IM PLENUM

EMPIRISCHE BILDUNGSFORSCHUNG

NOTWENDIGKEIT UND RISIKO

Jörg-Dieter Gauger
Josef Kraus (Hrsg.)

ISBN 978-3-941904-33-0

www.kas.de

Konrad Adenauer Stiftung

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.

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

INHALT

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 Botschaften 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 vergröbert 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

<< zurück zum Inhaltsverzeichnis

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 nachprüfbare 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 Unterricht[1] 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 Jugenderziehung* – 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 Heilpädagogen[5] 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' Zeitschriftengründung, 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 Laboratorien[8], 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 gegründeter Laboratory School in Chicago[10] 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 Intelligentests 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 Umfeldeinrichtungen 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 beschweren 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 Reformer 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 hilfsweise 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 prevoir*! 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 distanzierten 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 Enthaltsamkeit 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.).*

<< zurück zum Inhaltsverzeichnis

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 Grundschulkinder 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 Untersuchungen 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.

Spätestes 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 würden 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 eruierenden 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 Schlüsselelement 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 zufallsgesteuert 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 Lernerfolgen 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 Computerpogrammen (z.B. HLM 6: Raudenbush et al. 2004; vgl. Raudenbush / Bryk 1992), die inzwischen international als gleichsam obligatorisch durchzuführen galten.

Mehrebenen-Analsen 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, mehrebenenanalytisch 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 die. sem 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 und 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 Übergangsentscheidung 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 Bildungsbeteiligungsquoten 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 übergangen. 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. (Hrgs.) (2008): Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-Studierender und -Referendare – erste Ergebnisse zur Wirksamkeit der Lehrerausbildung. Münster.*

- *Bos, W. / Hornberg, S. / Arnold, K.-H. / Faust, G. / Friede, L. / Lankes, E.-M. / Schwippert, K. / Valtin, R. (2007): IGLU 2006. Lesekompetenzen von Grundschülern 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 Lese- 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 Stritzy, 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. (Hrgs.) (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. (Hrgs.) (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. (Hrgs.) (2001): Defining and Selecting Key Competencies. Göttingen / Bern, 45-65.*

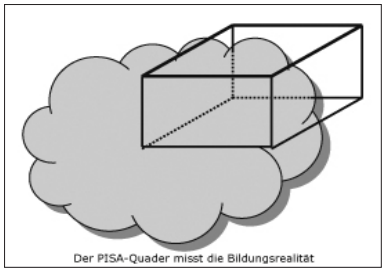
<< zurück zum Inhaltsverzeichnis

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.


Der PISA-Quader misst die Bildungsrealität

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 vergröbernden 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 wohlbekannt und in der Bildungsdebatte hochrelevant ist. In PISA werden zwar keine *primitiven* Fehler gemacht; bzw. die primitiven Fehler stammen von den manchmal naiven, oft eigennützigen 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 x 3 x 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 x 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.

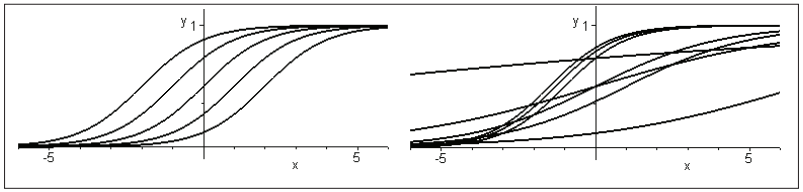
Die typische PISA-Aufgabe entsteht offenbar am Schreibtisch eines männlichen gebildeten Bürgers im angelsächsischen oder niederländischen Raum mit wenig schulischen und anscheinend oft eigentümlichen alltagspraktischen Erfahrungen. Wer als P&P wenig Affinität zu diesem Autorentyp aufweist, hat es eben 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 Lösungs*wahrscheinlichkeiten*, 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 Bruchrechung 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.

Bei den verbleibenden Aufgaben wird jedenfalls unterstellt, dass ihre Charakteristik i.W. wie oben aussieht, d.h. man arbeitet nicht mit den realen Lösungswahrscheinlichkeiten, sondern mit einem mathematischen Modell, dem sog. Rasch-Modell. Joachim Wuttke (2007) hat festgestellt, dass bei vielen Aufgaben die realen Abweichungen von einer idealen Charakteristik jedoch erheblich sind, und es fragt sich, wie weit sie akzeptabel sind. Georg Rasch selbst hat übrigens erklärt, dass sein Modell lediglich für die Untersuchung ganz primitiver Items geeignet ist und nicht für komplexe Fragen (wie etwa PISA-Mathematik-Aufgaben).

