Aus: Stefan Thomas Hopmann, Gertrude Brinek, Martin Retzl (Hg./ Eds.): PISA zufolge PISA - PISA According to PISA.
LIT Verlag, Wien 2007

Deutsche Pisa-Folgen

Thomas Jahnke

Deutschland: Universität Potsdam

In dieser Note werden die Beschlüsse der Kultusministerkonferenz zum ‚Bildungsmonitoring'und zu den ‚Bildungsstandards' in Mathematik als nationale Pisa-Folgen identifiziert. Eine Auseinandersetzung mit der Testforschung in den USA und eine Ernüchterung der veröffentlichten Meinung zu der Testwirklichkeit kann die Geltungsmacht von Pisa & Co in Deutschland möglicherweise eindämmen.

Die Teilnahme an der Dritten Mathematik- und Naturwissenschaftsstudie (TIMSS) und dem ersten Durchgang des Programme for International Student Assessment (PISA) hat zu einer grundlegenden Wende in der deutschen Bildungspolitik geführt. Das Unbehagen an der deutschen Schule ist messbar geworden und mit diesen Messungen ist auch der Weg, die Verhältnisse zu bessern, vorgezeichnet: die Messwerte müssen höher werden, dann wird es besser. Die Wucht, mit der dieser Gedanke die mediale und politische Öffentlichkeit durchrollte und vereinzelte Kritik an solchen Erkenntnisse und der einzuschlagenden Kur unter sich begrub, hatte lawinenartigen Charakter. Die Messergebnisse scheinen wirklicher als jede Theorie, und den „deskriptiven Befunde" haftet eine quasi-naturwissenschaftliche Objektivität und damit unwiderlegbare Wahrheit an: so liegen die Dinge – im Rahmen der Messgenauigkeit. Der Triumpf empirischen Denkens: die Wirklichkeit ist beziffert, digitalisiert, das Menetekel hat Dezimale bekommen und kann nun Steuerungsprozessen unterworfen werden, deren Ergebnisse wieder zu messen sind und so weiter.

In Deutschland wird kaum diskutiert, dass auch solchen Messungen eine – möglicherweise holprige, unausgesprochene, wenig durchdachte – Theorie zugrunde liegt und Begriffe wie ‚Kompetenzstufen' oder ‚Grundbildung' sich nicht messtechnisch ergeben, sondern „Realität" eher hervorbringen als

beschreiben. Ferner ist der Glaube, durch periodisierte Testungen würden die Leistungen deutscher Schülerinnen und Schüler steigen, weit und auch in der Bildungsadministration verbreitet. Kritik an Pisa wird häufig damit zurückgewiesen, dass ein Unternehmen dieser Größenordnung natürlich auch Schwächen und Ungereimtheiten aufweise, dass der ‚Pisa-Schock'aber grundsätzlich doch das Augenmerk auf die Schulwirklichkeit gelenkt und schon damit Bewegung gebracht und diverse Reformbestrebungen in Gang gesetzt habe. Verkannt wird dabei, dass es sich bei Pisa nicht um eine einmalige Testung handelt, deren Ergebnisse in ihrer Aussagekraft möglicherweise überschätzt einem reflektierenden Betrachter schon etwas erzählen könnten, sondern um ein Programm, das keineswegs den Blick für verschiedenste Reformansätze und –anstrengungen öffnet, sondern im Gegenteil den Weg durch das Ziel schon festgeschrieben hat: deutsche Schülerinnen und Schüler sollen bei den künftigen Tests besser abschneiden.

„Bildungsmonitoring"

Die *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring* liegt in doppelter Form vor: einmal als Beschluss der Kultusministerkonferenz vom 02.06.2006 [1], zum anderen als Broschüre [2], die vom Sekretariat der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland in Zusammenarbeit mit dem Institut zur Qualitätsentwicklung im Bildungswesen (IQB) 2006 herausgegeben wurde. Die Broschüre ist – schon auf dem Umschlag – mit ganzseitigen Farbfotos von Schülerinnen und Schülern illustriert, enthält ein Vorwort der Präsidentin der Kultusministerkonferenz und ein Inhaltsverzeichnis mit geänderter Nummerierungen der Abschnitte, ist um einen Abschnitt mit Aufgabenbeispielen angereichert. Offensichtlich hat man dem IQB zugestanden, sein Aufgaben- und Pflichtenbuch selbst zu überarbeiten und die Formulierungen des zugrunde liegenden Beschlusses sich passend zu glätten und auszulegen. Dies geschieht tatsächlich Absatz für Absatz. Aus der Formulierung *... in eine Reihe von Beschlüssen der KMK einzuordnen, die entsprechende Handlungsfelder beschreiben und gemeinsame zentrale Arbeitsbereiche nach Pisa 2003 festlegen* in dem Beschluss der KMK wird der Bezug auf Pisa 2003 gestrichen. Aus dem *Arbeitsbereich Bereitstellung von*

[1] Kultusministerkonferenz (KMK): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (Beschlüsse der Kultusministerkonferenz vom 02.06.2006)

[2] Kultusministerkonferenz (KMK): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (2006)

Fortbildungskonzeptionen und –materialien zur kompetenz- bzw. standardbasierten Unterrichtsentwicklung, vor allem Lesen, Geometrie, Stochastik wird der Bezug zum Lesen und der – nicht nachvollziehbare, verwunderliche – Bezug auf spezielle mathematische Bereiche, den man wohl nur durch die Abwesenheit und damit auch Verzichtbarkeit von ‚Fachkompetenz' erklären kann, gestrichen. Wir zitieren im Folgenden die etwas schlankeren Formulierungen des ursprünglichen Beschlusses. Schon der erste Absatz lässt wenig Zweifel unter welchen Auspizien Bildung heute betrachtet wird:

> Bildung nimmt eine Schlüsselrolle für die individuelle Entwicklung, für gesellschaftliche Teilhabe sowie berufliches Fortkommen, aber auch für den wirtschaftlichen Erfolg eines Landes ein. Die globalen Entwicklungen der vergangenen Jahrzehnte haben die grundlegende Bedeutung von Bildung für Deutschland noch einmal unterstrichen. Die Ausschöpfung aller Begabungspotentiale und die Sicherung und Entwicklung von Qualität im Bildungswesen sind daher zentrale Aufgaben der Bildungspolitik. (S. 1) [3]

Als *zentrale Instrumente der Kultusministerkonferenz für das Bildungsmonitoring* werden dann benannt

– Internationale Schulleistungsuntersuchungen
– Zentrale Überprüfung des Erreichens der Bildungsstandards in einem Ländervergleich
– Vergleichsarbeiten in Anbindung oder Ankoppelung an die Bildungsstandards zur landesweiten Überprüfung der Leistungsfähigkeit einzelner Schulen
– Gemeinsame Bildungsberichterstattung von Bund und Ländern. (S. 1/2)

Ob die bisher veröffentlichten ‚Bildungsstandards'(s.u.!) solchen Überprüfungen und Belastungen standhalten, kann man bezweifeln. In jedem Fall wird Deutschland durch diesen Beschluss zum Testland ausgerufen und erklärt: für die Jahre 2006 bis 2018 (!) werden in einer Tabelle 17 Testungen und 19 Berichterstattungen über diese terminiert, die sich allein durch die Teilnahme an PIRLS, TIMSS und PISA [4] sowie die Ländervergleiche bundesweit er-

[3] In ihrem Tenor und Jargon erinnert solche Funktionsbeschreibung von ‚Bildung'an die entsprechende Verlautbarungen der OECD. Überraschender Weise ist in der Überarbeitung des Textes (Siehe die o.a. Broschüre) in dem angeführten Zitat das Wort ‚Bildung'durch ‚Das Bildungssystem'ersetzt, als seien diese Begriffe synonym.

[4] PIRLS ist die Abkürzung für Progress in International Reading Literacy Study, die in Deutschland auch mit IGLU für Internationale Grundschul-Lese-Untersuchung bezeichnet wird. TIMSS war ursprünglich ein Akronym für Third International Mathematics and Science Study; seit TIMSS 2003 steht das Akronym für Trends in Mathematics and Science Study; PISA steht für Progamme for International Students Assessment.

geben. Dazu kommen noch die *länderspezifische und länderübergreifenden Vergleichsarbeiten in Anbindung oder Anlehnung an die Bildungsstandards* in Jahrgangsstufen 3 und 4 für Deutsch und Mathematik, in den Jahrgangsstufen 8 und 9 für den Hauptschulabschluss in Deutsch, Mathematik, Erste Fremdsprache (Englisch, Französisch) und in den Jahrgangsstufen 9 und 10 für den Mittleren Schulabschluss in Deutsch, Mathematik, Erste Fremdsprache (Englisch, Französisch), Biologie, Chemie, Physik.

Dass schulische Bildung in Deutschland solchem ‚Monitoring' nicht mehr entkommen kann, wird schließlich im letzten Abschnitt *Bildungsberichterstattung* gesichert:

> Kern der Bildungsberichterstattung ist ein überschaubarer, systematischer, regelmäßig aktualisierter Satz von Indikatoren, d.h. statistischen Kennziffern, die jeweils für ein zentrales Merkmal von Bildungsprozessen bzw. einen zentralen Aspekt von Bildungsqualität stehen. Diese Indikatoren werden aus amtlichen Daten und sozialwissenschaftlichen Erhebungen in Zeitreihe dargestellt, wenn möglich im internationalen Vergleich und aufgeschlüsselt nach Ländern. Um den Vergleich mit Entwicklungen in den Mitgliedstaaten der Europäischen Union und der OECD zu ermöglichen, wird Anschlussfähigkeit und Kompatibilität mit internationalen Berichtssystemen (. . .) angestrebt. (. . .)
>
> Durch die Verfügbarkeit individueller Verlaufsdaten und die regelmäßige Erfassung erworbener Kompetenzen soll die Leitidee der Bildungsberichtserstattung „Bildung im Lebenslauf" umgesetzt werden. Für einen einheitlichen Satz schulstatistischer Daten und die Sicherung der Anschlussfähigkeit an die internationale Bildungsstatistik haben die Länder bereits grundlegende Beschlüsse gefasst. So haben die Länder am 22.09.2005 vereinbart, längerfristig ihre Daten entsprechend den im Kerndatensatz vereinbarten Merkmalsausprägungen zur Verfügung zu stellen. Zumindest Daten der öffentlichen Schulen sollen für das Schuljahr 2008/2009 von allen Ländern vorliegen. (S. 14)

In der überarbeiten Broschüre zum Bildungsmonitoring wurde der zuletzt zitierte Absatz gestrichen. Es ist aber wohl kaum davon auszugehen, dass damit auch die angestrebte Datenbank nicht eingerichtet wird.

