen Instanz**, das bestimmt auch zu dem **Missbrauch durch Politik und Medien** beigetragen hat.

**Missfallen** hat mir außerdem die zunehmende **Geheimnis-Krämerei** um die Ergebnisse zu den konkreten Aufgaben. Seit den PISA-Tests sind jetzt drei und seit den IGLU-Tests zwei Jahre vergangen, und man hat noch nichts erfahren. Immerhin haben wir uns ja über die internationalen Aufgaben hinaus eine Menge "nationaler" Aufgaben ausgedacht als Grundlage für weitere Forschungen, die dann schon veraltet sein werden, ehe sie überhaupt beginnen können. Diese konkreten Forschungen wären mir viel wichtiger, als dass jetzt, 2003, die nächste PISA-Runde stattfindet, ein weiterer riesiger Daten-Berg aufgehäuft wird, zu dem lediglich wieder die Statistik veröffentlicht wird.

Trotzdem: Es war **überfällig, dass auch Deutschland sich an diesen internationalen Vergleichs-Untersuchungen beteiligt**. Auch wenn man meint, dass die Durchschnitts-Werte und Platz-Ziffern 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 S&S **international Mittelmaß** (für die SII s. T3.1, 140) und lediglich in der

Grundschule etwas besser sind. — Ganz **wichtig ist die Aufmerksamkeit, die Bildung, speziell auch mathematische Bildung, in Politik und Öffentlichkeit** gewonnen hat und die erfreulicher Weise bis heute anhält.

Die **mathematik-didaktische** Arbeit für die **Grundschule hat eine gewisse Bestätigung** erfahren, die auch die Richtung für die Sekundarstufe I vorgibt, wie z.B. im Projekt SINUS (Blum 2003) realisiert. Vor allem aber hat uns PISA die **Versäumnisse beim schwachen Viertel unserer Heranwachsenden** vor Augen geführt. Diese Versäumnisse sind allerdings nicht nur schulischer, sondern gesamt-gesellschaftlicher Natur, und nicht ausschließlich, aber zu einem großen Teil bestehen sie aus mangelnder Integration der Familien mit MH.

#### **TIMSS-, PISA-, IGLU-Veröffentlichungen**

(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

(IG, 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

### **Sonstige Literatur**

Bender, Peter (1980): Analyse der Ergebnisse eines Sachrechentests am Ende des 4. Schuljahres. In: Sachunterricht und Mathematik in der Primarstufe 8, 150-155, 191-198, 226-233

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

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

Klafki, Wolfgang (1958): Didaktische Analyse als Kern der Unterrichtsvorbereitung. In: Die Deutsche Schule 50, 450–471

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

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

Schmidt, Christoph M. & Michael Fertig (2003): Genaues Hinsehen lohnt. Die Determinanten des Abschneidens deutscher Schüler in der PISA 2000-Studie. In: Forschung & Lehre 2003, 313–315

Weinert, Franz Emanuel (1996): Thesenpapier zum Vortrag "Ansprüche an das Lernen in heutiger Zeit". München: Manuskript

Winter, Heinrich (1975): Allgemeine Lernziele für den Mathematikunterricht? In: Zentralblatt für Didaktik der Mathematik 7, 106–116