Man braucht ein ganzes Ensemble gut messender Aufgaben, deren Trennpunkte c sich einigermaßen gleichmäßig über den Bereich von 0 bis 100 verteilen (wie oben im linken Bild), und hat eine Skala für die Aufgabenschwierigkeit: Je höher der Trennpunkt, desto weniger P&P lösen die Aufgabe, desto schwieriger ist sie also.

Die Dualität zu der entsprechenden Skala für die Testleistungen der P&P liegt auf der Hand.

Die Testpunktzahlen werden noch so normiert, dass der Mittelwert 500 und die Standardabweichung 100 beträgt. Diese Normierung wird allein auf der Basis der OECD-Länder vorgenommen, d.h. unter Ausschluss der Daten der Partnerländer wie Brasilien. Sie wird außerdem für bestimmte Berechnungen auf die Mittelwerte jeweils früherer Durchgänge bezogen und weicht dann für den jeweils aktuellen Durchgang vom Wert 500 ab.

Unbedingt ist zu beachten, dass die PISA-Zahlen immer nur relativ zu verstehen sind. Wenn z.B. die deutschen Jugendlichen im Jahr 2003 im sog. Problemlösen 513 Punkte und in Mathematik 503 Punkte erzielten, heißt das nicht, dass sie in Problemlösen besser als in Mathematik sind (dieser Vergleich ist sowieso sinnlos), sondern nur, dass sie in PISA-Problemlösen im Vergleich zu den anderen Ländern besser abgeschnitten haben als in PISA-Mathematik im Vergleich zu den anderen Ländern.

Ein weiterer beliebter Fehlschluss besteht darin, die Länderpunktzahlen und damit die Rangplätze als exakt anzunehmen. Die Punktzahlen sind, als Ergebnis von Stichproben, aber immer mit dem sog. Standardfehler behaftet, der von PISA auch stets angegeben wird. Daher muss man nahe beieinander liegende Länder zu Clustern zusammenfassen, weil bei Variation der Stichproben die Reihenfolge sich ohne Weiteres um mehrere

Plätze verändern könnte, z.B. 2006 in Mathematik: Österreich 505, Deutschland 504, Schweden 502 und Irland 501.

Die Werte der P&P-Testleistungsskala werden nun folgendermaßen auf die Aufgabenschwierigkeitenskala übertragen: Für eine bestimmte Aufgabe wird für jede Testpunktzahl die Menge der P&P mit dieser Punktzahl betrachtet und ermittelt, wie hoch der Anteil derer, die die Aufgabe richtig gelöst haben, an *diesen* P&P ist. Es wird unterstellt – bei allen genannten Vorkehrungen wohl zu Recht –, dass mit zunehmender Testpunktzahl dieser Anteil wächst (je besser die P&P, desto eher lösen sie eine bestimmte Aufgabe). Dann wird diejenige Testpunktzahl, bei der der Anteil der Löserinnen und Löser erstmals 62% beträgt, als die Schwierigkeit dieser Aufgabe festgelegt.

Der Wert 62% ist willkürlich. Er gibt die Meinung eines anonymen, zufälligen, vorübergehenden Kollektivs von sog. Expertinnen und Experten (für: was weiß ich) darüber wieder, ab wann man mit der Lösungsquote eines Kollektivs wohl zufrieden sein kann.

Nun ist also eine gemeinsame Skala vorhanden. Zur weiteren Vereinheitlichung werden die P&P-Testleistungen und die Aufgabenschwierigkeiten unter den gemeinsamen Begriff „Kompetenzen" gefasst: die P&P *verfügen* über Kompetenzen, und die Aufgaben *erfordern* Kompetenzen, die mit der Skala simultan „gemessen" werden.

Sinnvollerweise hat das internationale PISA-Konsortium diese Skala (für die Inhaltsbereiche Mathematik, Lesen usw. sowie für einzelne Teilbereiche jeweils separat) in Stufen eingeteilt und zwar erklärtermaßen willkürlich, lediglich zum Zweck des leichteren Redens darüber, und die inhaltliche Beschreibung den nationalen Gruppierungen überlassen. So gesehen, ist nichts dagegen einzuwenden, dass

- (i) die Stufen alle gleich breit gemacht wurden,
- (ii) die Lage der Stufen und die gemeinsame Breite von der (bei verschiedenen Inhaltsbereichen bzw. Teilbereichen sowie bei verschiedenen Durchgängen allerdings unterschiedlichen) Anfangs- und Endpunktzahl (ca. 300 und ca. 700) sowie von der Anzahl der Stufen abhängt und
- (iii) P&P oder Aufgaben auf unterschiedlichen Stufen landen, je nach dem, wer alles am Test teilgenommen hat.