„Teaching to the Test"

Man muss konstatieren, dass die Kritik an den Testverfahren und an dem Gedanken, die ‚Erträge'schulischer Bildung könnten über periodisierte Tests in sinnvoller Weise gemessen und gesteigert werden, Deutschland nicht erreicht hat oder dass es auch nur zu einer sorgfältigen und redlichen Diskussion dieses

prima vista selbstverständlich erscheinenden Gedankens hierzulande gekommen ist. [5] Das ist auch nicht weiter erstaunlich, weil von den involvierten Testinstitute und den mit ihnen kooperierenden Wissenschaftler kaum zu erwarten ist, dass sie mit ihren Test- Knowhow auch die Test-Kritik auf den Markt bringen.

In dem vierzehnseitigen Beschluss der Kultusministerkonferenz zum Bildungsmonitoring vom 02.02.2006 wird auf Seite 13 unter der Zwischenüberschrift „Weiterentwicklung der Bildung, aber kein Teaching to the Test“ auf diese Problematik – wie folgt – kurz eingegangen:

> Neben der Funktion der Beschreibung von Leistungsanforderungen und der Leistungsmessung dienen die Bildungsstandards primär der Weiterentwicklung des Unterrichts und vor allem der individuellen Förderung aller Schülerinnen und Schüler. Die Länder sind sich darin einig, dass mit der Setzung der Bildungsstandards als übergreifenden Referenzrahmen eine Entwicklung hin zum „teaching to the test“ oder eine Verengung des Unterrichts aus die Anforderungen der Standards verhindert werden muss. (S. 13)

Diese Kürze ist trotz der beschworenen Einigkeit der Länder erstaunlich. Es liegt nahe, wenn man die ‚Weiterentwicklung des Unterrichts und vor allem der individuellen Förderungen der Schülerinnen und Schüler‘ durch Bildungsstandards befördern oder anordnen will, deren Erreichen im Wesentlichen durch Tests überprüft wird, die Erfahrungen von Ländern und darunter insbesondere der USA zu rezipieren, die seit Jahren oder Jahrzehnten eine solche Politik verfolgen.

> For several decades, some measurement experts have warned that high-stakes testing could lead to inappropriate forms of test preparation and score inflation, which we

[5] Es ist aufschlussreich und vermutlich nicht folgenlos, dass die Daten von Pisa in Australien aufgearbeitet werden und gleichsam deutschen (oder europäischen) Boden nie betreten. In Zeiten einer sich global verstehenden Forschung scheinen solche räumliche Distanzen ohne jede Auswirkung, u.a. weil der Zugriff auf Datenserver ubiquitär und ohne zeitliche Verzögerung möglich ist. Dennoch ist es von Bedeutung, ob und in welchem Rahmen und geistigen Raum die Verfahren zur Aufbereitung der Daten entwickelt, diskutiert und kritisiert werden, ob sie als die Ergebnisse der Form und dem Inhalt nach prägende Instrumente begriffen werden oder nur – als mehr oder minder schlecht dokumentierte – Routinen in Softwarepaketen erscheinen, ob sie überhaupt wissenschaftlich diskutiert oder schlicht als notwendige und doch arbiträre Essenzen eines ‚State of the Art‘ aufgefasst werden, ob den beteiligten Forschern daran liegt, ihre Verfahren zu verkaufen oder als Erkenntnisinstrumente in die Diskussion einzuführen und zu legitimieren etc.

define as a gain in scores that substantially overstates the improvement in learning it implies. (p. 99)

So leitet Daniel Koretz, Erziehungswissenschaftler an der Havard-Universität und assoziierter Direktor des Center of Research, Standards, and Student Testing (CRESST), seinen Aufsatz *Alignment, High Stakes, and the Inflation of Test Scores*[6] ein und beschreibt einen Ausgangspunkt, über den eine öffentliche Diskussion in Deutschlang bisher kaum hinausgekommen ist:

On common response to this problem has been to seek "tests worth teaching to". The search for such tests has led reformers in several directions over the years, but currently, many argue that tests well aligned with standards meet this criterion. If tests are aligned with standards, the arguments runs, they test material deemed important, and teaching to the test therefore teaches what is important. If students are being taught what is important, how can the resulting score gains be misleading? (p. 99)

Koretz begründet seinen Widerspruch gegen solche Naivität theoretisch und empirisch unter anderem eindrücklich mit Sägezahnkurven ("sawtooth pattern") für die gemessenen Leistungen der gleichen oder einer vergleichbaren Population, die sich in verschiedenen Erhebungen je nach den verwendeten Tests in unterschiedlichster Weise ergaben. Auch der Hoffnung, solche Effekte seien allein der Testkonstruktion und den Testumständen zuzuschreiben, widerspricht er:

The problem is not confined to commercial, off-the-shelf, multiple-choice tests. It has appeared as well with standards-based tests and with tests using no multiple-choice items. (p. 106)

Die Vorstellung, Schülerleistungen ließen sich in einem Test objektiv oder mit angebbaren Fehlermargen – gleichsam physikalisch messen, ist schlicht (und) irreführend. Folgerungen aus solcher Vorstellung mehr als fragwürdig. Wird dies in Abrede gestellt, verschwiegen oder das Gegenteil prätendiert, liegen in aller Regel massive Erkenntnisinteressen der Auftraggeber oder -nehmer der Testungen vor.

