

EMPIRISCHE
BILDUNGSFORSCHUNG

NOTWENDIGKEIT UND RISIKO

Jörg-Dieter Gauger
Josef Kraus (Hrsg.)

I M
P L E N U M

ISBN 978-3-941904-33-0

www.kas.de

Diese Publikation dokumentiert die gleichnamige Veranstaltung, die der Deutsche Lehrerverband (DL) in Zusammenarbeit mit der Konrad-Adenauer-Stiftung am 3. September 2009 in Berlin durchgeführt hat.

INHALT

- 5 | VORWORT
- 7 | EMPIRISCHE SCHULFORSCHUNG AUS
BILDUNGSHISTORISCHER SICHT
Heinz-Elmar Tenorth
- 21 | DIE NATIONALE UND INTERNATIONALE BEDEUTUNG
EMPIRISCHER BILDUNGSFORSCHUNG
Rainer H. Lehmann
- 41 | PROBLEMATIK DER MESSINSTRUMENTE
AM BEISPIEL JÜNGERER SCHULSTUDIEN
Peter Bender
- 71 | BEITRAG DER BILDUNGSÖKONOMIE ZUR
SICHERUNG DER QUALITÄT VON SCHULE
Manfred Weiß
- 85 | PÄDAGOGISCHE EMPIRIE AUS
BILDUNGSPHILOSOPHISCHER SICHT
Volker Ladenthin
- 103 | HERAUSGEBER UND AUTOREN
- 104 | ANSPRECHPARTNER IN DER
KONRAD-ADENAUER-STIFTUNG

*Das Werk ist in allen seinen Teilen urheberrechtlich geschützt.
Jede Verwertung ist ohne Zustimmung der Konrad-Adenauer-Stiftung e.V.
unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen,
Mikroverfilmungen und die Einspeicherung in und Verarbeitung durch
elektronische Systeme.*

© 2010, Konrad-Adenauer-Stiftung e.V., Sankt Augustin/Berlin

*Gestaltung: SWITSCH Kommunikationsdesign, Köln.
Druck: Druckerei Franz Paffenholz GmbH, Bornheim.
Printed in Germany.
Gedruckt mit finanzieller Unterstützung der Bundesrepublik Deutschland.*

ISBN 978-3-941904-33-0

VORWORT

Die Konrad-Adenauer-Stiftung (KAS) und der Deutsche Lehrerverband (DL) sind in den zurückliegenden Jahren mit mehreren gemeinsamen Veranstaltungen und Publikationen auf aktuelle bildungspolitische Fragen eingegangen. Im Jahr 2007 etwa galt eine Kooperationstagung dem Thema „Bildungschancen für Migranten“. Die Veranstalter dieser Tagung, deren Vorträge und Empfehlungen in einer Broschüre dokumentiert sind, wollten unter Beteiligung von Botschaftern anderer Zuwanderungsländer die Chancen und die Versäumnisse bei der schulischen Integration junger Migranten ausloten. Ebenfalls im Jahr 2007 veröffentlichten KAS und DL zusammen einen Beitrag zum Thema „Perspektiven der schulischen beruflichen Bildung“, um die Bedeutung unseres Berufsbildungssystems, das in den üblichen bildungspolitischen Debatten zu kurz kommt, zu unterstreichen.

Anliegen der hier dokumentierten Fachtagung vom 3. September 2009 zur empirischen Bildungsforschung und der damit im Zusammenhang stehenden Bildungsökonomie war und ist es, nach innen und nach außen ein differenziertes Bild dieser in der Öffentlichkeit häufig sehr vergrößert dargestellten Forschungszweige und ihrer Ergebnisse zu vermitteln.

Wenn wir an die Versuche politischer Kräfte zurückdenken, noch im Sommer 1999 die erste PISA-Studie 2000 zu verhindern, dürfen wir für uns in Anspruch nehmen, damals mit Entschiedenheit zu denjenigen gehört zu haben, die in Sachen Bildung Bilanzen einforderten und die daran anschließenden Debatten mit drei größeren Diskussionsbeiträgen, allesamt veröffentlicht in der Schriftenreihe der Konrad-Adenauer-Stiftung, konstruktiv begleitet haben. Denn Bildung kann keine Veranstaltung im Elfenbeinturm sein, sie muss sich natürlich auch öffentlich bewerten und praktisch verwerten lassen.

Unter dem Eindruck einer bisweilen eindimensional anmutenden Ausrichtung so mancher Bildungspolitik auf das Messen, auf das Zählen und auf Quoten sowie auf Fragen der ökonomischen Verwertbarkeit sehen wir uns aber auch veranlasst, eine Stimme in eine andere Richtung zu erheben: Man möge doch bitte aufpassen, dass vor lauter Messen und vor lauter Verwertungsdenken nicht vergessen wird, was umfassende Bildung und ihr Sinn ist!

Kurz: Den Veranstaltern der Fachtagung geht es bei der Bildung um Ausgewogenheiten:

- zwischen Zweckorientierung und Zweckfreiheit,
- zwischen Bilanzierung und Freiraum,
- zwischen Verwertungsdenken und Bildungsauftrag,
- zwischen ökonomischen Anforderungen und kultureller sowie politischer Teilhabe,
- zwischen Zielstrebigkeit und Entschleunigung.

Noch kürzer könnte man es auch auf den Nenner bringen: Leistungsbilanzen und bildungsökonomische Analysen – ja! Das Ganze aber mit Maß und Mitte!

Möge die vorliegende Dokumentation einen kleinen Beitrag für diese Ausgewogenheiten leisten!

Sankt Augustin / Berlin / Bonn, im Dezember 2009

Prof. Dr. Jörg Dieter Gauger
Konrad-Adenauer-Stiftung e.V.

Josef Kraus
Deutscher Lehrerverband

EMPIRISCHE SCHULFORSCHUNG AUS BILDUNGSHISTORISCHER SICHT

Heinz-Elmar Tenorth

VORBEMERKUNG: THEMA UND EXEMPEL

Empirische Bildungsforschung gibt es nicht erst, seit wir alle – besorgt oder zustimmend – PISA-Daten rezipieren, auch nicht erst mit TIMSS oder IEA, sondern – je nach dem Begriff der Forschung – bereits seit mehr als 200 Jahren, wenn auch in unterschiedlicher Gestalt und Nähe zur aktuellen empirischen Bildungsforschung. Auch die empirische Schulforschung, die mit dem Titel meines Vortrags angesprochen wird, existiert als wesentlicher Teil der empirischen Bildungsforschung nicht erst seit gestern oder nur im 21. Jahrhundert. Diesen Ausgangsbefund kann der Bildungshistoriker beitragen, wenn über „Notwendigkeit und Risiko“ der empirischen Bildungsforschung diskutiert werden soll. Gleichzeitig fühle ich mich auch entlastet, dass ich nicht die gesamte empirische Bildungsforschung in ihrer eigenen Tradition und Praxis historisch zu analysieren habe, sondern mich auf die Schulforschung beschränken kann, also dann doch nur die thematische Linie aufnehmen muss, in der auch die aktuell Ärgernis erregenden aktuellen Schulleistungsstudien stehen.

Ich ignoriere also im Folgenden alle empirischen Untersuchungen über Bildungsprozesse, die beginnend mit Carl-Philipp Moritz und seinem *Magazin zur Erfahrungsseelenkunde* von 1783 zur Biografie von Lernenden wie zur Entwicklung von Kindern, zu Lebensverläufen in bürgerlichen oder adligen Familien oder zur diagnostischen Analyse von Gehörlosigkeit oder von Lernschwierigkeiten oder sozial auffälligen Verhaltens und seiner Therapie erschienen sind, und das ist ein kontinuierlicher Strang der intensiven empirischen Bildungsforschung seit dem 18. Jahrhundert, den ich damit ausblende. Aber diese Arbeiten gehören natürlich zur Bildungsforschung; denn sie thematisieren in empirischer Methodik und unter dem Anspruch objektiver, mit distinkten Methoden unternommener, realitätserhellender Forschung die Konstitution von Subjektivität im historisch-gesellschaftlichen Prozess – versprechen uns also nachprüfbar Informationen über das, was Philosophen über Bildungsprozesse behaupten, aber nicht nachprüfbar zu sagen wissen. Aber diesen Streit will ich jetzt nicht ins Zentrum rücken.

Ich konzentriere meine Überlegungen auf das kleine Segment der Schulforschung, hier ganz ohne großen Aufwand definiert als das Totum der Untersuchungen zu Lehr-/Lernprozessen und ihren Formen, Voraussetzungen und Folgen, wie sie sich im Kontext institutionalisierten Lernens und Unterrichtens in modernen Bildungssystemen beobachten lassen. Auch dann beschränke ich mich noch auf solche Studien, die das allgemeinbildende Schulwesen thematisiert und zum Forschungsgegenstand gemacht haben, und lasse erneut eine Fülle anderer Studien außer Acht, z.B. solche über das berufliche Schulwesen, die man spätestens in empirischer Wendung seit dem ausgehenden 19. Jahrhundert findet, auch Arbeiten über das Hochschulsystem, seine Praxis und seine Effekte, die spätestens mit dem sog. „Überfüllungsproblem“ und der Entdeckung der Hochschullehre und der „Hochschulpädagogik“ als Problem im ausgehenden 19. Jahrhundert einsetzen und bis heute andauern; und schließlich: ich sage nichts über das Lernen im Kontext der Erwachsenenbildung und des Erwachsenenalters oder die außerschulischen Bildungswelten.

Meine letzte Vorbemerkung betrifft die Materialgrundlage und die Fokussierung meiner Überlegungen: Im Wesentlichen werde ich – material, in den Datengrundlagen, wenn auch natürlich nur in sehr knappen Auszügen – die Situation in der deutschsprachigen Forschung und auch da im Wesentlichen nur die Entwicklung seit dem ausgehenden 19. Jahrhundert behandeln, konzentriert auf zwei signifikante Phasen: um 1900 und

um 1960/70. Dabei werde ich nur knapp die einschlägige Forschungspraxis zumindest der USA seit dem 20. Jahrhundert mit berücksichtigen; denn die USA waren zwar nicht die Erfinder der empirischen Schul- (und Bildungs-)forschung (das geschah eher in Westeuropa), aber sie haben doch seit dem 20. Jahrhundert zunehmend für zentrale Themen nicht allein theoretisch und methodisch, sondern auch im Aufbau einer eigenständigen Forschungsinfrastruktur eine bedeutsame Vorreiterrolle gespielt. Meine Blickrichtung ist historisch, d.h. zunächst in dem Sinne, dass ich die einschlägige Tradition der Forschung in ihrer eigenen Praxis vorstellen will, in der – schon angesichts der verfügbaren 30 Minuten unvermeidlichen – Auswahl an dem Kriterium orientiert, solche Exempel zu wählen, die für den Wechsel und die Abfolge von Theorien und Methoden ebenso signifikant waren und sind wie in ihrer Leistung und Wirkung für den bildungspolitischen Kontext; ich nehme dann, zum Ende hin und nur knapp, die empirische Schulforschung auch insofern als historisches Thema, als ich das Modell „empirische Schulforschung“ selbst historisiere und seinen Status im Prozess der Analyse von schulischen Bildungswelten und individuellen Konstruktionen von Bildungsprozessen distanziert beobachte.

1. EMPIRISCHE SCHULFORSCHUNG UM UND SEIT 1900

Auch wenn ich den Ruhm und die Praxis der Philanthropen so wenig schmälern will wie die einschlägigen Anstrengungen der Schulmänner des 19. Jahrhunderts: in Mittel- und Westeuropa hat die empirische Schulforschung ihre erste große Blüte seit dem ausgehenden 19. Jahrhundert und bis in die Zeit nach dem Ersten Weltkrieg. Deutsche Pädagogen haben gelernt, dass es zwei prominente Namen sind, die man nennen muss: Ernst Meumann, Wundt-Schüler und als Professor von Leipzig bis Zürich, Münster bis Hamburg tätig, und den Karlsruher Lehrer und Lehrerbildner Wilhelm August Lay. Sie werden bemüht, um das Phänomen „empirische Pädagogik“ und seinen Kontext zu bezeichnen, auch, um Kontroversen auszubreiten. Ich ziehe es vor, nicht so personalisiert zu denken, sondern von gesellschaftlich einflussreichen Organisationen, z.B. in Deutschland vom Bund für Schulreform bzw. dem Deutschen Ausschuss für Erziehung und Unterricht¹ auszugehen, dann von im Blick auf verschiedene Konferenzen und Kongresse, die der „Bund“ und andere Akteure, z.B. der „Verein für Kinderforschung“, seit dem ausgehenden 19. Jahrhundert veranstaltet haben. Dieser Zugang belegt deutlicher die gesellschaftlich für notwendig gehaltene Unterstützung „empirischer

Pädagogik“, sie zeigt damit den Wissensbedarf, den es um 1900 gab, und die über die Lehrer hinaus als problematisch eingeschätzten Themen und Probleme; und von hier aus kann man auch erkennen, dass einige der Zeitgenossen nicht ohne Grund das „Ende der philosophischen Pädagogik“ ausgerufen haben, weil die Philosophie – so wenig wie die traditionelle Schulmännerklugheit – die Probleme bearbeiten konnten, die sich in der Sozialisationsordnung stellten.