Die deutsche PISA-Mathematik-Gruppe hat allerdings aus dieser ersichtlich zufälligen Stufeneinteilung Großes gemacht und ein ganzes Kompetenzstufenmodell darauf gegründet. Jeder Stufe wurden gewisse Kompetenzen zugewiesen, und idealerweise soll sich dann allein aus der PISA-Punktzahl bzw. -Stufe quasi naturgesetzlich erschließen lassen, welche Kompetenzen ein Mensch hat bzw. eine Aufgabe erfordert.

Diese Gruppe geht anscheinend davon aus, dass

- (i) sich die möglichen Kompetenzen überhaupt sinnvoll linear anordnen lassen und
- (ii) der Aufgabensatz von PISA geeignet ist, diese Ordnung treu auf die Punkteskala zu übertragen und sie damit zu metrisieren, bzw. dass ein solcher Aufgabensatz wenigstens denkbar ist.

Beide Annahmen sind höchst naiv und werden in den ausführlichen Analysen auch nicht wirklich substantiiert.

Ob die anderen Länder eigentlich zum selben Modell gekommen sind (was sie ja eigentlich wegen dessen Naturgesetzlichkeit müssten)? Es sollte im ureigenen Interesse von PISA liegen, einmal die Kompetenzstufenmodelle der ca. 50 PISA-Länder zu vergleichen. Derartige Vergleiche existieren m.W. nicht, vermutlich weil die anderen Länder sich nicht die Mühe gemacht haben, solche Modelle breit zu entwickeln, sondern bestenfalls ein paar (mehr oder weniger triviale) Stichworte aufgeschrieben haben.

Tatsächlich landen Menschen mit ähnlichen PISA-Punktzahlen auf ein und derselben Stufe, und wenn sie noch so unterschiedliche Kompetenz*profile* besitzen.

Eine noch stärkere Mehrdeutigkeit besteht bei den Aufgaben: Schon verschiedene Aufgaben*teile* können unterschiedliche Kompetenzen erfordern und dadurch auf verschiedene Stufen gehören. Eine ähnliche Uneindeutigkeit wird vom spezifischen Wissen der P&P, von ihrer Vertrautheit mit der jeweiligen Aufgabe bzw. dem Kontext oder *vom jeweils eingeschlagenen Lösungsweg* erzeugt (Wolfram Meyerhöfer).

Beispiel D: Die Grundfläche einer *Pyramide* ist ein Quadrat. Jede Kante der skizzierten Pyramide misst 12 cm [Zeichnung]. Bestimme den Flächeninhalt einer der dreieckigen Seitenflächen.

Man muss das ebene Problem in der räumlichen Situation sehen, und das fällt einem umso leichter, je mehr man sich in der Schule mit räumlichen Sachverhalten befasst hat. Dann braucht man nur den Flächeninhalt eines gleichseitigen Dreiecks zu kennen, und es handelt sich um eine reine Wissensaufgabe. Aber auch wenn man diesen zuerst noch herleiten muss, liegt hier kein „komplexes Modellieren" vor, was aber für die Kompetenzstufe charakteristisch sein soll, auf der diese Aufgabe wegen ihres hohen Schwierigkeitsgrads von 810 Punkten landet.

Beispiel E: Wie kannst du einen Geldbetrag von genau 31 Pfennig hinlegen, wenn du nur 10-Pfennig-, 5-Pfennig- und 2-Pfennig-Münzen zur Verfügung hast?

Wegen ihrer hohen Punktzahl von 797 wird diese Aufgabe unter „begriffliches Modellieren und Problemlösen" eingeordnet. Ich kann diese Einordnung inhaltlich nicht nachvollziehen. Es handelt sich doch um eine begrifflich völlig anspruchslose Abzählaufgabe an vorgestellten oder aufgezeichneten konkreten Objekten, und die Schwierigkeit liegt nur in der Erfassung aller Fälle.