---

[6] Koretz, D.: Alignment, High Stakes, and the Inflation of Test Scores. Yearbook of the National Society for the Study of Education (2005) 104 (2), 99–118.
(Online erhältlich unter: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1744-7984.2005.00027.x)

Auch die Auswirkungen von Testungen auf den Unterricht werden in den USA seit Jahrzehnten untersucht. Koretz zum Beispiel beschreibt und charakterisiert in dem zitierten Papier *Reallocation, Alignment* und *Coaching:*

Reallocation. Reallocation refers to shifts in instructional resources among the elements of performance. Research has shown that when scores on a test are important to teachers, many of them will reallocate their instructional time to focus more on the material emphasized by the test. (. . .) Many observers believe that reallocation is among the most important factors causing the sawtooth pattern (. . .).

Alignment. Content and performance standards comprise material – performance elements, in the terminology used here – that someone (not necessarily the ultimate user of scores) has decided are important. If the material is emphasized in the standards, that implies that users should give this material substantial weight in the interference they draw about student performance. Alignment gives this same material high weights in the test as well. (. . .)

Coaching. The term "coaching" is used in a variety of different ways in writings about test preparation. Here it is used to refer to two specific, related types of test preparation, called substantive and non-substantive coaching.

Substantive coaching is an emphasis on narrow, substantive aspects of a test that capitalizes on the particular style or emphasis of test items. The aspects of the tests that are emphasized may be either intended or unintended by the test designers. For example, in one study of the author's, a teacher noted that the state's test always used regular polygons in test items and suggested that teachers should focus solely on those and ignore irregular polygons. The intended interferences, however, were about polygons, not specifically regular polygons. (. . .)

Nonsubstantive coaching refers to the same process when focused on nonsubstantive aspects of a test, such as characteristics of distracters (incorrect answers to multiple-choice items), substantively unimportant aspects of scoring rubrics, and so on. Teaching test-taking tricks (process of elimination, plug-in, etc.) can also be seen as nonsubstantive coaching. In some cases – for example, when first introducing young children to the op-scan answer sheets used with multiple-choice tests – a modest amount of certain types of nonsubstantive coaching can increase scores and improve validity by removing irrelevant barriers to performance. In most cases, however, it either wastes time or inflates scores. (p. 110-112)

An anderer Stelle findet sich ähnliche Kritik. So fasst Brian M. Stecher sein Kapitel 4 *Consequences of large-scale, high-stakes testing on school and classroom practices* in dem von ihm mitherausgegebenen Buch *Making Sense of Test-Based Accountability in Education* [7] folgendermaßen zusammen:

[7] Stecher, B. M.: Consequences of large-scale, high-stakes testing on school and classroom

The net effect of high-stakes testing on policy and practice is uncertain. Researchers have not documented the desirable consequences of testing – providing more instruction, working harder, and working more effectively – as clearly as the undesirable ones – such as negative reallocation, negative alignment of classroom time to emphasize topics covered by a test, excessive coaching, and cheating. More important, researchers have not generally measured the extent or magnitude of the shifts in practice that the identified as a result of high-stakes testing.

Overall, the evidence suggests that large-scale high-stakes testing has been a relatively potent policy in terms of bringing about changes within schools and classrooms. Many of these changes appear to diminish students' exposure to curriculum, which undermines the meaning of the test scores. (p. 99/100)

Der im letzten Absatz angesprochene Antagonismus scheint der deutschen Kultusministerkonferenz möglicherweise von ihren Beratern vorenthalten worden zu sein. Das Gleiche gilt vermutlich für das *Position Statement on High Stakes Testing in PreK-12 Education* der American Evaluation Association (AEA), in dem es heißt:

High stakes testing leads to under-serving or mis-serving all students, especially the most needy and vulnerable, thereby violating the principle of "do no harm." The American Evaluation Association opposes the use of tests as the sole or primary criterion for making decisions with serious negative consequences for students, educators, and schools. The AEA supports systems of assessment and accountability that help education.

Recent years have seen an increased reliance on high stakes testing (the use of tests to make critical decisions about students, teachers, and schools) without full validation throughout the United States. The rationale for increased uses of testing is often based on a need for solid information to help policy makers shape policies and practices to insure the academic success of all students. Our reading of the accumulated evidence over the past two decades indicates that high stakes testing does not lead to better educational policies and practices. There is evidence that such testing often leads to educationally unjust consequences and unsound practices, even though it occasionally upgrades teaching and learning conditions in some classrooms and schools. The consequences that concern us most are increased drop out rates, teacher and administrator deprofessionalization, loss of curricular integrity, increased cultural insensitivity, and disproportionate allocation of educational resources into testing programs and not into hiring qualified teachers and providing sound educational programs. The deleterious

practices. In L. S. Hamilton, B. M. Stecher, and S. P. Klein (Eds.): Making Sense of Test-Based Accountability in Education. RAND. Santa Monica 2002. P. 79-100
(Online unter: http://www.rand.org/pubs/monograph_reports/MR1554/index.html)

effects of high stakes testing need further study, but the evidence of injury is compelling enough that AEA does not support continuation of the practice.

While the shortcomings of contemporary schooling are serious, the simplistic application of single tests or test batteries to make high stakes decisions about individuals and groups impede rather than improve student learning. Comparisons of schools and students based on test scores promote teaching to the test, especially in ways that do not constitute an improvement in teaching and learning. Although used for more than two decades, state mandated high stakes testing has not improved the quality of schools; nor diminished disparities in academic achievement along gender, race or class lines; nor moved the country forward in moral, social, or economic terms. The American Evaluation Association (AEA) is a staunch supporter of accountability, but not test driven accountability. AEA joins many other professional associations in opposing the inappropriate use of tests to make high stakes decisions.