Die zentralen Themen dieser Zeit und damit zugleich die Problemlagen im Bildungssystem lassen sich sehr gut an den Kongressen des Bundes ablesen: 1910 widmet sich der erste Kongress zwei Themen: „Die Arbeitsschule“² sowie „Intelligenzproblem und Schule“³, der zweite Deutsche Kongress für Jugendbildung und Jugendkunde in München im Oktober 1912 widmet sich den Themen: *Das Wesen der Bildung, Die Schultypen und Die Vorbildung auf das Lehramt* – und man kennt die Kontinuität von Problemlagen bis heute. Der dritte Deutsche Kongress für Jugendbildung und Jugendkunde in Breslau im Oktober 1913 behandelt nur ein aber gewichtiges Thema: *Der Unterschied der Geschlechter und seine Bedeutung für die öffentliche Jugendziehung* – und es ist keineswegs eine traditionelle oder gar antifeministische Perspektive, die hier eingenommen wird, sondern eher eine differenztheoretische Argumentation, die man fast schon in Gender-Begrifflichkeit übersetzen könnte.

Der Deutsche Ausschuss für Erziehung und Unterricht, wie sich der „Bund“ im Ersten Weltkrieg nennt, stellt in den kriegerischen Zeiten keineswegs seine Arbeit ein, sondern erweitert seine Interessen um ein neues Thema: *Der Aufstieg der Begabten* (und Peter Petersen ediert den einschlägigen Tagungsband – Leipzig/Berlin 1916). Der Kriegslage pädagogisch entsprechend ist dieses Thema aus der Absicht der Klassenversöhnung geboren, orientiert an dem Gedanken, auch in den Unterschichten die Begabten zu identifizieren und den Vorwurf des sozialen Klassenschulsystems auszuräumen.

Man erkennt in der Sequenz der Themen auch sehr gut die Intention, der die empirische Schulforschung im Ursprung um 1900 folgte, d.h., in der Formulierung des Bundes, in der Absicht, „das gesamte Bildungswesen im Sinne der Anpassung an die Natur der jugendlichen Persönlichkeit umgestalten zu helfen“ und „folgerichtig die Forderung“ zu bedienen, „dass sich die pädagogischen Reformen auf eine wissenschaftliche Erkenntnis des jungen Menschen gründen sollen“. Verwissenschaftlichung

der Schulpolitik und d.h. gleichzeitig Entideologisierung und Objektivierung der Lehrerpraxis. Die Fragen des ersten Kongresses heißen deshalb: „1. Entspricht die mit dem Schlagwort ‚Arbeitsschule‘ gekennzeichnete Art des Unterrichts- und Erziehungsbetriebes den Zielen des Bundes? 2. Berechtigen und nötigen die bisherigen Ergebnisse der wissenschaftlichen Jugendkunde den Bund zur Unterstützung von Vorschlägen, die auf Organisationsänderungen im praktischen Betrieb des öffentlichen Bildungswesens abzielen?“⁴

In ähnlicher Weise denkt und arbeitet auch der „Verein für Kinderforschung“, mit leicht veränderten Adressaten im Bereich der als „Hilfsschulpädagogik“ und der Fürsorge adressierten Klientel. Auf seine Einladung treffen sich z.B. 1906 in Berlin Anstaltspädagogen zusammen mit Heilpädagoginnen⁵ und diskutieren über Begabung und Intelligenz, Förderung der Kinder und Fürsorge. Das erinnert zugleich daran, dass die ersten Impulse zu einer empirisch orientierten, an der Psychologie theoretisch und methodisch geschulten empirisch orientierten Erziehungswissenschaft schon früh von den Überlegungen zur „Schulhygiene“⁶ ausgegangen sind, von Ärzten ganz wesentlich beeinflusst, aber auch von der Schulverwaltung.

In einem umfassenden disziplinären Sinne kommt die Inspiration der empirischen Pädagogik dann aber aus der Psychologie, z.B. schon 1899 bei dem Berliner Oberlehrer Ferdinand Kemsies. Er kooperierte mit Carl Stumpf, dem empirisch arbeitenden Psychologen und Dilthey-Kollegen an der Berliner Universität, fühlte sich von der Kinderforschung inspiriert, arbeitet aber von Beginn an auch schulbezogen. Kemsies' Zeitschriften-gründung, die *Zeitschrift für Pädagogische Psychologie*, setzte deshalb auch das erste sichtbare Zeichen für eine Abkehr der pädagogischen Forschung und Reflexion nicht nur von der herbartianischen, sondern auch von der nur philosophischen Pädagogik.⁷ Für die Intensität der Beziehungen von empirischer Pädagogik und Lehrerschaft, die sich darin andeutet, sind vor allem aber die Aktivitäten der Lehrer seit dem ausgehenden 19. Jahrhundert selbst ausschlaggebend. Das zeigt sich vor allem in der Gründung eigener Laboratorien⁸, in denen sie die empirische Schulforschung unterstützten und die Effekte empirischer Arbeitsformen nutzten, denn „die experimentelle Arbeit bringt ja stets Arbeitsgemeinschaft und Arbeitsteilung mit sich“.⁹ Solche Laboratorien nehmen Vorbilder aus dem Ausland auf, wie sie u.a. in der 1896 gegründeten Laboratory School in Chicago¹⁰ gegenwärtig sind, aber auch aus der eigenen

Lehrertradition, wie sie in den herbartianischen Seminaren bestanden haben, die Forschung und Praxis, meist in einer Übungsschule, miteinander verbunden haben (obwohl die späteren Institute und Laboratorien meist nicht über eine Versuchsschule verfügten, sondern Kooperationsbeziehungen eingingen). In Deutschland sind sie nach solchen Vorbildern auch nicht allein in Leipzig, dem meist genannten Ort, und seinem „Institut für experimentelle Pädagogik und Psychologie“ Lehrer und Forscher gemeinsam tätig, sondern auch in anderen Orten: in Frankfurt, getragen von der Vereinigung für Kinderkunde, das „Pädagogisch-psychologische Institut München“ arbeitet seit 1910 als Institut des Bezirkslehrervereins und in Kooperation mit Aloys Fischer, das „Institut für Jugendkunde“ hat in Bremen der Oberlehrer Theodor Valentiner gegründet und seit 1913 gibt es das gleichnamige „Institut für Jugendkunde“ in Hamburg, an dem Meumann, Deuchler und die Sterns, die prominenten pädagogischen Psychologen des ersten Drittels des 20. Jahrhunderts also, mit ihrem Schülerkreis bis hin zu Martha Muchow, beteiligt waren.

Die wesentlichen Ziele, Akteure und Referenzfelder der empirischen Schulforschung um 1900 zeichnen sich so deutlich ab, dass ich ein erstes Resümee versuchen kann: Es sind Lehrer, Schulverwaltungsleute, Administratoren, außerschulische Pädagogen, Ärzte und Wissenschaftler im Umkreis v.a. der Psychologie, die „empirische Schulforschung“ betreiben. Der „Bund für Schulreform“, man kann ihn als Beleg für die Netzwerke dieser Forschungspraxis nehmen, wurde im Oktober 1908 in Berlin gegründet, um alle zu organisieren, „die am Fortschritt der nationalen Bildungsarbeit in Schule und Haus lebhaft interessiert“ sind.¹¹ Die reformorientierten Volksschullehrer organisieren sich entsprechend in der „Pädagogischen Zentralstelle des DLV“, später wird die „Erziehungswissenschaftliche Hauptstelle“ des DLV der Ort der Kooperation von Lehrerschaft und Wissenschaft; die katholischen Lehrer gründen dafür das „Deutsche Institut für wissenschaftliche Pädagogik“, die preußische Staatsregierung noch im Weltkrieg das „Pädagogische Zentralinstitut“ in Berlin, das dann nach 1918 die Adaptation der Intelligenztests für die deutsche Bildungsszene leistet (mit Hylla, Bobertag u.a. also sehr aktiven preußischen Schulreformern).

Im Blick auf diese Akteure und die weitere Geschichte ist allerdings auch sichtbar, welche Umgewichtung sich bei den Adressaten und Rezipienten bald nach 1910 ereignet: Die Lehrer in der Praxis entwickeln bald Distanz, weil die Entlastung im Alltag, auch die Legitimation im Alltag,

z.B. bei Selektionsentscheidungen, dann doch nicht kommt, die man sich früh versprochen hatte; die neue Forschung ist eher auf der Ebene der Leitung und Steuerung gefragt, sie wird dort rezipiert, weniger im Alltag der Lehrerarbeit selbst – dafür ist sie nicht relevant genug.

Das überrascht etwas, weil die empirisch-pädagogische Forschung in ihren leitenden Ideen nahe bei der Alltagspraxis zu arbeiten scheint, denn in ihren Methoden und Theorien ist die frühe empirische Schulforschung im Wesentlichen psychologisch (und in der Auswertungsmethodik statistisch), aber in den Fragestellungen eher weniger soziologisch, auch wenn man die „Milieu“-bezogenen Studien etwa von Adolf Busemann nicht ignorieren darf oder die professionsbezogenen Analysen bei Aloys Fischer oder gesellschaftskritische Arbeiten bei Siegfried Bernfeld. Aber das inspiriert eher die Kritik, als die berufliche Realität der Lehrer (anders als die Lehrerverbände in ihren Organisationen).

Unbeschadet des Problems der Alltags- und Praxisrelevanz – diese Studien in Deutschland und Europa haben als Forschungsprogramme und -leistungen internationale Geltung, ja Vorbildcharakter. Die *Themen* und die zugehörigen Methodenüberlegungen und Forschungspraktiken, die sich z.B. in den Arbeiten von Meumann finden lassen, sind typisch für die Zeit und die Arbeit der empirischen Pädagogik und der Pädagogischen Psychologie (und ich nehme nur die dreibändigen „Vorlesungen“ als Referenz): Bd. 1: *Körperliche und geistige Entwicklung des Kindes und Jugendlichen, der einzelnen Fähigkeiten – in der Schule: Gedächtnis, Vorstellung, Sprache, Gefühl, Willen*; Bd. 2: *Individuelle Unterschiede – Begabung und Begabungslehre und -forschung, Intelligenz, pädagogische Bedeutung*; Bd. 3: *„Geistige Arbeit des Schulkindes und ihre Beziehung zur Methodik des Lehrers“, „Geisteshygiene der Schularbeit“, spezielle Didaktik der Unterrichtsfächer: sprachlich, Schreiben, Rechnen, Zeichnen, „höhere Unterrichtsfächer“*. Man sieht, dass die Mikro-, Meso- und Makroebene des Bildungssystems berücksichtigt wird, wenn auch mit Schwerpunkt auf der Ebene der unterrichtlichen Interaktion – aber auch das schmälert die Distanz zur Lehrerschaft nicht.

Es gibt eine Leerstelle, eine eher vernachlässigte Thematik, und das ist das Curriculum: Diese Fragen sind in Deutschland eher noch Streitthema in der politischen Arena, im Wesentlichen vorentschieden durch die Schulstruktur und die Typen der höheren Schule, die sich ja über Lehrpläne definieren und abgrenzen – anders als z.B. in den USA, wo

die Curricula bereits im frühen 20. Jahrhundert intensiv behandelt worden sind, auch flächendeckend, genauso wie es hier auch schon früher Test-Studien gab, weil die nationalen Prüfungen und ihre Standardisierung, die man in Deutschland im Abitur hat, hier ja vollständig fehlen, deshalb ersetzt die Testpraxis in zentralen Aspekten (wenn auch nicht beim Übergang in Hochschule oder Beruf) das deutsche Berechtigungssystem – Ausleseprüfungen und ihre testtheoretische Optimierung fehlen allerdings auch in Deutschland nicht. In Berlin waren z.B. die Fragen der Übergangsauslese in das Gymnasium die Kollegen Moede, Piorkowski und Wolf sehr aktiv oder in Mannheim Anton Sickinger bei der Begründung von Leistungsdifferenzierung; und die Auslesediagnostik hat auch ein starkes Standbein in der Sonderpädagogik – aber noch diese Arbeiten belegen die kulturelle und bildungssystemische Differenz: trotz der Anstrengungen des Zentralinstituts kommt es nie zu einem *Educational Testing Service*, wie wir ihn seit 1947 in den USA haben – und heute mit allen Folgen diskutieren können.¹²

Differenzen zu Deutschland zeigen sich in der Forschungspraxis der USA außer in den Themen auch bei den Akteuren, die gesellschaftliche Praxis und ihre Organisationen – und zwar jenseits der Lehrerschaft – sind hier sehr viel stärker beteiligt: Es sind die großen Stiftungen, die diese Forschung befördern: die Russell Sage Foundation, Rockefeller-Foundation, Carnegie-Foundation, National Science Foundation (etc.). Mit der Einrichtung des *Educational Testing Service* (1947) betritt man aber schon die zweite Phase, die ich diskutieren will. Sie zeigt nach 1945 dann eine Internationalisierung der Schulforschung, v.a. in den Theorien und leitenden Modellen, in den Methoden, Verfahren und Techniken. Und Deutschland schließt sich – insgesamt mit ein wenig Verspätung, mit dem DIPF in Frankfurt aber schon seit 1950 – bald diesem Trend an. Schulforschung gehört jetzt zum Standard der nationalen Wissenschaftssysteme und – in Deutschland zuerst mehr außeruniversitär – zu den basalen Einrichtungen der Beobachtung der nationalen Bildungssysteme und ihrer politischen Gestaltung.