Am schönsten zeigt sich die Unzulänglichkeit dieses Kompetenzstufen-Ansatzes m.E. an folgender Aufgabe, die zwar dem Naturwissenschaftentest von IGLU 2001 (mit 10-jährigen P&P) entnommen ist, dessen Stufenmodell aber denselben Prinzipien gehorcht wie das von PISA-Mathematik.

Es hat folgende Stufen: 0 „Vorschulisches Alltagswissen", I „Einfache Wissensreproduktion", II „Anwenden alltagsnaher Begriffe", III „Anwenden naturwissenschaftsnaher Begriffe", IV „Beginnendes naturwissenschaftliches Verständnis", V „Naturwissenschaftliches Verständnis und Lösungsstrategien".

Beispiel F: Welches Tier säugt seine Jungen? Huhn, Frosch, Affe, Schlange?

Diese Aufgabe gehört ersichtlich auf Stufe I oder gar 0, je nach den Fernseh-Erfahrungen der Kinder, auf keinen Fall aber auf II oder gar III. – Aus welchen Gründen auch immer, sie fällt den Kindern recht schwer und landet mit 474 Punkten doch auf III.

Natürlich gibt es schon bei der Bewertung der Aufgabenlösungen Unklarheiten noch und noch. Z.B. würde ich bei der Terrassen-Aufgabe die Antwort 1275 als inkorrekt und dafür 1269 als korrekt bewerten.

Aber i.W. sind die P&P-Punktzahlen doch harte Daten, auch wenn sie nicht nur das Ergebnis von ML-Kompetenzen sind, sondern auch von Rate-Vermögen, Erfahrungen mit Tests (die in Deutschland – noch – geringer ausgeprägt sind), Abfolge der Aufgaben, Verteilung auf die Testhefte, Tageszeit, zu der der Test stattfindet, Konzentrationsfähigkeit usw.

3. HARTE SOZIOMETRIE MIT EXTREM WEICHEN DATEN IN PISA

PISA hat aber höhere Ambitionen, und zwar sollen die Testergebnisse mit dem ökonomischen, sozialen und kulturellen Status der P&P verknüpft werden. Die Daten dazu hat man u.a. mit Fragebögen gewonnen, auf denen die Jugendlichen selbst über sich und ihre Familie Auskunft geben sollten.

Da hat man sich nach dem Vorhandensein gewisser Haushaltsgeräte erkundigt oder Fragen folgender Art gestellt: „Wie viele Bücher habt ihr zu Hause?" oder „Wie oft kommt es im Allgemeinen vor, dass deine Eltern mit dir über Bücher, Filme oder Fernsehsendungen diskutieren?" – Das ergibt offensichtlich extrem weiche Daten.

So haben z.B. – völlig neben der Realität – die Schweizerinnen und Schweizer ihre Lage schlechter eingeschätzt als die Deutschen.

Oder: Bei PISA 2003 haben in Schleswig-Holstein 43,0%, in Bayern 24,8% und in Deutschland 23,9% der Jugendlichen angegeben, schon einmal eine Klasse wiederholt zu haben (PISA 2005, 169ff, Klemm, 9, Abb. 2). Diese Werte spiegeln sich in den W&W-Quoten, die für die Schuljahre 1995/96 und 2000/01 in Klemm (19, Tab. 3) abgedruckt sind, völlig unterschiedlich wieder. Bei Schleswig-Holstein würde man besonders hohe Quoten erwarten; sie sind aber niedriger als die von Bayern. Wenn man die anderen Bundesländer einbezieht, wird das Bild noch viel uneinheitlicher. – Haben die Jugendlichen in Schleswig-Holstein ein anderes Verständnis vom Sitzenbleiben als die in Bayern? Bekennen sie sich eher zum Sitzenbleiben?

Oder: Nur 40% aller Jugendlichen in Deutschland haben als genauen Beruf ihres Vaters denselben angegeben wie dieser; und auch wenn sie ihn nur grob nennen sollten, lag die Übereinstimmung noch unter 70%. Jeweils etwas größer ist die Kohärenz bezüglich der Mutter, weil diese häufiger keiner Erwerbstätigkeit nachgeht und dieser Status von den Jugendlichen leichter erkannt wird.