In einer Endnote zu diesem Text wird auf weitere Organisationen verwiesen, die ebenfalls dagegen opponieren, weit reichende Entscheidungen auf Grund von Testergebnissen zu fällen.

AEA joins many other professional associations, teacher unions, parent advocacy groups in opposing the inappropriate use of tests to make high stakes decisions. These include, but are not limited to the American Educational Research Association, the National Council for Teachers of English, the National Council for Teachers of Mathematics, the International Reading Association, the College and University Faculty Assembly of the National Council for the Social Studies, and the National Education Association [8]

Für den deutschen Betrachter ist es kaum nachvollziehbar, mit welchen Besserungs- wenn nicht gar Heilserwartungen gleich in welcher Richtung die hiesige Bildungspolitik umfangreichste Testprograme einführt, während der sicherlich nicht zimperliche angelsächsische Evaluations-Pragmatismus sich in kaum zu übertreffender Deutlichkeit nach mehr als zwanzigjähriger Erfahrung von solchen Bestrebungen distanziert.

„Bildungsstandards“

Während in dem Beschluss der Kultusministerkonferenz zum Bildungsmonitoring noch wie oben zitiert davon die Rede ist, dass die *Bildungsstandards primär der Weiterentwicklung des Unterrichts und vor allem der individuellen*

[8] American Evaluation Association AEA): Position Statement on HIGH STAKES TESTING In PreK-12 Education. 2002 (Online unter: http://www.eval.org/hst3.htm)

Förderung aller Schülerinnen und Schüler dienen heißt es in der *Vereinbarung über Bildungsstandards für den Mittleren Schulabschluss* [9] der gleichen Organisation deutlicher und weniger pädagogisch verschleiert:

Die Kultusministerkonferenz sieht es als zentrale Aufgabe an, die Qualität schulischer Bildung, die Vergleichbarkeit schulischer Abschlüsse sowie die Durchlässigkeit des Bildungssystems zu sichern. Bildungsstandards sind hierbei von besonderer Bedeutung. Sie sind Bestandteil eines umfassenden Systems der Qualitätssicherung, das auch Schulentwicklung, externe und interne Evaluation umfasst. Bildungsstandards beschreiben erwartete Lernergebnisse. Ihre Anwendung bietet Hinweise für notwendige Förderungs- und Unterstützungsmaßnahmen. (S. 3)

Standards [10] und Tests bescheinigen sich gegenseitig ihre Notwendigkeit in einem Maße, dass sie gleichsam aus purer Logik existent werden: Tests benötigen Standards, an denen sie oder auf die sie ausgerichtet sind; Standards benötigen Tests zur Überprüfung ihrer Einhaltung oder Erreichung oder ihres Verfehlens. Eine kurze – weniger logische – Geschichte der Standards in Deutschland hat Hans Dieter Sill 2006 skizziert. [11] Er kommt zu dem Schluss:

Die Standards sind nicht im Resultat gründlicher wissenschaftlicher Analysen internationaler und nationaler Entwicklungen entstanden, sondern sind Ergebnis eines politisch motivierten Beschlusses auf ministerialer Ebene, der in sehr kurzer Zeit umzusetzen war. Es bestanden weder zeitliche noch personelle Ressourcen, um den wissenschaftlich außerordentlich anspruchsvollen Prozess der Entwicklung nationaler Standards in der notwendigen Tiefe und Gründlichkeit zu gestalten. (S. 299/200)

Die Ergebnisse solcher Knappheit kennzeichnen zum mindestens rechnerisch nicht die *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss*, die am 4.12.2003 von der Kultusministerkonferenz beschlossen wurden. Durch die Setzung von sechs sich ohne jede Trennschärfe oder auch nur

---

[9] Vereinbarung über Bildungsstandards für den Mittleren Schulabschluss (Jahrgangsstufe 10) – (Beschluss der Kultusministerkonferenz vom 4.12.2003) in: Kultusministerkonferenz (KMK): Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 4.12.2003

[10] In Achtung der großen deutschen Bildungstheoretiker des 18., 19. und 20. Jahrhunderts versuche ich das Wort Bildungsstandard zu vermeiden. Als wäre Ohr und Verstand mit dem Kompositum ‚Bildungsstandards' noch nicht ausreichend gequält, wird in dem Beschluss der Kultusministerkonferenz zum Bildungsmonitoring an zahlreichen Stellen noch von der notwendigen *Normierung* und *Nachnormierung* der Bildungsstandards gesprochen.

[11] Sill, H. D.: PISA und die Bildungsstandards. In: Jahnke, Th.; Meyerhöher, W. (Hrsg.): Pisa & Co – Kritik eines Programms. Franzbecker Verlag. Hildesheim 2006. S. 293 – 330.

eigene Charakteristik überlappenden *Kompetenzen*, von fünf – seit Jahren bekannten – *mathematischen Leitideen* und drei *Anforderungsbereichen* ergeben sich aus Gründen der Multiplikation neunzig verschiedene Möglichkeiten eine Aufgabe zu kennzeichnen. Sind – wie wohl meistens zu erwarten – mehrere Kompetenzen oder Leitideen gefragt, dann ergeben sich mehrere hundert solcher Klassifikationen.