2. EMPIRISCHE SCHULFORSCHUNG UM UND SEIT 1960: DAS BÜNDNIS VON EGALITÄR ORIENTIERTER REFORM UND BILDUNGSFORSCHUNG

Wir kennen und nennen für die deutsche Situation alle das dann leitende, stilprägende Gremium, das ist der Deutsche Bildungsrat, mit den Gutachten und Studien, die er seit 1965 in Auftrag gegeben hat. Das waren insgesamt mehr als 50 und keineswegs nur „Begabung und Lernen“, die 1968 als Band 4 der Gutachten erschienene Bibel der egalitär und gesamtschulisch orientierten Schulreform. Wir kennen von daher auch die universitären und außeruniversitären Umfuleinrichtungen der empirischen Bildungsforschung von Frankfurt bis Berlin, von Konstanz bis Hamburg, von Landau bis Dortmund, die Universitäten und Wissenschaftler im Umkreis der sozialdemokratischen Bildungsreform, die deutsche Gesellschaft für Erziehungswissenschaft und ihre Arbeitsgemeinschaft für empirisch-pädagogische Forschung, die Soziologen und alle weiteren; aber man sollte auch andere Institutionen nicht übersehen, das Comenius-Institut der evangelischen Kirchen z.B., das Pädagogische Zentrum in Berlin als Muster eines forschenden, reformstützenden Landesinstitute (etc.).

Sie, diese Akteure und Institutionen bestimmen die leitenden *Themen*: Auslese als Focus, der Zusammenhang von sozialer Herkunft und Schulerfolg, die Rolle von Bildungssystem und Profession in diesem Prozess, internationale Systemvergleiche, Curriculumstudien in unterschiedlicher theoretischer Modellierung zwischen Klafki und Robinsohn, die Bildungsökonomie, aber auch wieder psychologisch orientierte Instruktionsforschung (Prozess-Produkt-Paradigma; Behaviorismus vs. Kognitionspsychologie), komparative Gesamtschuluntersuchungen, exemplarisch bei Helmut Fend, in denen soziologische und psychologische, professionsbezogene und curriculumtheoretische Arbeiten gebündelt wurden, und natürlich: ein Begabungskonzept, das unmittelbar nach der Gesamtschule zu rufen schien.

Die einschlägigen Handbücher beschwerten noch heute die Regale der ehemaligen Studierenden und sie legen Zeugnis ab von der *Einheit im Forschungs- und Gestaltungs-, ja Veränderungswillen* der damaligen Adepten der Bildungs-Reform und der theoretisch-methodischen Innovation; ja, es schien sogar so, als sei die empirische Bildungsforschung eindeutig auf Seiten der Reformen zu platzieren. Das war aber nicht nur

Politik, dabei wurden auch neue theoretische und methodische Ambitionen realisiert. Die Karriere der einschlägigen Begriffe, vor allem der von Interaktion, Kommunikation und Sozialisation legt davon in einer breiten Öffentlichkeit Zeugnis ab; eher schon expertenhaft bedeutsam dagegen war das begleitende methodische Interesse. Fragen von Korrelation und Kausalität, Probleme der Messung wurden ebenso diskutiert wie kausalanalytische Attribuierungen, die sich mit den Pfadanalysen eröffneten (Walter Müllers Arbeiten über die Reproduktion sozialer Ungleichheit waren ein Meilenstein aller Debatten), und die Attributionsforschung hat uns die Welt anders sehen gelehrt.

Mit den späten 1970er Jahren zögerlich, sicherlich deutlich in den 1980ern und danach breit und nachhaltig sind wir aus dem verbreiteten euphorischen Taumel der Allmacht und des Allwissens in der Gestaltung der Bildungswirklichkeit, wie sie durch empirische Forschung ermöglicht schien, ebenso erwacht wie aus einer Form der vermeintlich wissenschaftlich beglaubigten Kritik. Diese Art der Kritik, Kultur- und Ideologiekritik, erweist sich retrospektiv – so kann man heute bei Ralf Konersmann im Blick auf Kulturkritik lesen und für die Pädagogik übersetzen – als eine Kritik, die „sich als Inhaberin des überlegenen Standpunktes wähnte – eines Standpunktes, der sich klassischerweise auf Mastersubjekte wie die Wahrheit, die Vernunft und die Geschichte berief. Mit dieser Art der Kritik und dem Gestus der starken Behauptung, der Einschüchterung und der Unterwerfung, ist es nun vorbei.“¹³

Insofern, es war nicht etwa nur die empirische Bildungsforschung, die solche Ambitionen genährt hat, es war die Bildungsforschung insgesamt, auch die Erziehungsphilosophie, die überzogene Versprechen und überzogene Erwartungen in die Welt gesetzt hat – und deshalb enttäuschte Hoffnungen als Erbe hinterlassen musste. Der Realismus der aktuellen empirischen Bildungsforschung rührt auch daher. Sie versucht die Politik nicht mehr zu überbieten oder zu ersetzen, sondern beansprucht nur die Form einer methodisch kontrollierten Beobachtung der Wirklichkeit, als die man Wissenschaft beanspruchen kann.

3. HISTORISIERUNG UND KONTEXTUALISIERUNG

Die beiden Exempel aus vergangenen Gegenwarten erlauben ein historisch begründetes Fazit über Form und Leistungsfähigkeit der empirischen Schulforschung, quasi „Über Nutzen und Nachteil der Empirie für

das pädagogische Leben“, vielleicht auch nur für das Bildungssystem und den Bildungsprozess. Ich tue das eher in Stichworten, gelegentlich auch etwas zugespitzt, damit die kritische Diskussion ein Objekt hat:

Notwendigkeit und Funktion sind eindeutig: in Krisenphasen und bei Reformhoffnungen setzen Gesellschaften wie unsere seit dem ausgehenden 18. Jahrhundert punktuell, seit dem 19. Jahrhundert systematisch auf empirische Bildungsforschung. Sie ist Teil des breiten Prozesses der „Verwissenschaftlichung des Sozialen“¹⁴, und dann wird in gleicher Weise Wissen und Legitimation nachgefragt und erzeugt, Steuerungsrelevanz und Alltagsunterstützung erwartet.

Das definiert auch die wesentlichen *Referenzsysteme*: Ausgehend bei den Praktikern und der Bildungsadministration gibt es eine breite Palette von Rezipienten, aber die Nachfrage ist schwankend, Praxisrelevanz nicht dauerhaft gegeben, selbst bei den scheinbar handlungsnahen Forschungen psychologischen Typs zeigt sich immer wiederkehrend auch Distanz der pädagogischen Akteure gegenüber den Beobachtern aus der Bildungsforschung. Unverkennbar sind *Themen und damit auch Theorie- und Methodenpräferenzen* – jenseits der Tatsache, dass es um Schulforschung geht – auch eindeutig national und kulturell geprägt, abhängig von den Problemen und Defiziten, die je nationale Bildungssysteme und der Alltag der Professionen erzeugen: Es gibt immer Auslese- und Übergangsforschung, aber wer das Abitur hat, kann sehr lange auf extensive Testdiagnostik und -forschung verzichten, wer landesweite und in Schultypen abgestützte Curricula kennt, muss nicht hilfswiese die Curriculumforschung bemühen (etc.).

Auch die *Forschungsinfrastruktur* ist variabel: Außeruniversitäre Institute, Universitäten, projektbezogene und drittmittelbezogene Forschung, Planungsgremien – wie den Bildungsrat – das gibt es überall, die Gewichtung ergibt sich aus Traditionen der Universitäten, der Sozialwissenschaften, der Forschungsfinanzierung (etc.). Ein Einheitsmodell ist auch bis zum Ende des 20. Jahrhunderts dabei nicht entstanden, die Argumentationskultur in der Bildungsforschung kennt eindeutig nationale Stile.¹⁵

Für die Gründungsphase und dann erneut nach 1960 ist das *Bündnis empirischer Bildungsforschung mit der reformorientierten Fraktion der pädagogischen Profession* überraschend stark, und zwar international. Das ist offenbar alter positivistischer Geist: *Savoir pour prévoir!* Auf der

Ebene von Organisationen und Verbänden und in der Programmrhetorik relativ stabil, findet dieses Bündnis aber im Alltag der Lehrer keine dauerhafte Unterstützung, hier regieren andere praxisbezogene, Wissenssysteme mit anderen Gütekriterien und einer anderen Programmrhetorik (die in der „Aktionsforschung“ unseligen Angedenkens eine Form gefunden hatte, die jeden Standard empirischer Schulforschung im politischen Bewusstsein der eigenen Fortschrittlichkeit manifest dementierte).

Den Bündnispartnern und ihren Erwartungen entsprechend, aber auch relativ zur eigenen Programmatik changiert die *Funktion und Wirkung empirischer Schulforschung* konstant zwischen der Objektivierung und distanzierter Analyse von Problemlagen und -dimensionen und der offenkundigen Ideologieaffinität. Damit wird auch das Risiko dieser Art von Forschung bewusst: sie tendiert dazu, ihren eigenen Modus der Erfahrung der pädagogischen Welt zu monopolisieren, andere Reflexionsformen, z.B. die Klugheit der Akteure und der Erfahrung ihrer eigenen Praxis, abzuwerten und als Wissenschaft nur gelten zu lassen, was ihrem eigenen Wissensmodell entspricht - entsprechend ist die Geschichte der empirischen Schulforschung immer begleitet von metatheoretischen Kontroversen.

Dazu gehören natürlich auch die klassischen Einwände der Erziehungsphilosophie gegen die Empirische Pädagogik, die Kritik an ihrer normativen Enthaltensamkeit und die Unterstellung, sie zeige besondere Anfälligkeit gegenüber Weltanschauungen, zumal naturalistischen. Eduard Spranger hat diesen Vorwurf schon 1915 gegen die Kinderpsychologie vorgetragen, als die er die empirische Pädagogik codierte:

„Die Kinderpsychologie ist doch nur ein kleiner Ausschnitt aus dem großen Reich der Pädagogik, und wer wirklich erziehen will, der muss nicht nur das Kind verstehen, sondern vor allem das große Leben rings um ihn herum, das uns trägt, und das er hineinbilden soll in werdende Herzen, um ihnen Wert zu geben und Inhalt. Auf diese Werte und diesen Inhalt drängt die neue Pädagogik hin. Auch wenn die Kinderpsychologie weniger handwerksmäßig und mit mehr Geist getrieben würde, als es meist der Fall ist, auch dann würde sie an solche Aufgaben nicht heranreichen. ... Es ist ein Irrtum, dass eine Weltanschauung gewonnen werden könnte von naturwissenschaftlicher Grundlage aus.“¹⁶

Weltanschauung, das aber ist der Wissensbedarf, den Spranger nicht nur im Kriege als primär definiert, dem er vielmehr seine ganze öffentliche Arbeit als Erziehungsphilosoph gewidmet hat. Wissen dieser Art erzeugt dann seine eigene Geltung, aber auch eindeutige Risiken: „Auch der Lehrer lebt nicht von den Brosamen der Methoden und Experimente, sondern von einer jeglichen großen Bewegung, die sein Herz ergreift. Aus diesem Ergriffensein folgt von selbst unsre Arbeit für 1915: Die Beschäftigung mit den Wissenschaften der Kultur und die Deutung der Pädagogik nicht als Psychologie oder gar Naturwissenschaft, sondern als einer Kulturwissenschaft, die die Wege zeigt, wie das große objektive Leben mit seinen Inhalten und Werten in die Jugend hineingebildet werden kann.“

Wir kennen das Problem dieses schönen Zitats: Was hier als Kulturwissenschaft 1915 angepriesen wird, ist nicht anders als preußische Staatsmetaphysik in scheinmodernem Gewande. Oder, und zur Rettung der hier kritisierten empirischen Pädagogik: Über die Leistung der empirischen Schulforschung wird nicht in ideologiekritischen Zuschreibungen entschieden, die ihre eigenen Risiken nur schwer kaschieren können, sondern nur an der Forschungspraxis selbst – und dann wissen wir. Sie beobachtet klug, aber sie belehrt pädagogisch, also in den wichtigen pragmatischen Kontexten nur sehr begrenzt.