3.1. Der „soziale Gradient" ist unbrauchbar

Der „höchste" Beruf in der Familie spielt aber eine ganz wichtige Rolle: der für PISA relevante Status der Familie wird reduktionistisch i.W. in Form des sog. HISEI mit ihm identifiziert (ISEI: *International Socio-Economic Index of Occupational Status*; HISEI: *Highest ISEI*). Man hatte 2003 einen anderen Parameter verwendet, den ESCS (*Economic, Social, and Cultural Status*). Wegen eines total uneinheitlichen Bildes beim Vergleich von 2000 mit 2003 hat man im Bericht über 2006 wieder den HISEI für den Vergleich der *drei* Durchgänge herangezogen (PISA 2007, 323).

Für jedes Land wird *seine* Abhängigkeit der Variablen „PISA-Punktzahl" (und zwar beim Lesen) von der Variablen „sozialer Status der Familie" (dem HISEI) mittels einer linearen Regression dargestellt. Bei einer solchen Analyse mit zwei Variablen entsteht eine Punktwolke, und diese wird durch eine Gerade repräsentiert. Je größer deren Steigung (der sog. soziale Gradient) ist, desto ausgeprägter erscheint die Abhängigkeit.

Die beim Lesetest erzielten Punktzahlen sind zwar (mit gewissen Abstrichen) in voller Genauigkeit vorhanden. Die Variable „Sozialstatus" dagegen ist, wie gesagt, wachsweich. Deren Anordnung auf einer linearen Skala ist eine erneute fragwürdige Reduktion. Wie man dann noch darauf eine Metrik pfropfen kann, ist mir unbegreiflich. Da werden ja *Abstände* etwa zwischen Professor und Astrologe oder zwischen Botschafter und Tänzer (um einmal einige der Berufe zu nennen) als Zahlen definiert, und mit diesen wird auf zwei Stellen hinter dem Komma genau Regressionsrechnung getrieben, zwecks exakter Vermessung des PISA-Quaders.

Wenn man sich aber einmal auf diese Vorgehensweise einlässt, dann ist klar, dass große Gruppen mit sehr niedrigen Punktzahlen in Verbindung mit niedrigem sozialem Status die Steigung des sozialen Gradienten erhöhen. Und daran haben unsere Jugendlichen mit Migrationshinter-

grund (MH; wenigstens ein Elternteil im Ausland geboren), und zwar vornehmlich die mit doppeltem MH aus bestimmten Ländern, zusätzlich zum schwachen Viertel unserer eingeborenen Jugendlichen einen erheblichen Anteil.

Während sich Deutschlands sozialer Gradient von 2000 bis 2006 von 45 über 38 bis 35 deutlich verringerte, hat es neben ähnlich starken positiven wie negativen Entwicklungen in anderen Ländern in diesen 6 Jahren auch ausgeprägte Berg- und Talfahrten gegeben, und das trotz der einheitlichen Verwendung des HISEI (ebenda, 323):

Island	19	12	18
Kanada	26	22	25
Korea	15	19	17
Österreich	35	40	35
Portugal	38	31	39
Schweiz	40	30	32
Tschechien	43	32	46

Hier hätte deutlich ausgesprochen gehört, dass dieser Gradient unbrauchbar, weil zu labil, ist. Das tun die Autoren nicht; sie geben (ebenda, 323) „Veränderungen in der Stichprobenausschöpfung und veränderte Anteile von fehlenden Werten" (ein PISA-Euphemismus für „Mängel in der Erhebung") als mögliche Ursache an und stellen damit redlicherweise auch gleich noch ihr Stichprobenauswahlverfahren in Frage.

Dieser Gradient war ja in der deutschen Bildungsdiskussion nach 2000 von interessierten Kreisen als Grundlage für die Parole von der in Deutschland besonders großen Abhängigkeit der Schulleistungen vom sozialen Status benutzt worden, die bis heute als Begründung für die Einheitsschule herhalten muss. Es ist wohl zu viel verlangt, sich selbst und diesen Kreisen deutlich zu machen, dass man da i.W. einem Artefakt aufgesessen ist.

3.2. Die „relative Wahrscheinlichkeit des Gymnasialbesuchs" ist nichtssagend

Ein noch fragwürdigerer Parameter ist die sog. „relative Wahrscheinlichkeit des Gymnasialbesuchs" (rWG), sehr grob gesprochen: der Quotient aus dem Verhältnis der Anzahl der „reichen" 15-jährigen Gymnasiastinnen und Gymnasiasten (G&G) zu der Anzahl der „reichen" 15-jährigen Nicht-G&G und dem Verhältnis der Anzahl der „armen" 15-jährigen G&G zu der Anzahl der „armen" 15-jährigen Nicht-G&G.