Auf 24 der 36 Seiten der Broschüre sind Aufgabenbeispiele und *Lösungsskizzen mit der Angabe von Leitideen und allgemeinen mathematischen Kompetenzen sowie deren Zuordnung zu Anforderungsbereichen* abgedruckt. Während die Klassifikation der Aufgaben wenig zwingend oder aufschlussreich, eher selbstverständlich und für die Bearbeitung von zu vernachlässigender Bedeutung ist, erschreckt die magere Qualität der Aufgaben, die Material, wie es in neueren, gut durchgearbeiteten und aufbereiteten Schulbüchern zu finden ist, nicht einmal im Ansatz erreicht.

In Aufgabe (1) wird unangemessen modelliert.

In Aufgabe (2) wird das Ungeschick eines Grafikers, das diesem vermutlich durch einen fehlerhaften Umgang mit einer Tabellenkalkulationssoftware unterlaufen ist, nicht thematisiert, sondern hingenommen.

In Aufgabe (3) wird ein nicht symmetrisch gezeichneter Stern als symmetrisch bezeichnet und dann nach der Zahl seiner Symmetrieachsen gefragt.

In Aufgabe (4) erstaunen die künstliche Fragestellung und die Klassifikation.

In Aufgabe (5) ist eine mit „Lohnerhöhung in EURO“ beschriftete Achse in Zehnerschritten von 0 bis 50 bezeichnet, aber zugleich in 30 Teile geteilt, so dass ein Teilabschnitt 1 2/3 € entspricht und die Achsenbeschriftungen nicht an den Teilstrichen sitzen (können).

In Aufgabe (6) wird ein Punkt mit P(y;x) bezeichnet und dann bemerkt, dass „x die erste Koordinate des Punktes P ist“.

In Aufgabe (7) erstaunen die Teilfragen c) und d).

In Aufgabe (8) wird vor allem die Anstrengung deutlich, eine Leitidee unterzubringen.

In Aufgabe (9) ist die Fragestellung c) undurchsichtig.

In Aufgabe (10) wird der Taschenrechner fragwürdig benutzt.

In Aufgabe (11) soll man sich mit den fünf Schüleräußerungen in Sprechblasen auseinandersetzen, die man außerhalb der Schule wohl kaum mathematisch aufarbeiten würde.

In Aufgabe (12) werden Fragestellungen zu Linearen Funktionen behandelt, denen man wenig Sinn abgewinnen kann.

In Aufgabe (13) wäre im Ansatz einmal eine Modellierung möglich, wenn sie nicht im Text schon vorgegeben wäre.

In Aufgabe (14) vereint mühsam Fragestellungen, die wenig gemein haben.

Die Aufgaben sind durchweg eher hölzern formuliert, die Grafiken lieblos und fehlerhaft, die Lösungsskizzen wenig hilfreich und zum Teil falsch (Z.B. in gravierender Weise bei Aufgabe (3) und Aufgabe (5)). Innovative Anregungen gehen von solchem Material nicht aus. Warum sind diese Aufgaben, die die ‚Bildungsstandards für den Mittleren Schulabschluss' deutschlandweit exemplifizieren sollen, über die blass konturierten Kompetenzen hinaus deren Inkarnation darstellen, so voller Mängel? Die einzige rationale Antwort auf diese Frage liegt darin, dass es in den Standards nicht um Kompetenzen, Leitideen und Anforderungsbereiche geht, dass das Musterhafte dieser Aufgaben sich nicht auf ihren Inhalt bezieht. Es geht gar nicht darum, sie und ihre Lösungsmöglichkeiten sich gründlich anzuschauen, sie also ernst zu nehmen, sondern den Lehrpersonen und den Schülerinnen und Schüler klar zu machen, dass es jetzt einen neuen administrativ-zwingenden Begriff gibt, nämlich den der Standards gibt, den es ohne Widerworte einzuhalten gilt, der keinen Widerspruch ob gegen Tests oder Vergleichsarbeiten und deren Inhalte duldet. Ernst zunehmen sind also nicht die Aufgaben, sondern die Kandare, an die Lehrerinnen und Lehrer wie Schülerinnen und Schüler genommen werden: ihr müsst das jetzt können, sonst setzt es etwas, sei es durch Publikation der mageren Ergebnisse der Schüler, der Lehrer oder der Schule, sei es durch andere Zwangsmaßnahmen. Jetzt wird Ernst gemacht und dieser Ernst heißt eben Standard. Es mag schon sein, dass das Aufziehen dieser neuen Saiten – gleichsam als Lob der Ernsthaftigkeit staatlicher Bildungsvorgaben – manchem zu Pass kommt und mancher davon profitiert zum Beispiel als staatlich bestellter Bildungsforscher oder Testentwickler, aber Mathematikdidaktik ist das nicht. In der Vereinbarung über *Bildungsstandards für den Mittleren Schulabschluss (Jahrgangsstufe 10)* heißt es (auf Seite 4 in der zitierten Broschüre):

> Die Standards und ihre Einhaltung werden unter Berücksichtigung der Entwicklung in den Fachwissenschaften, in der Fachdidaktik und in der Schulpraxis durch eine von den Ländern gemeinsam beauftragte wissenschaftliche Einrichtung überprüft und auf der Basis validierter Tests weiter entwickelt. (S. 4)

Eine inhaltliche Weiterentwicklung hat seither nicht stattgefunden; offensichtlich besteht auch gar kein Bedürfnis nach einer breiten und tiefen fachlichen, fachdidaktischen oder schulpraktischen Diskussion.