- 1| Korrekt heißt er: „Bund für Schulreform des Allgemeinen Deutschen Verbandes für Erziehungs- und Unterrichtswesen“, nach 1914 publiziert er als „Deutscher Bund für Erziehung und Unterricht“.
- 2| Erster Deutscher Kongreß für Jugendbildung und Jugendkunde zu Dresden am 6., 7. und 8. Oktober 1911 – Erster Teil. Die Arbeitsschule. Vorträge und Verhandlungen am Freitag, dem 6. Okt. 1911. Leipzig/Berlin 1912 (Arbeiten des Bundes für Schulreform 4).
- 3| Erster Deutscher Kongreß für Jugendbildung und Jugendkunde zu Dresden am 6., 7. und 8. Oktober 1911 – Zweiter Teil. Intelligenzproblem und Schule. Leipzig/Berlin 1912 (Arbeiten des Bundes für Schulreform 5).
- 4| Erster Deutscher Kongreß, 1912, a.a.O., Erster Teil, Vorwort.
- 5| Schaefer, K.L. (Hg.) (1907): Kongress für Kinderforschung und Jugendfürsorge in Berlin (1.- 4. Oktober 1906). Langensalza.
- 6| Die Beziehung von Medizin und Pädagogik analysiert Stroß, A.M. (2000): Pädagogik und Medizin. Weinheim, für die Zeit um 1900 bes. 228ff.
- 7| Zu Kemsies bereits Kallendorf, F. (1975): Die Hinwendung der Pädagogik zu den Erfahrungswissenschaften, gespiegelt am Ansatz der „Zeitschrift für Pädagogische Psychologie“. Diss.phil. PH Westfalen/Lippe (Bielefeld).
- 8| Zur Geschichte der Laboratorien u.a. Ingenkamp, K. (1987): Das Institut des Leipziger Lehrervereines 1906-1933 und seine Bedeutung für die Empirische Pädagogik. In: Empirische Pädagogik 1, 60-70 sowie zum gleichen Institut und

- seinem professionspolitischen Kontext Naumann, G. / Pehnke, A. / Uhlig, C. (1987): *Das Ringen des Leipziger Lehrervereins um allseitige Lehrerbildung*. In: *Wiss. Zeitschrift der PH ‚Clara Zetkin‘ Leipzig III*, 36-40 (dort zum Institut 38f.); als Übersicht zur Gesamtheit der „kommunikativen und organisatorischen Infrastruktur“ Dudek (Anm.:11) 90ff., zu den Instituten auch Hopf (Anm.:11) 87ff. (zu Leipzig, München und Hamburg).
- 9| So Meumann im Vorwort zum 1. Bd. seiner Vorlesungen 1907 (vgl. Meumann, 1916, III).
- 10| Zu deren Geschichte und zur Rolle Deweys und seiner Frau bis 1904, als sie die Schule verließen, vgl. DePencier, I.B. (1967): *The History of the Laboratory Schools. The University of Chicago 1896-1965*. Chicago.
- 11| Die beste und umfassendste Darstellung liefert Dudek, P. (1990): *Jugend als Objekt der Wissenschaften. Geschichte der Jugendforschung in Deutschland und Österreich 1890-1933*. Opladen, 100-115, zit. 100; das kurze Kapitel zum „Bund für Schulreform“ bei Hopf, C. (2004): *Die experimentelle Pädagogik. Empirische Erziehungswissenschaft in Deutschland am Anfang des 20. Jahrhunderts*. Bad Heilbrunn, 252-254 kann dagegen nicht überzeugen, schon weil es zu sehr aus der Perspektive Meumanns geschrieben ist und aus der Hamburger Situation heraus die Rolle des Bundes verzeichnet, ansonsten aber – obwohl sie eingangs auf ihn verweist – den Reichtum der Informationen leider nicht nutzt, der seit Dudek bereitliegt.
- 12| Lemann, N. (1999): *The Big Test: The Secret History of the American Meritocracy – und die umfangreiche und kritische Diskussion zu diesem Buch*, u.a. in: *The Atlantic*, 7. Oct. 1999.
- 13| Konersmann, R. (2008): *Kulturkritik*. Frankfurt a.M., dort in der Einleitung die Zitate.
- 14| Raphael, L. (1996): *Die Verwissenschaftlichung des Sozialen als methodische und konzeptionelle Herausforderung für eine Sozialgeschichte des 20. Jahrhunderts*. In: *Geschichte und Gesellschaft* 22, 165-193.
- 15| Vgl. meinen Beitrag (2009): *Struktur der Erziehungswissenschaft*. In: *Andresen, S. u.a. (Hg.): Handwörterbuch Erziehungswissenschaft*. Weinheim/Basel, 850-865.
- 16| Spranger, E. (1915): *Zum Geleit für 1915*. In: *Die Deutsche Schule* 19, 1-5 (ND in Herrlitz, 1987, a.a.O.).

DIE NATIONALE UND INTERNATIONALE BEDEUTUNG EMPIRISCHER BILDUNGSFORSCHUNG

Rainer H. Lehmann

Empirische Bildungsforschung im Sinne einer systematischen Erfassung, Analyse und Erklärung der Dynamik eines Bildungssystems hätte in Deutschland längst einen viele Jahrzehnte übergreifenden Traditionsstrang bilden können, wäre es hier nicht – ähnlich wie in der Soziologie – zu zeitgeschichtlich bedingten Brüchen gekommen. Deren Überwindung hat es den deutschen Kolleginnen und Kollegen erst in neuerer Zeit erlaubt, Anschluss an die internationale *scientific community* zu finden.

Der Tübinger Soziologentag 1961 mit der Auseinandersetzung zwischen Karl Popper und Theodor Wiesengrund Adorno (vgl. Maus / Fürstenberg 1972) hat indirekt eine ganze Generation von Vertretern des Fachs Pädagogik an den Hochschulen der (alten) Bundesrepublik mit methodologischen Grundüberzeugungen ausgestattet, die Karlheinz Ingenkamp 1989 als „Test-Aversion des deutschen Intellektuellen“ gebrandmarkt hat. Lange hat solche undifferenzierte Ablehnung quantitativer Lernstandsfeststellungen, die die Grundlage der empirischen Bildungsforschung bilden, die autochthone Entwicklung dieser Disziplin nicht nur eingeschränkt, sondern geradezu verhindert.

Der Umstand, dass die empirische Bildungsforschung als Wissenschaftszweig mit klar definierten Rationalitätskriterien ihren Weg in die deutsche Hochschul- und Politiklandschaft erst auf dem Wege des Re-Imports aus dem dominanten anglophonen Kontext gefunden hat, verlangt nahezu zwingend die Umkehrung des hier vorgegebenen Themas. Ehe gründlicher von der nationalen Bedeutung empirischer Bildungsforschung gehandelt werden kann, ist zunächst deren internationale Bedeutung zu thematisieren. Erst auf dem Umweg über den Ansatz einer empirisch fundierten vergleichenden Erziehungswissenschaft mit dem Leitkonzept der Welt als Experimentierfeld und seinen Implikationen – vgl. die These von Benjamin Bloom, Arnold Anderson, Mary Jean Bowman, Torsten Husén u.a. zur „*world as a single educational laboratory*“ (vgl. Heyneman 2003; Husén / Postlethwaite 1967, 27) – hat sich das bildungspolitische Potenzial dieses Ansatzes auch in Deutschland durchgesetzt.

Deshalb soll in einem ersten Schritt anhand der großen international vergleichenden Studien die Reichweite empirischer Bildungsforschung umschrieben werden, die selbst ohne die Entwicklung solcher Spezialfelder wie der vergleichenden Bildungsökonomie inzwischen sehr beträchtlich ist.

In einem zweiten Schritt gilt es, die analytischen Möglichkeiten und bildungspolitischen Implikationen zu skizzieren, die sich aus der Existenz der so generierten Datensätze ergeben.

Drittens schließlich sollen die so begründeten Positionen auf den nationalen Kontext bezogen werden, auch unter dem Gesichtspunkt wachsenden Interesses an regionalen Fragestellungen.

1. ZUR REICHHALTIGKEIT DES DATENMATERIALS INTERNATIONALER BILDUNGSFORSCHUNG

Es gehört zu den allgemeinen Grundüberzeugungen, zum populären Diskurs weit über die Zünfte der Erziehungswissenschaft und der Bildungspolitik hinausgehend, dass die Hauptkriterien schulischer Bildungsbemühung bzw. Qualifikationsanstrengung in den „Domänen“ Leseverständnis, Mathematik und Naturwissenschaften zu finden sind, die durch internationale Schulleistungsvergleiche, allen voran das *Programme of International Student Assessment* (PISA) der OECD, untersucht werden. Hier vor allem wird die „Produktivität“ von Bildungssystemen und ihren

Komponenten öffentlich und namentlich bildungspolitisch thematisiert (vgl. Husén / Postlethwaite 1967). Es ist vielleicht nicht überflüssig zu fragen, warum die Bestandsaufnahme gerade in diesen drei Kompetenzbereichen auf Dauer gestellt worden ist.

Bereits in der allerersten Studie der IEA („*International Association for the Evaluation of Educational Achievement*“) von 1959 bis 1962, der so genannten „*Pilot Twelve-Country Study*“ oder „*Feasibility Study*“, in der 13-jährige Schülerinnen und Schüler aus zwölf Ländern untersucht wurden, ging es um Mathematik, Leseverständnis, Geographie, Naturwissenschaft und nonverbale kognitive Fähigkeiten im internationalen Vergleich (Foshay et al. 1962). Man mag es als Zugeständnis an die vergleichsweise gute Beherrschbarkeit testmethodischer Probleme in diesen Domänen betrachten, dass hier der Anfang gemacht wurde. Dass damit aber – ohne den bestimmten Artikel – Kernaufgaben der Allgemeinbildung angesprochen sind, wird – ungeachtet mancher *zusätzlicher* Desiderate – im Grunde bis heute nicht bestritten.

Namentlich durch die IEA ist im Laufe der Zeit das ursprünglich begrenzte Programm international vergleichender Bildungsforschung sehr erheblich erweitert worden. Bereits das nächste IEA-Großprojekt – die *First International Mathematics Study* von 1964 (FIMS: Husén / Postlethwaite 1967) bezog auch die Oberstufenmathematik in die Untersuchung ein, und die so genannte *Six Subjects Study* von 1970/71 (Walker 1976) berücksichtigte zusätzlich Grundschul Kinder und die neuen Domänen Französisch als Fremdsprache, Englisch als Fremdsprache und Politische Bildung. Seither ist die Untersuchung dieser Bereiche in mehrere Zyklen eingebunden, zum Teil mit relativ hoher Frequenz (*Trends in International Mathematics and Science Study*: TIMSS, vierjährig; *Progress in Reading Literacy Study*: PIRLS, fünfjährig). PIRLS ist übrigens in Deutschland als „Internationale Grundschul-Lese-Untersuchung“ unter dem Akronym *IGLU* publiziert worden (vgl. Bos et al. 2003; Bos et al. 2007). Teilweise sind die Zyklen aber auch eher langfristig angelegt, so z.B. der Bereich Politische Bildung (*Civic Education*: zuletzt Amadeo et al. 2002; Torney-Purta et al. 2001; Händle, Oesterreich, & Trommer, 1999) mit Zeiten zwischen den Replikationen von zehn Jahren und mehr. Zu erwähnen ist auch die Befassung der IEA mit Kompetenzen im IT-Bereich (Law, Pelgrum / Plomp 2008) und neuerdings die Erweiterung des Blickwinkels über Aspekte der Schulleistungen hinaus etwa auf die Genese professioneller Kompetenz von Lehrkräften in der Ausbildung; vgl. die

Teacher Education and Development Study in Mathematics TEDS-M (Tatto et al. 2008) mit der Vorläuferstudie *Mathematics Teaching for the 21st Century* MT 21 (Blömeke / Kaiser / Lehmann 2008). Als ein Glanzstück international vergleichender, empirischer Bildungsforschung darf das *IEA Pre-Primary Project* (1986-2003; vgl. Montie / Xiang / Schweinhart, 2007) gelten, einer über fast zwanzig Jahre betreuten Längsschnitt-Untersuchung, in der es gelungen ist, die hohen Renditen von Investitionen in die frühkindliche Erziehung nachzuweisen, namentlich in problematischen sozialen Kontexten.

Unbedingt zu erwähnen sind thematische Ausweitungen, die zielgruppenbezogen und auch methodisch den von der IEA erschlossenen Erfahrungsraum überschritten haben. Beispielhaft hierfür sei der von *Statistics Canada und Educational Testing Service* initiierte Forschungsstrang *International Adult Literacy Survey* (IALS: OECD & Statistics Canada 2000) bzw. *Adult Literacy and Life Skills Survey* (ALL: OECD 2006) genannt, innerhalb dessen die Grundqualifikationen Erwachsener gemessen und zur individuellen ökonomischen und sozialen Lage in Beziehung gesetzt worden sind. Letzteres ist insoweit von besonderer Bedeutung, als ja gar nicht sicher ist, in welchem Verhältnis die von den großen Vergleichsstudien erhobenen *schulischen* Kompetenzen zu den nachschulischen Lebensverläufen stehen. Und weiter: Es gibt sehr ernst zu nehmende Anzeichen dafür, dass dieses Verhältnis zwischen verschiedenen Gesellschaften erheblich variieren kann. Auch kann nicht ausgeschlossen werden, dass sich das Grundqualifikationsniveau in Deutschland zumindest relativ zu den Vergleichsländern, vielleicht aber auch absolut, verringert hat (vgl. Lehmann 2000).