Warum nicht Anteile (von G&G an *allen* „reichen" 15-Jährigen bzw. an *allen* armen 15-Jährigen), sondern diese Verhältnisse („odds") zueinander in Bezug gesetzt werden („odds ratios" gebildet werden), ist mir nicht klar. Vermutlich hat sich das irgendwo „bewährt" (eine Begründung, die in den PISA-Berichten immer wieder einmal auftaucht). Jedenfalls wachsen die „odds" überproportional mit den Anteilen, und zwar zunehmend rasanter: wenn der Anteil gegen 100% geht, gehen die „odds" gegen ∞. Beispiel: Beträgt der Anteil bei den „Reichen" 60% und bei den „Armen" 20%, dann ist der rWG nicht 60/20 = 0,6/0,2 = 3, sondern die „odds" lauten 60/40 = 1,5 sowie 20/80 = 0,25 und die rWG = 1,5/0,25 = 6. Bei hoher Gymnasialbeteiligung der „Reichen" wird also durch Verwendung der „odds" der Zahlenwert der rWG deutlich erhöht.

Da dieser Parameter ersichtlich nicht für Vergleiche mit dem Ausland gedacht ist, stellt die Verwendung der „odds ratios" ein einfaches Mittel dar, höhere Zahlenwerte für die „sozialen Disparitäten" im deutschen Bildungssystem zu erhalten. Das Hervorbringen „schlechter" Nachrichten gehört ja zum Erfolgsrezept von PISA, und die Veröffentlichung der Werte des rWG von 2003 führten zu Schlagzeilen wie „Chancenungleichheit in Deutschland wächst" bzw. „Chancenungleichheit in Bayern am größten". Der damals für Bayern besonders hohe Wert geht übrigens nicht auf das eben beschriebene Phänomen mit den „odds ratios" zurück, da in Bayern ja die Gymnasialbeteiligung in fast allen sozialen Schichten niedriger als in Deutschland insgesamt ist, insbesondere auch bei den „Reichen". – Aber eben auch bei den „Armen", und das führte, jedenfalls 2003, zu dem hohen Wert.

In den PISA-Durchgängen 2000 und 2003 war Bayern das beste PISA-Bundesland, eines der besten Länder der Welt und bei Kontrolle der Migrationsquote (MQ) sogar mit das beste Land der Welt überhaupt.

Außerdem hatte Bayern damals unter den alten Bundesländern den zweitniedrigsten sozialen Gradienten, und das alles mit einem konservativen, betont dreigliedrigen Schulsystem. Aber bei der rWG war Bayern 2003 in der Variante „mit Kontrolle der Lese- und Mathematikkompetenz" das schlechteste Bundesland. Obwohl es in der Variante „ohne Kontrolle von Kovariaten" in der Nähe des Durchschnitts lag und Sachsen-Anhalt sowie Bremen Spitze waren und im PISA-Bericht konzediert wurde, dass die Wahrheit wohl irgendwo zwischen den beiden Varianten liegt (PISA 2005, 262), wurden an die breite Öffentlichkeit nur die Werte der erstgenannten Variante gebracht, und zwar mit den o.a. reißerischen Aufmachern.

Nun war in gewissen Kreisen der Jubel groß, als 2006 Sachsen mit seinem zweigliedrigen Schulsystem Bayern als PISA-Spitzenbundesland ablöste, hatte man doch endlich einen Beleg für die Überlegenheit der Einheitsschule (jedenfalls wollte man dieses Ergebnis so interpretiert wissen). Allerdings ließ man außer Acht, dass die neuen Bundesländer allesamt eine viel geringere MQ als die alten Bundesländer haben (ca. 5% gegenüber 22% bis 41%) und dass außerdem dort die Migrations*struktur* erheblich günstiger ist. Bei Kontrolle der MQ liegt natürlich nach wie vor Bayern deutlich vorne, und mit der Zweigliedrigkeit seines Schulsystems hat der Erfolg Sachsens herzlich wenig zu tun.