Risse in der öffentlichen Geltungsmacht

Bei der sorgfältig arrangierten Veröffentlichung der ersten Pisa-‚Ergebnisse' in Deutschland wurde – wie in geringerem Ausmaß schon bei der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie (TIMSS) – in den Medien im Kern nur das Entsetzen über das Abschneiden der deutschen Schülerinnen und Schüler ausgerufen und in Szene gesetzt (‚Pisa-Schock'). Die Ergebnisse selbst, ihre Interpretation oder die angewandten Verfahren zu ihrer Gewinnung wurden auf den Pressekonferenzen, in den zugehörigen Berichten und Kommentare nicht einmal simpelsten Plausibilitätsprüfungen unterzogen. Ein Elchtest, der diesen komplexen Untersuchungsapparat im Ansatz oder auch nur die technischen Details des Tests in der Schule (zeitliche Länge, Art und Anzahl der Fragen) und dessen wunderliche Aussagekraft näher befragt hätte, blieb aus. Es galt nur das Ausmaß des deutschen Versagens zu beklagen und auf Abhilfe je nach Couleur des Kommentators und seiner Organisationszugehörigkeit zu sinnen. Zwar waren zuweilen recht ernüchternde Berichte von direkt an dem Test beteiligten Schülerinnen und Schülern sowie Lehrerinnen und Lehrern zu lesen, aber solche Augenzeugenkolportagen wurden als lokale Ausrutscher wider die Handbücher und vorgegebenen internationalen Verfahrensregeln bezeichnet und ihre Erwähnung oder Betrachtung als unwissenschaftlich gebrandmarkt. Sie gingen in der Dramatik und Wucht der globalen Untersuchung unter. Jegliche Kritik an Pisa wurde medial nur als ein untauglicher Versuch gesehen, die Misere der deutschen Bildung schön zu reden oder sie gar ganz zu leugnen. Die Geltungsmacht von Pisa hatte die Medien wie auch die Politik fest im Griff.

Bei der zweiten Pisa-Welle ließ sich keine vergleichbare Dramatik in den Medien mehr aufbauen. Auch halbherzige, bildungspolitisch forcierte Versuche, aus dem Vergleich der Ergebnisse der beider Durchläufe Schlüsse auf ein erstes Wirken deutscher Maßnahmen zu ziehen, erwiesen sich als verfahrenstechnisch gewagt und inhaltlich weder glaubwürdig noch überhaupt plausibel und in der Tendenz sogar kontraproduktiv. Nicht einmal die (unsinnigspektakulären) Länderrankings ließen sich noch verwerten, so dass ein neues Debakel die mediale Aufmerksamkeit sichern musste, dass nämlich in Deutschland die territoriale und soziale Herkunft in besonderer Weise auf die Bildungschancen ‚durchschlage'. Auch hier unterblieben übrigens einfache Nachfragen, wie denn dieses Forschungsergebnis zustande gekommen sei, welche Größen oder Indikatoren man hier gemessen, verrechnet oder gegeneinander aufgetragen habe und in welcher Weise die deutschen Ergebnisse

die vergleichbarer Länder über- oder untertrafen. Medial handelte es sich also nicht um ein Resultat einer komplexen Untersuchung, deren Verfahren zumindest im Groben zu erläutern seien, sondern um eine moralische Katastrophe, an deren Beseitigung man ohne Nachfrage und Aufschub zu arbeiten habe. [12]

Inzwischen ist auch die Blendkraft dieser Nachricht dahin. Der folgende Artikel zeigt beispielhaft, dass der schiere Glaube, dem jähen Entsetzen über das vermessene deutsche Schulbildungsdebakel würde sich nun mit der gleicher vollmundigen Bestimmtheit und Kennerschaft eine Besserung der Verhältnisse anschließen, in den Medien zu bröckeln beginnt.

Langer Anlauf ohne Sprung

Die wahren Pisa-Sieger sind gar nicht die Finnen. Die wahren Pisa-Sieger sitzen in Berlin, Dortmund und Bielefeld. In den Schulen sieht man sie selten. Meist brüten sie über Testbögen, ersinnen Prüfungsfragen oder erforschen mit Hingabe die Wirkung ihrer eigenen Forschung. „So viele Daten hatten wir noch nie“, freut sich der Bielefelder Erziehungswissenschaftler Klaus Jürgen Tillmann, Mitglied im deutschen Pisa-Konsortium, „als empirischer Bildungsforscher bin ich natürlich ganz begeistert.“ An Fördergeldern herrscht kein Mangel, neue Forschungsstätten werden gegründet, etwa das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) an der Berliner Humboldt-Universität. Allein der Forschungsgegenstand selbst dämpft noch die Wissenschaftlereuphorie: „Den Schulen bringt das leider nichts“, sagt Pädagoge Tillmann.

Gegen miese Testergebnisse, scheinen Deutschlands Schulminister zu glauben, helfe vor allem Testen. Zwar hat sich die Kultusministerkonferenz als Reaktion auf den Pisa-Schock sieben Verbesserungsstrategien vorgenommen – darunter Sprachkurse für Migrantenkinder, mehr Ganztagsschulen und gezielte Leseförderung –, doch konsequent umgesetzt haben sie bislang nur eine einzige: Tests. „Entwicklungen gibt es zwar in allen sieben Bereichen“, sagt Tillmann, „aber flächendeckend in allen Ländern sind nur die zentralen Prüfungen in den Schulen angekommen.“ (. . .)