Damit ist bereits angedeutet, wie durch analytische Ausdifferenzierung der Studien nicht nur neue Fragestellungen entstanden, sondern die Paradigmata des internationalen Vergleichs selbst modifiziert worden sind.

In der Umschreibung des ursprünglichen Konzepts der „*world as a single educational laboratory*“ ist oben nicht zufällig die Übersetzung „Experimentierfeld“ gewählt worden. Bekannt und berüchtigt ist die Abqualifizierung quantitativ-empirischer Forschung in der Erziehungswissenschaft als dem „*agricultural-botany paradigm*“, dem „Saatzuchtparadigma“, verpflichtet (Parlett / Hamilton 1977). In der Tat beruhten die frühen internationalen Vergleichsstudien wie andere großflächig angelegte Untersu-

chungen auf der vergleichsweise schlichten Logik eines Mehr oder Weniger, eines auf der Ebene der Bildungssysteme saldierten Ertrages: des „*yield*“ (auch dies ein Begriff, der in agro-ökonomischen Untersuchungen heimisch ist). So wurde schon bei der Konzipierung der ersten internationalen Mathematikstudie (FIMS) von 1964 die Produktivitätsfrage nicht zuletzt als curriculares Problem definiert, nämlich als Frage nach dem Nutzen der ‚Neuen Mathematik‘, für die eine Antwort aus den Ländervergleichen gesucht wurde. Der der Studie zugeschriebene praktische Stellenwert wurde durch die Forschungsförderung der Schwedischen Reichsbank unterstrichen.

Spätestens für die Sechs-Fächer-Studie von 1971/72 war es dann in verschiedenen teilnehmenden Ländern von erheblicher Bedeutung, ob sich die neu eingerichteten Gesamtschulen bzw. Gesamtschulsysteme nach ihrem bildungsmäßigen „Ertrag“ mit den etablierten traditionellen Schulsystemen würdigen messen können. Es fehlte freilich auch nicht an frühen Äußerungen von Bildungspolitikern, die an ihrer politischen Entscheidung zugunsten eines Gesamtschulsystems ganz unabhängig von den Vergleichsergebnissen festgehalten haben oder hätten.

Der Inbegriff dieser einfachen – aber, wie noch zu zeigen ist, äußerst problematischen – Vergleichsform ist die ‚Liga-Tabelle‘, die als solche keine Schlüsse auf *Gründe* für beobachtete Differenzen erlaubt. Selbst in der inzwischen üblich gewordenen Form des Vergleichs von Perzentilbändern, bei dem dann auch die relative Positionen einzelner Leistungssegmente sowie die Leistungsstreuung insgesamt berücksichtigt werden können, bleiben die Hintergründe beobachteter Leistungsdifferenzen ungeklärt. Nicht zuletzt hier liegt einer der Gründe dafür, dass diese Tabellen oft selektiv, manchmal sogar wahrheitswidrig und jedenfalls häufig in Abhängigkeit von vorgefassten ‚Überzeugungen‘ gelesen und missbraucht werden.

An der Stelle solcher Bindung an *a priori* und darum oft auch blind getroffene Entscheidungen müssen aber – soll international vergleichende Bildungsforschung auch bildungspolitisch ernst genommen werden – begründete Interpretationen der Befunde treten, die den zunächst zu eruiierenden systemimmanenten analytischen Zusammenhängen Rechnung tragen. Dies gilt für alle drei Ebenen der Nutzung internationaler Studien für die Zwecke bildungspolitischer Steuerung (vgl. Prenzel / Baumert / Klieme 2008):

1. Internationale Vergleichsstudien ermöglichen eine empirisch gesättigte Bestandsaufnahme nicht nur im Sinne einer relativen Positionsbestimmung unter den vergleichbaren Bildungssystemen hinsichtlich der erreichten mittleren Lernstände, sondern als ‚absolute‘ Beschreibung der vorhandenen Kompetenzen und ihrer Verteilung in der Zielgruppe.
2. Unter der Voraussetzung, dass Vergleichbarkeit angenommen werden kann, gelingt eine evidenzgestützte Bestimmung dessen, was im eigenen Bildungssystem prinzipiell, mit einer geeigneten Strategie, erreicht werden könnte.
3. In den rekurrenten, fest institutionalisierten Vergleichsuntersuchungen wie PIRLS/IGLU, TIMSS und PISA können mittel- und längerfristige Trends im Bildungssystem herausgearbeitet werden, wird also zu allererst die Rekonstruktion der Dynamik des Bildungssystems möglich, von der oben die Rede war.

Der Übergang von deskriptiven zu theoretischen Darstellungsformen ist nicht scharf zu ziehen, setzt doch jede Deskription mit einiger Tiefenschärfe bereits die Definition von Teilgruppen voraus, die im Regelfall ohne vorgängige theoretische Verständigung nicht geleistet werden kann. Reduzieren sich in den Liga-Tabellen die an einer Studie teilnehmenden Länder auf je einen Datenpunkt oder allenfalls einen kurzen Vektor von Kennwerten, so werden spätestens hier die Relationen zwischen implizit oder explizit definierten Teilgruppen thematisiert. Es folgen einige Beispiele:

1. Fragt man nach der Rendite von Bildungsinvestitionen (als dem ökonomischen Inbegriff der Produktivität des Systems), so ist letztlich die Berücksichtigung unterschiedlicher Qualifikationsniveaus und damit eine theoretisch fundierte Definition von Kompetenzstufen unabdingbar. Dasselbe gilt übrigens auch in dem Falle, dass man den ökonomistischen Jargon vermeidet und – wie PISA etwa – von politischen, sozialen, kulturellen und wirtschaftlichen „Partizipationschancen“ spricht. Insoweit gewinnt insbesondere die davon abhängige Ermittlung des Umfangs von Gruppen, die durch ein so bestimmtes Qualifikationsniveau charakterisiert sind, unmittelbare politische Steuerungsrelevanz.

2. Jede Auseinandersetzung mit den Hintergründen der beobachteten interindividuellen (Kompetenz-) Differenzen setzt ihrerseits eine klare, wiederum theoretisch fundierte Bestimmung von sozialen Schichten, Lagen oder Klassen voraus, denen die untersuchten Kinder und Jugendlichen entstammen. Namentlich im Umfeld der PISA-Studien sind hier – etwa im Ringen um einen auch international geeigneten Indikator für den sozialen Hintergrund der untersuchten Jugendlichen – große Fortschritte gegenüber den eher pragmatischen Vorgaben früherer Studien gemacht worden.
3. Die Unterscheidung individueller Laufbahnen – „Trajektorien“ – durch das Bildungssystem ist ohne fundiertes Hintergrundwissen über die Angebotsstrukturen im eigenen Land und ohne Kenntnis struktureller Äquivalenzen in den Vergleichssystemen nicht möglich; diese kann nicht aus singulären Beobachtungen abgeleitet werden, sondern beinhaltet notwendig hypothetische Elemente. Erst recht kommt eine aussagekräftige Analyse individueller und gruppenspezifischer Einkommensprofile über die Zeit („age-earnings profiles“) nicht ohne ein erhebliches Maß theoretischer Annahmen aus, schon gar nicht, wenn es um die Erarbeitung von Prognosen zu (noch) nicht beobachteten Zeiträumen geht.
4. Die Diskussion von Fragestellungen der Gender-Forschung, auch und gerade, wenn neue Befunde gängigen oder zumindest lange und kreativ gepflegten Stereotypen widersprechen, verweist auf die Formulierung theoretischer Alternativen, die sich den in der einschlägigen Literatur vertretenen, auf Plausibilität sich stützenden Deutungen als überlegen zeigen.

Diese Liste lässt sich nahezu beliebig verlängern. Sie repräsentiert einen Katalog von Fragestellungen, unter denen die riesigen Dateien der großen internationalen Studien bearbeitet werden können und langfristig auch analysiert werden müssen. Aus dem scheinbar einfach zu überschauenden Ensemble von zwölf Vergleichsländern in der Machbarkeitsstudie der IEA von 1959 ist 50 Jahre später ein Forschungsprogramm geworden, das nicht nur Arbeitsmöglichkeiten für viele Hundert beteiligte Wissenschaftler bietet, sondern auch ein umfangreiches ‚Pflichtenheft‘ zu berücksichtigender Fragestellungen für die bildungspolitischen Entscheidungsträger.

2. ZUR BEDEUTUNG DER INTERNATIONALEN BILDUNGSFORSCHUNG FÜR DEN VOLLZUG DER WENDE ZU EINER OUTPUT-ORIENTIERTEN, EVIDENZBASIERTEN BILDUNGSPOLITIK

Neben den eben genannten Themen empirisch zu bearbeitender Fragen sind auch die im Kontext international vergleichender Studien realisierten und den teilnehmenden Ländern erschlossenen neuen methodischen Ansätze von besonderer Bedeutung für die Entstehung einer output-orientierten, evidenzbasierten Bildungspolitik. „*Capacity building*“ als Schlüsselement internationaler Kooperation ist eine auch und gerade im Kontext der Bildungsforschung und Bildungspolitik zu konstatierende Notwendigkeit, die zumal in Deutschland schon seit Jahrzehnten bestanden hat und der im Übrigen in den letzten Jahren das Bundesministerium für Bildung und Forschung, die zuständigen Landesministerien und nicht zuletzt die Deutsche Forschungsgemeinschaft durch beträchtliche Unterstützungsleistungen Rechnung getragen haben. Ich nenne einige Beispiele für solchen forschungsmethodischen Technologie-Transfer.

2.1. Stichprobeneffekte

Unter den gegenwärtig noch aktiven und einflussreichen Bildungsforschern haben nicht wenige ihre methodische Expertise auf dem Felde der experimentellen Psychologie mit den dort üblichen zufallsgesteuerten Experimental-Kontrollgruppen-Designs erworben. Die statistische Signifikanz beobachteter Effekte konnte dort nach den Vorgaben einfacher Zufallsstichproben überprüft werden. Nun sind aber die zu behandelnden Zielgruppen nicht nach diesem Schema strukturiert: Schülerinnen und Schüler lernen nicht in zufällig entstandenen Lerngruppen nach gemeinsamem Lernarrangement; vielmehr besitzen die verwendeten Stichproben intakter Schulklassen, zu denen es schon aus pragmatischen Gründen keine Alternative gibt, intern eine hierarchische Struktur mit mehreren Aggregationsebenen (Regionen, Schulformen, Schulen, Klassen). Konsequenz ist, dass die *effektive Stichprobe* hierzulande etwa um den Faktor 5 kleiner ist als die Anzahl der getesteten Schüler.

Die Theorie, die solche Stichprobenstruktur berücksichtigt und damit die Erhebungskosten für *lege artis* angelegte Untersuchungen nicht unerheblich in die Höhe getrieben hat, ist in den sechziger Jahren des vorigen Jahrhunderts von Leslie Kish (1965) entwickelt worden. Noch um 1990

fiel es prominenten Fachvertretern in Deutschland schwer, diese Implikation zu akzeptieren: Man sei, so eine informelle Reaktion auf die Berücksichtigung der Theoreme von Kish, doch nicht so töricht, sich durch solchen Purismus die Signifikanz der eigenen Befunde zu zerstören. Heute dagegen, unter dem Zwang, mit international akzeptablen Stichproben zu arbeiten, ist es selbstverständlich, den von Kish so genannten „Design-Effekt“ in die Planung einzubeziehen. Dabei ist es nicht uninteressant zu sehen, dass hierzulande inzwischen die für die Stichprobenziehung und die entsprechende Berechnung der Stichprobenfehler eingesetzten Routinen sowohl von den für die internationalen Vergleiche federführenden Instituten als auch von den Qualitätsagenturen auf Bundes- und Länderebene fast ausnahmslos mit dem *Data Processing Centre* (DPC) der IEA an eine Institution vergeben werden, deren Vorgehensweisen ursprünglich – wie angedeutet – als eher befremdlich gegenüber den im Inland etablierten Forschungsroutinen gehalten wurde.

2.2. Kontexteffekte: Mehrebenen-Analysen

Mit der hierarchischen Struktur von Bildungssystemen, die in Regionen, Schulformen, Schulen und Schulklassen gegliedert sind, hängt die Beobachtung zusammen, dass gleiche individuelle Anfangsausprägungen lernrelevanter Merkmale u.U. mit beträchtlich unterschiedlichen Lernerefolgen einhergehen, wofür Kontexteffekte verantwortlich sein können und oft auch sind. Auch führen mangelhafte Unterscheidungen zwischen den Effekten der verschiedenen Aggregatebenen zu Fehlschlüssen.

Erste Ansätze, hiermit methodisch sauber umzugehen, gehen in Deutschland auf Ansätze von Treiber und Weinert (1985) zurück. Der Durchbruch, nach dem die dann in so genannten „hierarchischen linearen Modellen“ realisierten Mehrebenen-Analysen zum Bestandteil normaler Forschung geworden sind, gelang aber erst mit der Übernahme der Algorithmen in leicht zugänglichen Computepogrammen (z.B. HLM 6: Raudenbush et al. 2004; vgl. Raudenbush / Bryk 1992), die inzwischen international als gleichsam obligatorisch durchzuführen galten.

Mehrebenen-Analysen erlauben es beispielsweise, sogenannte „Kompositionseffekte“ statistisch sichtbar zu machen, denen zufolge lernstarke Schülerinnen und Schüler in entsprechend zusammengesetzten Lerngruppen mehr und schneller lernen als in ausgeprägt heterogenen Kontexten (Baumert / Köller / Schnabel 2000; Köller / Baumert 2008, 750).

Auch und gerade in Anbetracht des Umstands, dass diese Frage derzeit bildungspolitisch äußerst kontrovers diskutiert wird, und zumal dann, wenn dann unter Umständen konkurrierende analytische Zugriffsweisen ins Spiel gebracht werden (Lehmann / Lenkeit 2008; Baumert et al. 2009), ist es Pflicht, entsprechende Entscheidungen in Auseinandersetzung mit einschlägigen Erkenntnissen zu treffen. An Beispielen für einen unterhalb solcher Differenzierungen ansetzenden, selektiven Umgang mit partiellen Befunden auf der Suche nach bildungspolitischer Munition mangelt es nicht (Prenzel / Baumert / Klieme 2008).

Überhaupt ist die Vorstellung, dass Merkmalszusammenhänge – z.B. zwischen Merkmalen der sozialen Herkunft und Lernerfolg – quasi gesetzesartigen Charakter tragen und in allen Kontexten gleich geartet sind, mehrbenenanalytisch zu prüfen und gegebenenfalls aufzugeben. Hier mag es lehrreich sein, sich zu vergegenwärtigen, dass der oft thematisierte „Sozialgradient“ als Indikator für eben jenen Zusammenhang in Deutschland nicht in allen Bundesländern so steil ist wie regelmäßig beklagt und dass er selbst innerhalb eines Bundeslandes beträchtlich zwischen den Schulformen variieren kann.

2.3. Kompetenzmodelle

Der öffentliche Diskurs über die Erträge der Bildungsbemühungen in diesem Lande ist gesättigt mit der Verwendung des Begriffs „Kompetenz“, der übrigens ebenfalls einen Import aus der internationalen empirischen Bildungsforschung darstellt. Er verdankt seinen Stellenwert einem Gutachten, das Franz Weinert im Jahr 2000 für das Projekt „*Defining and Selecting Key Competencies*“ (DeSeCo; Rychen / Salganik 2001) erarbeitet hat und das – vermittelt über das deutsche „Forum Bildung“ – nachfolgend maßgeblich war für die so genannte „Klieme-Expertise“ zur Entwicklung nationaler Bildungsstandards (Klieme et al. 2003).

Auch hier sind die methodischen Voraussetzungen und deren Übernahme aus dem Kontext der internationalen Bildungsforschung zu betonen. Zwar ist der für die *Kompetenzmessung* inzwischen maßgebliche probabilistische Ansatz zur Testtheorie, der 1960 von dem dänischen Mathematiker und Psychologen Georg Rasch entwickelt worden ist und inzwischen unter dem Etikett *Item Response Theory* (IRT) zum methodischen Kernbestand der Zunft gehört, in Deutschland vor allem am Institut für die Pädagogik der Naturwissenschaften (IPN) bzw. an der Universität Kiel

schon relativ früh zur Kenntnis genommen und weiter entwickelt worden (vgl. die Arbeiten von Jürgen Rost und Hans Spada), doch zum Durchbruch ist dieser Ansatz hierzulande erst durch die internationalen Vergleichsstudien gelangt, die ausreichend umfangreiche und komplex strukturierte Stichproben sowie gut nachvollziehbare Präsentationsformen für die Ergebnisse bereitgestellt haben. Zu nennen ist als erste auch Deutschland betreffende Realisierung die Internationale Lesestudie der IEA von 1991 (Lehmann et al. 1995), dann aber vor allem die bereits erwähnte *Third International Mathematics and Science Study* (TIMSS), deren erste Ergebnisse 1996 publiziert wurden und damit erstmals ein bildungspolitisches Erdbeben – den „TIMSS-Schock“ – ausgelöst haben (vgl. Baumert et al. 1997). Die unmittelbar damit zusammenhängenden „Konstanzer Beschlüsse“ der Kultusministerkonferenz (KMK) vom Oktober 1997 gingen hierauf zurück.

Es können hier nicht die nachfolgenden Bemühungen um die weitere Ausarbeitung für viele unterschiedliche Zielgruppen, viele Domänen oder Fächer, auch unterschiedliche Schulformen, geschildert werden. Das *proficiency scaling* als ein Ansatz, IRT-Skalen zu einem Instrument der von Pädagogen schon lange geforderten – allerdings manchmal schon wieder problematisierten – kriteriumsorientierten Leistungsmessung zu machen, ist längst in allen Anwendungsbereichen obligatorisch. Die Möglichkeit, auf dieser Basis zeitlich und/oder nach dem Anforderungsniveau gestaffelte Testverfahren zu entwickeln und namentlich in Längsschnittstudien einzusetzen, sei hier nur erwähnt.

2.4. Bildungssoziologische und bildungsökonomische Effekte

Als abschließendes Beispiel für die Chance, die deutsche Bildungsforschung durch die Teilnahme an internationalen Studien global anschlussfähig zu machen, sei der Stellenwert der so entstandenen Datencorpora skizziert. Die Notwendigkeit, Effekte sozialer Benachteiligung etwa international vergleichsfähig darzustellen, hat es erforderlich gemacht, über die Hintergrundfragebögen standardisierte Operationalisierungen zu übernehmen. Größen wie die *International Standard Classification of Occupations* (ISCO) der International Labour Organization (ILO, 1988), die *International Standard Classification of Education* (ISCED), des *International Socio-Economic Index* (ISEI: Ganzeboom / Treiman 2003) und die soziale Klassifikation nach Erikson-Goldthorpe-Portocarero (EGP: Erikson / Goldthorpe / Portocarero 1979) sind Beispiele hierfür.

Die Bedeutung solcher Analysen ergibt sich aus dem hohen Stellenwert, den Fragen der Bildungsangebote, ihrer Zugänglichkeit in Abhängigkeit von der jeweiligen sozialen Lage und ihrer tatsächlichen Nutzung durch die internationalen Vergleiche auch unter diesem Aspekt in der öffentlichen Diskussion – wieder – erlangt haben.

3. ZUR NATIONALEN BEDEUTUNG DER EMPIRISCHEN BILDUNGSFORSCHUNG: DAS PROGRAMM DER „EMPIRISCHEN WENDE“ ZU EINER OUTPUT-ORIENTIERTEN, EVIDENZBASIERTEN BILDUNGSPOLITIK (H. LANGE)

Der ehemalige Hamburger Staatsrat Hermann Lange (1939-2008) hat die Auswirkungen der großen internationalen Vergleichsstudien auf das deutsche Bildungswesen mit seiner Formel von der „empirischen Wende der Bildungspolitik“ wohl am prägnantesten bezeichnet. Die mit den Konstanzer Beschlüssen der KMK von 1997 eingeleitete Einbeziehung der einzelnen Bundesländer in die Vergleiche mit den dann festzustellenden erheblichen Differenzen birgt potentiell erheblich mehr politische Brisanz noch als der Bezug der Bundesrepublik insgesamt zu Bezugswerten der OECD, der Europäischen Union oder auch einzelner europäischer und außereuropäischer Systeme. So war es nur konsequent, dass die Bundesländer inzwischen dazu übergegangen sind, eigene Bildungsforschungsinstitute („Qualitätsagenturen“) zu gründen, entsprechende Studien in Auftrag zu geben und regelmäßig Bildungsberichte herauszugeben. Mit beträchtlichem technischen Aufwand wird also zumindest auf dieser Ebene der alten Forderung nach „Rechenschaftslegung“ begegnet, dem englischen, der Verbraucherschutzbewegung verpflichteten Begriff der „*accountability*“ entsprechend. Es verbindet sich damit die Erwartung, bildungspolitische Entscheidungen rationaler als bisher, nämlich unter Berücksichtigung erwartbarer Konsequenzen, zu treffen.

3.1. Anstehende bildungspolitische Entscheidungen: Strukturentwicklung

Obwohl sich die empirische Bildungsforschung lange Zeit bildungspolitischer, namentlich schulstruktureller Empfehlungen enthalten hat, spricht doch viel dafür, dass diese Dimension in der überschaubaren Zukunft an Bedeutung gewinnen wird. Vor allem demographische Entwicklungen – die Entvölkerung bestimmter Regionen ebenso wie die sich wandelnde Zusammensetzung der Bevölkerung in den Ballungszentren – lassen es

als fraglich erscheinen, ob es ratsam oder überhaupt möglich ist, ein vielgliedriges Sekundarschulsystem flächendeckend als Angebotsstruktur vorzuhalten. Die Situation wird auf der Nachfrageseite durch veränderte Elternwahlentscheidungen – bei steigender Konkurrenz um gymnasiale Bildungsgänge – noch verschärft. Zweifellos wird in diesem Zusammenhang nicht nur die Frage nach der Anzahl unterschiedlicher Sekundarschulformen, sondern auch die nach dem oder den optimalen Zeitpunkt(en) für die Übergangentscheidung empirisch fundiert zu thematisieren sein, wofür der – zumeist selektive – Hinweis auf andere Regelungen in anderen Ländern schwerlich ausreicht. Generell sind aber die hochgradig von politischen, erfahrungsresistenten Grundüberzeugungen getragenen Präferenzen dringend auf empirisch fundierte Entscheidungsgrundlagen angewiesen, wozu übrigens auch die Untersuchung und Berücksichtigung vorhandener und möglicher Optionen für die Durchlässigkeit der Institutionen für alternative Bildungswege gehört.

3.2. Weitere bildungspolitische Entscheidungen: Optimierung des Bildungssystems vor dem Hintergrund subgruppenspezifischer Probleme

Die internationalen Vergleichsstudien liefern ebenso wie nationale und regionale Erhebungen wertvolles Datenmaterial, das die besondere Lage von Subgruppen zu beschreiben und ggf. zu verbessern erlaubt. Hierher gehört z.B. die erst durch solche Studien veranlasste Revision der teilweise historisch verständlichen ‚Grundüberzeugung‘, dass weibliche Kinder und Jugendliche im (deutschen) Bildungswesen generell benachteiligt seien, oder auch die Relativierung der lediglich mit Beteiligungquoten argumentierenden These einer „institutionellen Diskriminierung“ (Gomolla / Radtke 2002; vgl. auch Karakaşoğlu-Aydin 2001; Vernor Muñoz 2007). Namentlich dieses Argument beruht auf der starken Voraussetzung, dass kognitive Ressourcen und speziell Fachleistungen in den jeweils verglichenen Gruppen gleich verteilt sind. Die Vergleichsstudien bieten nun, ebenso wie ausreichend differenziert instrumentierte nationale und regionale Erhebungen, die Möglichkeit, eben solche Voraussetzungen zu prüfen. Die auf solcher Basis konzipierten Interventionen unterscheiden sich dann notwendig danach, ob die Prämissen empirisch belastbar sind.

Stets muss es darum gehen, jeder Einzelgruppe – Kindern und Jugendlichen aus zugewanderten Familien, u.U. auch differenziert nach Herkunftsländern und Migrantenstatus, solche mit sonderpädagogischem Förderbedarf, Risikogruppen mit extrem schwachen Fachleistungen und entsprechenden Ausbildungs- und Beschäftigungschancen – Zugang zu einem Maximum substanzieller Kompetenz zu eröffnen. Was in dieser Hinsicht als möglich und aussichtsreich gelten kann, erschließt sich zumindest *auch*; häufig aber *zu allererst* durch empirisch gesättigte Einsichten in das Bedingungsgeflecht, dem sich hohe Lernerfolge verdanken.

4. FAZIT

Sollte bis hierher der Eindruck entstanden sein, die benötigten Erkenntnisse seien nur auf dem Wege des Imports international verwendeter, als bewährt betrachteter Paradigmata und Methoden zu gewinnen, so würden damit wesentliche Beiträge gerade auch deutscher Forschungsgruppen zum ‚*state of the art*‘ der empirischen Bildungsforschung übergegangen. In der Entfaltung des PISA-Programms vom ersten Durchgang im Jahre 2000 über die Erhebungen von 2003, 2006 und 2009 hinweg lassen sich Rückwirkungen des hierzulande belebten Interesses an solchen Untersuchungen nachweisen, etwa in der Aufnahme von Verfahren zur Erfassung von Problemlösekompetenzen der Jugendlichen oder in der Betonung der Bedeutung und Entwicklung metakognitiver Fähigkeiten, die inzwischen, auf entsprechende Initiative und unter Verweis auf deutsche Zusatzuntersuchungen im ersten OISA-Durchgang hin, in das internationale Testprogramm aufgenommen worden sind. In den Erträgen dieser multilateralen Forschungs- und Entwicklungslinie wird man nicht die geringste Bedeutung der international-kooperativen Vergleichsstudien und ihrer nationalen Ergänzungen sehen.