Tatsächlich wird an den deutschen Schulen so viel evaluiert, verglichen und inspiziert wie nie zuvor. Schon vor der Einschulung müssen Vierjährige häufig zum Deutschtest antreten, in sieben Bundesländern schwitzen dann die Drittklässler über „Vera“ („Vergleichsarbeiten“) Tests, in der Mittelstufe folgen vielerorts weitere Vergleichsarbeiten. Dazwischen kommen alle Jahre wieder internationale Studien wie Pisa, Iglu oder Timss und je nach Land Erhebungen mit phantasievollen Namen wie „Quasum“, „Desi“, „Tosca“, „Markus“, „Ulme“ oder „Lau“. (. . .)

[12] Es geht hier keineswegs darum, deutsche Defizite im Umgang und in der Beschulung mit Schülerinnen und Schülern mit Migrationshintergrund (o.a.) in Abrede zu stellen, sondern darum deren heftige Moralisierung als einen wesentlichen Grund für die Existenzberechtigung von Pisa & Co zu akzeptieren.

„Nach Pisa wollte sich kein Kultusminister vorwerfen lassen, dass er nicht auf Leistung setzt“, erklärt Forscher Tillmann. „Dahinter steht die vage Hoffnung, dass vom Überprüfen auch alles irgendwie besser wird.“ Doch noch fehlt den Lehrerkollegien das Know-how, um aus der Datenflut Konzepte abzuleiten. „Da muss dringend was geschehen“, sagt Tillmann, „sonst bleibt das Ganze ein langer Anlauf, ohne dass gesprungen wird.“ (. . .)

Besonders weit auf dem Weg, sinnvolle Lehren aus den vielen Tests zu ziehen, glaubt sich Nordrhein-Westfalens Bildungsministerin Sommer. Sie rühmt ihr Schulsystem als das „modernste in Deutschland“. So will NRW als erstes Bundesland noch in dieser Legislaturperiode Schulrankings einführen. Zugleich können Eltern an Rhein und Ruhr neuerdings aussuchen, wo sie ihr Kind einschulen – und sich dabei möglicherweise an den Listen orientieren. „Wir wollen einen fairen Wettbewerb“, sagt Sommer.

Doch gerade darin sehen viele Wissenschaftler die größte Gefahr der neuen Testkultur: „Wenn die Schulen nur noch auf ihre Listenplätze schauen, findet überhaupt keine Schulentwicklung mehr statt“, warnt Wilfried Bos, Chef des Dortmunder Instituts für Schulentwicklungsforschung. Faire Rankings, die etwa den sozialen Hintergrund der Schülerschaft berücksichtigen, sind kaum möglich, wenn wie etwa bei Vera nach Herkunft und Familie der Kinder gar nicht gefragt werden darf.

Zudem erwies sich schon die Prämierung der Vera-Besten im vergangenen Jahr als Flop: Viele Schulen hatten sich gute Ergebnisse erschummelt – sie hatten die Testaufgaben vorher mit den Schülern trainiert (SPIEGEL 27/2006). „Wenn es erst mal richtige Rankings gibt“, glaubt Schulleiterin Borns aus Münster, „dann wird noch viel mehr gemogelt.“

Julia Koch in Der SPIEGEL 24/2007

Vermutlich werden solche Artikel die öffentliche Geltungsmacht von Pisa in Deutschland mehr erschüttern und eher zerrütten als eine wissenschaftliche Kritik an den Methoden und Verfahren der Untersuchung, die als unbedeutender innerwissenschaftlicher, von Laien angezettelter Zwist abgetan werden kann, die Öffentlichkeit kaum erreicht und eine Bildungspolitik, die sich Pisa gleichsam verschworen hat, nicht irritieren kann.

Literatur

American Evaluation Association AEA): Position Statement on HIGH STAKES TESTING In PreK-12 Education. 2002
(Online unter: http://www.eval.org/hst3.htm)

Kultusministerkonferenz (KMK): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (Beschlüsse der Kultusministerkon-

ferenz vom 02.06.2006). (Online unter: http://www.kmk.org/aktuell/Gesamtstrategie%20Dokumentation.pdf)

Kultusministerkonferenz (Hrsg.) in Zusammenarbeit mit dem Institut zur Qualitätsentwicklung im Bildungswesen: Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Berlin 2006.
(Online unter: http://www.kmk.org/schul/Bildungsmonitoring_Brosch%FCre_Endf.pdf)

Kultusministerkonferenz (KMK): Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 4.12.2003. (Online unter: http://www.kmk.org/schul/Bildungsstandards/Mathematik_MSA_BS_04-12-2003.pdf)

Koch, Julia: Langer Anlauf ohne Sprung. Der SPIEGEL 24/2007

Koretz, D.: Alignment, High Stakes, and the Inflation of Test Scores. Yearbook of the National Society for the Study of Education (2005) 104 (2), 99–118.
(Online erhältlich unter: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1744-7984.2005.00027.x)

Sill, H. D.: PISA und die Bildungsstandards. In: Jahnke, Th.; Meyerhöfer, W. (Hrsg.): Pisa & Co – Kritik eines Programms. Franzbecker Verlag. Hildesheim 2006. S. 293-330

Stecher, B. M.: Consequences of large-scale, high-stakes testing on school and classroom practices. In L. S. Hamilton, B. M. Stecher, and S. P. Klein (Eds.): Making Sense of Test-Based Accountability in Education. RAND. Santa Monica 2002. P. 79-100.
(Online unter: http://www.rand.org/pubs/monograph_reports/MR1554/index.html)