Hiernach ergeben sich die Bedeutung und der Stellenwert der empirischen Bildungsforschung letztlich aus ihrer *methodisch verankerten, inhaltlichen, Überzeugungskraft* – aus der kooperativ erzielten (um eine zunehmend beliebte Formulierung zu wählen) „Belastbarkeit“ ihrer Befunde. Diese freilich ist, wie wir seit Karl Popper wissen, immer nur und allenfalls vorläufig gegeben. Insoweit mahnt die Betonung der methodischen Grenzen auch zu gegenseitigem Respekt und Vorsicht in bildungspolitischem Diskurs und innovatorischer Praxis.

LITERATUR

- Amadeo, J.-A. / Torney-Purta, J. / Lehmann, R. / Husfeldt, V. / Nikolova, R. (2002): *Civic Knowledge and Engagement. An IEA Study of Upper Secondary Students in Sixteen Countries. Amsterdam (The International Association for the Evaluation of Educational Achievement).*
- Baumert, J. / Becker, M. / Neumann, M. / Nikolova, R. (2009): *Frühübergang in ein grundständiges Gymnasium: Übergang in ein privilegiertes Entwicklungsmilieu? Ein Vergleich von Regressionsanalyse und Propensity Score Matching. In: Zeitschrift für Erziehungswissenschaft 12, 189-215.*
- Baumert, J. / Köller, O. / Schnabel, K. (2000): *Schulformen als differentielle Entwicklungsmilieus – eine ungehörige Fragestellung? In: Schriftenreihe des Bildungs- und Förderungswerks der GEW im DGB, Heft 14.*
- Baumert, J. / Lehmann, R. H. / Lehrke, M. / Schmitz, B. / Clausen, M. / Hosenfeld, I. / Köller, O. / Neubrand, J. (1997): *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen.*
- Blömeke, S. / Kaiser, G. / Lehmann, R. (Hrsg.) (2008): *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-Studierender und -Referendare – erste Ergebnisse zur Wirksamkeit der Lehrerbildung. Münster.*
- Bos, W. / Hornberg, S. / Arnold, K.-H. / Faust, G. / Friede, L. / Lankes, E.-M. / Schwippert, K. / Valtin, R. (2007): *IGLU 2006. Lesekompetenzen von Grundschulern im internationalen Vergleich. Münster.*
- Bos, W. / Lankes, E.-M. / Prenzel, M. / Schwippert, K. / Walther, G. / Valtin, R. (2003): *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Münster.*

- Erikson, R. / Goldthorpe, J. H. / Portocarero, L. (1979): *Intergenerational class mobility in three Western European societies: England, France and Sweden*. In: *British Journal of Sociology* 30, 341-415.
- Foshay, A. W. / Thorndike, R. L. / Hotyat, F. / Pidgeon, D. A. / Walker, D. A. (1962): *Educational Achievements of Thirteen-year-olds in Twelve Countries. Results of an international research project 1959-61. Hamburg (Unesco Institute for Education: International Studies in Education)*.
- Ganzeboom, H. B.J. / Treiman, D. J. (2003): *Three internationally standardised measures for comparative research on occupational status*. In: Hoffmeyer-Zlotnik, J. / Wolf, Chr. (Hrsg.): *Advances in cross-cultural comparison*. New York, 159-193.
- Gomolla, M. / Radtke, F.-O. (2002): *Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule*. Opladen.
- Händle, C. / Oesterreich, D. / Trommer, L. (1999): *Aufgaben Politischer Bildung in der Sekundarstufe 1. Studien aus dem Projekt Civic Education*. Opladen.
- Heyneman, S. P. (2003): *Comment*. In: *Brookings Papers on Educational Policy*, 332-335.
- Husén, T. / Postlethwaite, T. N. (1967): *Chapter 1: Intentions and Background of the Project*. In: Husén, T. (Hrsg.): *International Study of Achievement in Mathematics. A Comparison of Twelve Countries*. Stockholm / New York, Vol. I, 25-34.
- Ingenkamp, K. (1989): *Die Test-Aversion des deutschen Intellektuellen. Eine Streitschrift*. Weinheim.
- International Labour Organization (ILO) (Hrsg.) (1988): *International standard classification of occupations - (ISCO-88)*. Geneva (ILO).
- Karakaşoğlu-Aydin, Y. (2001): *Kinder aus Zuwandererfamilien im Bildungssystem*. In: Böttcher, W. / Klemm, K. / Rauschenbach, T. (Hrsg.): *Bildung und Soziales in Zahlen. Statistisches Handbuch zu Daten und Trends im Bildungsbereich*. Weinheim und München, 273-302.

- Kish, L. (1965): *Survey Sampling*. New York.
- Klieme, E. / Avenarius, H. / Blum, W. / Döbrich, P. / Gruber, H. / Prenzel, M. / Reiss, K. / Riquarts, K. / Rost, J. / Tenorth, H.-E. / Vollmer, H. (2003): *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn (BMBF: Bildungsreform 1).
- Köller, O. / Baumert, J. (2008): *Entwicklung schulischer Leistungen*. In: Oerter, R. / Montada, L. (Hrsg.): *Entwicklungspsychologie*. 6. vollst. überarb. Auflage Weinheim, 735-768.
- Lange, H. (1999): *Qualitätssicherung in Schulen*. In: *Die Deutsche Schule* 91, 144-159.
- Law, N. / Pelgrum, W. J. / Plomp, T. (Hrsg.) (2008): *Pedagogy and ICT use in schools around the world: Findings from the IEA SITES 2006 study*. Hong Kong.
- Lehmann, R. H. (2000): *Anregungen für bildungspolitische Konsequenzen der empirischen Forschung*. In: Stark, W. / Fitzner, T. / Schubert, C. (Hrsg.): *Von der Alphabetisierung zur Leseförderung*. Stuttgart, 153-165.
- Lehmann, R. H. / Lenkeit, J. (2008): *ELEMENT. Erhebung zum Leseverständnis und Mathematikverständnis. Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin. Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien*. Berlin (Senatsverwaltung für Bildung, Jugend und Sport: http://www.berlin.de/imperia/md/content/sen-bildung/schulqualitaet/element6_bericht_komplett.pdf).
- Lehmann, R. H. / Peek, R. / Pieper, I. / von Stritzky, R. (1995): *Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen*. Weinheim und Basel.
- Maus, H. / Fürstenberg, F. (Hrsg.) (1972): *Der Positivismusstreit in der deutschen Soziologie*. 2. Auflage. Neuwied / Berlin.
- Montie, J. E. / Xiang, Z. / Schweinhart, L. J. (Hrsg.) (2007): *Role of preschool experience in children's development in 10 countries*. Ypsilanti, MI.

- Muñoz Villalobos, V. (2007): *Implementation of General Assembly Resolution 60/251 of 15 March 2006 entitled „Human Rights Council“. Report of the Special Rapporteur on the right to education. Addendum „Mission to Germany“. New York, NY (United Nations, General Assembly: A/HRC/4/29/Add.3).*
- Organisation for Economic Co-Operation and Development (OECD) (Hrsg.)(2006): *Learning a Living: First Results of the Adult Literacy and Life Skills Survey. Paris.*
- Organisation for Economic Co-Operation and Development (OECD) & Statistics Canada (Hrsg.)(2000): *Literacy in the Information Age. Final Report of the International Adult Literacy Survey. Paris/Ottawa.*
- Parlett, M. / Hamilton, D. (1977): 'Evaluation as Illumination'. In: D. Hamilton / D. Jenkin / C. King / B. MacDonald / M. Parlett (Hrsg.): *Beyond the Numbers Game. London, 6-22.*
- Prenzel, M. / Baumert, J. / Klieme, E. (2008): *Steuerungswissen, Erkenntnisse und Wahlkampfmunition: Was liefert die empirische Bildungsforschung? – Eine Antwort auf Klaus Klemm. Volltext: <http://de.scientificcommons.org/48399817>, geladen am 21.10.2009.*
- Raudenbush, S. W. / Bryk, A. S. / Cheong, Y. F. / Congdon, R. T. (2004): *HLM 6: Hierarchical Linear and Nonli-near Modeling. Chicago, IL.*
- Rychen, D. S. / Salganik, L. H. (Hrsg.) (2001); *Defining and Selecting Key Competencies. Göttingen/Bern.*
- Tatto, M. T. / Schwille, J. / Senk, S. / Ingvarson, L. / Peck, R. / Rowley, G. (2008): *Teacher Education and Development Study in Mathematics (TEDS-M): Conceptual framework. East Lansing, MI (Teacher Education and Development International Study Center, College of Education, Michigan State University).*
- Torney-Purta, J. / Lehmann, R. / Oswald, H. / Schulz, W. (2001): *Citizenship and Education in Twenty-eight Countries. Civic Knowledge and Engagement at Age Fourteen. Amsterdam (The International Association for the Evaluation of Educational Achievement).*

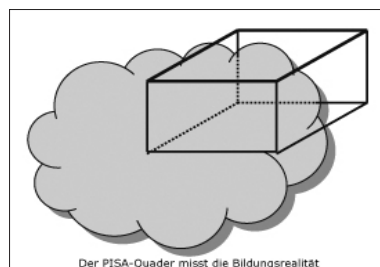
- Treiber, B. / Weinert, F. E. (1985): *Gute Schulleistungen für alle? Psychologische Studien zu einer pädagogischen Hoffnung. Münster.*
- UNESCO Institute of Statistics (ed.; 1997): *International Standard Classification of Education. Montreal, Quebec.*
- Walker, D. A. (1976): *The IEA Six Subject Survey: An Empirical Study of Education in Twenty-One Countries. Stockholm/New York (International Studies in Education IX).*
- Weinert, F. E. (2001): *Concept of Competence: A Conceptual Clarification. In: Rychen, D.S. / Salganik, L. H. (Hrsg.) (2001): Defining and Selecting Key Competencies. Göttingen / Bern, 45-65.*

PROBLEMATIK DER MESS- INSTRUMENTE AM BEISPIEL JÜNGERER SCHULSTUDIEN

Peter Bender

„Just another opinion“. Damit tun die Vertreterinnen und Vertreter der quantitativen Bildungsforschung gerne Ergebnisse ab, die nicht mit ihren Methoden erzielt werden. Stattdessen machen sie sich z.B. mit PISA anheischig zu demonstrieren, wie man die Bildung der 15-Jährigen „misst“. Mir kommt dieses Unternehmen allerdings so vor, wie wenn ein Erdölfeld erschlossen werden soll, die Ingenieurinnen und Ingenieure sich ein quaderförmiges Modell davon machen und ihre Aufgabe darin sehen, die Kanten dieses Quaders zu bestimmen. Am Schluss haben sie zwar nicht das Ölfeld, aber dafür einen schönen Quader nach ihren Vorstellungen genau vermessen (siehe Abbildung Seite 42).

Viele im Bildungsbereich mit quantitativen Methoden erzielte Erkenntnisse haben, im Gegensatz zum Aufwand für ihre Gewinnung, zur Überzeugung ihrer Protagonistinnen und Protagonisten und zu den wegen der Drittmittel-Förderung erlangten höheren Weihen im Wissenschaftsbetrieb, keine große Aussagekraft. Für diese Behauptung kann ich aus meinem Bereich der Mathematikdidaktik aus den letzten Jahrzehnten zig Belege liefern.



Viele dieser Arbeiten sind mit methodischen Fehlern gespickt. Vor allem fehlt es immer wieder an der Repräsentativität der Stichproben; wichtige Einflussgrößen werden bei Durchführung und Interpretation außer Acht gelassen; die Unschärfe von Antworten auf „weiche“ Fragen wird ignoriert;

es steht immer wieder in Zweifel, ob die Forschungsfragen und die veröffentlichten Antworten einerseits sowie das Untersuchungsdesign andererseits sich wirklich entsprechen (Validität); u.v.a.m. – Ein anderer großer Mängelbereich tut sich bei der Interpretation der Ergebnisse i.w.S. auf, zu der auch schon die Auswahl der Literatur sowie das Verständnis von deren jeweiligen Aussagen gehört. – Nicht immer merken die Autorinnen und Autoren, wie sehr sie dabei Vorurteilen unterliegen, besonders wenn sie ihre Zahlenwerte, die ja oft das Ergebnis von weitgehenden Annahmen, stark vergrößernden Schätzungen und Wahrscheinlichkeitsbetrachtungen sind, auf fünf wesentliche Stellen angeben, als ob sie eine naturwissenschaftliche Messung durchgeführt hätten. Oft genug bedienen sie aber *bewusst* Ideologien, politische Ziele oder ganz utilitaristische Absichten. – Wer hier an die Objektivität von Wissenschaft glaubt, ist naiv.

Im Folgenden beziehe ich mich i.W. auf PISA, das ja uns allen wohl bekannt und in der Bildungsdebatte hochrelevant ist. In PISA werden zwar keine *primitiven* Fehler gemacht; bzw. die primitiven Fehler stammen von den manchmal naiven, oft eigennützigem Exegetinnen und Exegeten. Aber in subtilerer Form treten einige der genannten Fehlertypen sehr wohl auf, und ihre Analyse ist durchaus lehrreich.

Aus aktuellem Anlass gehe ich am Schluss noch auf eine Studie im Auftrag der Bertelsmann-Stiftung zu dem „teuren und unwirksamen“ Sitzenbleiben ein, deren tendenziöse Aussagen kürzlich durch den deutschen publizistischen Blätterwald gejagt wurden.

Zunächst möchte ich aber noch einmal ausdrücklich das Paradigma in Frage stellen, das dem ganzen Unternehmen „PISA“ zugrunde liegt, nämlich die Meinung, man könne (und solle) Bildung *messen*; hier: die Bildung des Kollektivs der 15-Jährigen eines Lands. PISA drückt sich da scheinbar bescheidener aus: man „untersucht, wie gut fünfzehnjährige Schülerinnen und Schüler auf die Anforderungen der Wissensgesellschaft

vorbereitet sind“ (Buchrücken von PISA 2007 und 2008). Für mich ist das sehr wohl die Frage nach der Bildung der 15-Jährigen, und die Antwort in Form von gemessenen und abgeleiteten Zahlenwerten aus einem ganz schmalen Bereich halte ich für unangemessen und vermessen. – Trotz meiner grundsätzlichen Bedenken will ich mich aber im Folgenden auf den Mess-Ansatz von PISA einlassen. Für viele Zitate verweise ich auf (Bender 2007).

1. DIE MATHEMATIKDIDAKTIK IN PISA

Anders als noch bei TIMSS findet bei PISA ein „Verzicht auf transnationale curriculare Validität“ statt, stattdessen führen die Tests „ein didaktisches und bildungstheoretisches Konzept mit sich, das normativ ist“, angelehnt an die NCTM-Standards aus den USA. Der Erfolg des deutschen Mathematikunterrichts wird also an einem US-amerikanischen Curriculumsentwurf gemessen.

Grundlegend ist dabei das Konstrukt der *Mathematical Literacy* (ML; „mathematische Grundbildung“): „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.“

Diese „Definition“ passt durchaus zur Tradition der deutschen bildungstheoretischen Didaktik, wie sie z.B. vom *alten* Wolfgang Klafki (1958) verkörpert wird. Sie ist so gefasst, dass der reale Mathematikunterricht, wie er über weite Strecken in Deutschland und tendenziell wohl weltweit stattfindet, nämlich konzentriert auf das Ausführen von Verfahren und weniger auf Verstehen und Anwenden, ihr nur unzureichend gerecht wird.

Die Aufgaben, die in PISA gestellt sind, entsprechen in ihrer Gesamtheit aber ebenfalls dieser Definition nicht, d.h. zu ihrer Lösung wird vielleicht die Kompetenz zum Entkleiden von eingekleideten Rechenaufgaben gebraucht, nicht aber ML. Wer viele PISA-Punkte erzielt, kann gut PISA-Aufgaben lösen, zeigt aber nicht notwendig ML (insbesondere den deutschen Jugendlichen fehlten da, zumindest in den ersten Durchgängen, auch gewisse Techniken und Strategien auf mehreren Ebenen). Diese

ML-Ferne der PISA-Aufgaben haben zahlreiche Kollegen im In- und Ausland (Bender, Braams, Gellert, Hagemeister, Kießwetter, Meyer, Meyerhöfer, Wuttke) in zahlreichen Analysen dargestellt. Bezogen auf die grundsätzliche Forschungsfrage von PISA, nämlich nach dem Vorhandensein von ML, ist der PISA-Aufgabensatz also *nicht valide*. – Hierzu ein typisches Beispiel, das Uwe Gellert aus einer OECD-Schrift von 2000 ausgegraben hat, von dem ich natürlich nicht weiß, ob es jemals in einem PISA-Test eingesetzt wurde:

Beispiel A „Terrasse“: Nick möchte die rechteckige Terrasse seines neuen Hauses pflastern. Die Terrasse ist 5,25 Meter lang und 3,00 Meter breit. Er benötigt 81 Pflastersteine pro Quadratmeter. – Berechne, wie viele Pflastersteine Nick für die ganze Terrasse braucht.

Gedacht ist an eine Lösung der Art $5,25 \times 3 \times 81 = 1275,75$, und als korrekte Antworten sollen 1275, 1275,75 und 1276 akzeptiert werden. Klassifiziert wird diese Aufgabe so:

- „Kompetenzstufe 2: Beziehungen und Zusammenhänge zum Zwecke des Problemlösens“ [wo kommt so etwas nicht vor?];
- „Fundamentale mathematische Ideen: Raum und Form“ [eigentlich geht es um Arithmetik];
- „Erfahrungsbereich: Alltag“ [na ja].

Angeblich könne einem eine solche Aufgabe in vielen Situationen des Alltags und der Arbeitswelt begegnen und passe sie gut zur Definition der ML, wofür ja die Anwendung von Mathematik in „authentischen“ Situationen wesentlich sei.

So weit meine Übersetzung aus dem Englischen. Es handelt sich um eine eingekleidete Aufgabe, bei der es nicht auf die Lösung eines Sachproblems ankommt, sondern auf das Erkennen und Ausführen der erforderlichen arithmetischen Operation (die zweifache Multiplikation). Dies wird besonders deutlich an der Zulässigkeit der Lösung 1275, die ja mit der Pflasterung der Terrasse nichts zu tun hat, sondern lediglich aus arithmetischer Sicht, aber auch da nur mit Mühe, akzeptiert werden kann.

Wenn man einmal unterstellt, dass die Pflastersteine quadratisch sind und die Seitenlänge $1/9$ m haben, dann hat man, im Sinne der vorgegebenen Aufgabenlösungen, mit dem nicht-ganzen Teil der Terrassenlänge Probleme, weil man zu dessen Auslegen einige Steine noch vierteilen müsste, und zwar in Rechtecke mit Seitenlängen $1/36$ m und $1/9$ m.

In der Realität würde man jedoch beim Pflastern in der Länge einen kleinen Rand lassen oder aber die Fugen leicht verbreitern und dann wohl nur 47 Steine legen, wodurch dann $47 \times 27 = 1269$ Steine gebraucht würden. Diese Zahl erscheint mir, so gesehen, noch am „richtigsten“.

Aber wofür ist sie überhaupt von Interesse? Nach meiner Erfahrung werden solche Pflastersteine nach Flächeneinheiten verkauft. Aber selbst wenn sie stückweise verkauft würden, dann bestimmt nicht einzeln, sondern vielleicht in 81er- oder 100er-Gebinden. Außerdem werden auf Terrassen üblicherweise viel größere Steine verwendet. Und so weiter.

Unter sämtlichen Gesichtspunkten sind Situation und Fragestellung nicht authentisch. Darüber hinaus ist die Beschreibung der Kompetenzstufe nichtssagend, und was die angesprochenen mathematischen Ideen betrifft, so ist die Arithmetik von erheblich größerer Bedeutung als die Geometrie; – von „Raum“ kann sowieso keine Rede sein.

Selbstverständlich haben solche Textaufgaben ihren Platz im Mathematik-Curriculum; aber ihre Funktion dort ist von ML im Sinne von PISA himmelweit entfernt; und das Kritische ist: die PISA-Expertinnen und -Experten haben offensichtlich dafür kein Gespür.

Das liegt aber nicht nur an deren mangelnder mathematikdidaktischer Expertise, sondern ist in der Sache selbst begründet: Natürlich kann ein Test mit weltweit 250.000 Probandinnen und Probanden (P&P) nur in Form von Häppchen-Aufgaben, wohl oder übel viele im Multiple-Choice-Format, durchgeführt werden. Eigentlich kein einziger Aspekt der ML-Definition kann sich in solchen Aufgaben wiederfinden: Es ist nirgends nötig, eine vorgelegte Situation überhaupt auf Mathematisierbarkeit zu prüfen; denn es ist immer klar, dass zu mathematisieren ist. Es kann nirgends das Erkennen und Verstehen der Rolle der Mathematik in der Welt wirklich aufgezeigt werden. Keine einzige dieser Aufgaben, sei sie noch so komplex aufgebaut, stellt ein authentisches Sachproblem dar, gar ein Problem der P&P selbst; denn Allen ist klar, dass es um einen Test

geht. Natürlich ist keine Aufgabe wirklich offen; es ist lediglich immer wieder der Versuch erkennbar, ein direktes Anwenden von Faktenwissen und Fertigkeiten durch häufig textlastige Einkleidungen zu verhindern, wobei die Autorinnen und Autoren immer wieder über ihre eigenen Füße stolpern.

Beispiel B „Fläche eines Kontinents“: Hier siehst du eine Karte der Antarktis. Schätze die Fläche der Antarktis, indem du den [mit abgedruckten] Maßstab der Karte benutzt.

Diese Aufgabe ist ja ganz nett. Aber sie ist symptomatisch für die ML-Ferne von PISA. Wer den Flächeninhalt der Antarktis wissen will und nicht im Lexikon oder im Internet nachschaut, sondern anfängt, die Karte mehr oder weniger genau auszumessen, verfügt, mit Verlaub, über wenig ML! – Die Kompetenz zur Nutzung externer Informationsquellen kann mit einem Test à la PISA eben nicht gemessen werden.

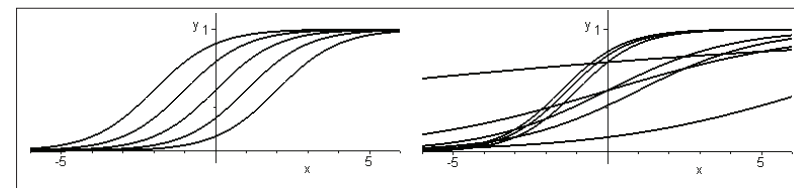
Die typische PISA-Aufgabe entsteht offenbar am Schreibtisch eines männlichen gebildeten Bürgers im angelsächsischen oder niederländischen Raum mit wenig schulischen und anscheinend oft eigentümlichen alltagspraktischen Erfahrungen. Wer als P&P wenig Affinität zu diesem Autorentyp aufweist, hat es eben ein bisschen schwerer mit einem aus einer fremden Sprache übersetzten Aufgabentext, mit geringerer Vertrautheit mit der angelsächsischen Kultur und Mentalität sowie der gehobenen Schicht des Autors und nicht zuletzt mit dessen Bild von der Mathematik und der Realität.

Alle diese aufgeführten mathematikdidaktischen Probleme sind bei einem Unternehmen wie PISA vermutlich unvermeidlich, und das spricht gegen es.

2. DIE PSYCHOMETRIE UND DAS UNZULÄNGLICHE KOMPETENZSTUFENMODELLE IN PISA

Allerdings kommt es in PISA auf die mathematikdidaktische Qualität gar nicht so sehr an, sondern eher darauf, ob eine Aufgabe im psychometrischen Sinn „gut“ misst. – Eine Aufgabe misst gut, wenn sie möglichst trennscharf ist, d.h. wenn es eine Zahl c ($0 < c < 100$) gibt, so dass die c Prozent PISA-schlechtesten Jugendlichen die Aufgabe alle nicht lösen und die anderen sie alle lösen.

Da immer nur Stichproben betrachtet werden, geht es um Lösungswahrscheinlichkeiten, und man ist auch schon mit Aufgabencharakteristiken (Lösungswahrscheinlichkeit als Funktion der P&P-Testleistung) der Form $1/(1+\exp(c-t))$ zufrieden; jedenfalls dürfen sie nicht wie im rechten Bild aussehen; oder gar Bereiche mit negativer Steigung haben.



Letzteres gibt es gar nicht so selten. Ein schönes Beispiel stammt von Wartha (2009):

Beispiel C (nicht aus PISA): Herr Brinkmeier hat bei einer Fernsehlotterie gewonnen. Er möchte den sechsten Teil seines Gewinns einem Kinderheim spenden. Sein Gewinn beträgt 2400 €. Wie viel Geld spendet er?

Im 5. Schuljahr betrug die Lösungshäufigkeit in einer Stichprobe in Bayern 76% (Gymnasium 89%, Realschule 81%, Hauptschule 59%), im 7. betrug sie nur noch 59% (G 76%, R 53%, H 45%). Im Text war allerdings für das 7. Schuljahr eine kleine Veränderung vorgenommen worden: „den sechsten Teil“ war ersetzt worden durch „ein Sechstel“. – Tatsächlich haben die Älteren häufig formalisierte, und damit oft fehlerbehaftete Bruchrechnung eingesetzt, mit der sie ja in der Zwischenzeit intensiv konfrontiert worden waren, während die Jüngeren – adäquat – viel elementarer gerechnet haben. Wartha erklärt die unterschiedliche Vorgehensweise mit dem veränderten Text. Ich will das nicht komplett ausschließen, aber ich meine, dass – i.W. unabhängig von der Formulierung – die Jüngeren einfach unbefangener herangegangen sind.

Bei PISA werden schlecht messende Aufgaben in Pilotstudien identifiziert und dann eliminiert. Ob das übrigbleibende Ensemble noch ein adäquates Bild von ML-Mathematik liefert, also valide für ML ist, ist offenbar zweitrangig. Die Mathematikdidaktikerinnen und -didaktiker im deutschen PISA-Team haben sich jedenfalls beklagt, dass sie an dieser Stelle gegen das Diktat der Psychometrikerinnen und -metriker nicht ankommen.