Nachdem nun also die Ergebnisse des PISA-Durchgangs 2006 vorlagen, wurde für die Bundesländer die rWG für die Durchgänge 2000 und 2006 explizit verglichen (PISA 2008, 338). Zusammen mit den Werten für 2003 ergibt sich folgendes Bild:

	2000		2003		2006	
	o. Kontr.	(mit K.)	o. Kontr.	(mit K.)	o. Kontr.	(mit K.)
Baden-Württemberg	5,8	(3,2)	8,41	(4,40)	5,6	(4,0)
Bayern	10,5	(6,5)	7,77	(6,65)	4,3	(2,7)
Berlin			4,46	(2,67)		
Brandenburg	3,2	(1,9)	3,71	(2,38)	4,8	(4,3)
Bremen	6,1	(3,0)	9,06	(2,83)	4,8	(3,2)
Hamburg			7,53	(3,55)		
Hessen	6,5	(2,7)	5,70	(2,71)	5,6	(3,4)
Mecklenburg-Vorpommern	6,0	(4,0)	7,96	(3,47)	3,2	(2,3)
Niedersachsen	7,8	(5,0)	6,45	(2,63)	4,8	(4,8)

	2000		2003		2006	
	o. Kontr.	(mit K.)	o. Kontr.	(mit K.)	o. Kontr.	(mit K.)
Nordrhein-Westfalen	6,5	(3,1)	8,07	(4,35)	6,7	(4,5)
Rheinland-Pfalz	9,1	(5,1)	8,28	(4,60)	4,0	(2,6)
Saarland	6,0	(3,5)	6,71	(3,48)	5,5	(4,1)
Sachsen	3,1	(2,2)	4,49	(2,79)	3,9	(2,8)
Sachsen-Anhalt	4,4	(3,1)	10,44	(6,16)	3,3	(3,0)
Schleswig-Holstein	8,1	(5,8)	6,24	(2,88)	5,0	(2,9)
Thüringen	4,0	(3,2)	5,13	(3,23)	3,0	(2,2)
Deutschland	6,0	(3,2)	6,87	(4,01)	4,6	(3,2)

O. Kontr. = ohne Kontrolle von Kovariaten; mit K. = mit Kontrolle der Lesekompetenz und 2003 zusätzlich mit Kontrolle der Mathematikkompetenz. Da Hamburg und Berlin wegen fehlender Repräsentativität im Durchgang 2000 nicht extra ausgewiesen worden waren, wurden auch ihre Ergebnisse aus 2006 nicht publiziert. Warum eigentlich nicht?

In den Durchgängen 2000 und 2006 einerseits sowie 2003 andererseits wurden die Klassen, aus denen dann jeweils zwei für den Vergleich zwischen „reich" und „arm" ausgewählt wurden, unterschiedlich definiert. 2000 und 2006 ging es nach sog. EGP-Klassen (ebenda, 322f), und aus den 7 Klassen wurden zunächst die Klassen V und VI (Facharbeiter und Arbeiter mit Leitungsfunktion) zusammengefasst und dann alle anderen Klassen mit dieser verglichen. In der o.a. Tabelle sind die Zahlen für den Vergleich der Oberen Dienstklasse (I) mit dieser zusammengesetzten Klasse angegeben. Das sind genau die Zahlen, die von PISA und von der Öffentlichkeit als die entscheidenden angesehen werden. 2003 ging es nach Quartilen der sog. ESCS-Klassifikation, und alle Quartile wurden mit dem dritten Quartil verglichen. Auch hier wird nur der Vergleich des ersten mit dem dritten Quartil als wesentlich angesehen, und diese Zahlen sind oben aufgeführt.

Für eine Einschätzung der Nützlichkeit der rWG ist es prinzipiell unerheblich, nach welchen Gesichtspunkten die Klassen festgelegt sind. Wichtig ist, dass es da stark unterschiedliche Möglichkeiten gibt, auf deren Basis jedes Mal ein scheinbar sinnvoller Parameter definiert werden kann, der „rWG" genannt werden kann (und bei PISA auch so genannt wird). Dass dann unterschiedliche Zahlenkollektionen herauskommen, ist zu erwarten und spricht noch nicht gegen die unterschiedlichen Definitionen. Aber wenn man sich den fast zufälligen Zahlensalat in der obigen Tabelle bei waagrechten und lotrechten Vergleichen anschaut, stellen sich doch

erhebliche Zweifel an irgendeiner Aussagekraft dieses Parameters ein, wohlgemerkt, auch wenn man die Unterschiedlichkeit der Definitionen 2000 und 2006 einerseits sowie 2003 andererseits ins Kalkül zieht.

In (Bender 2007, 291) habe ich, z.T. die Argumente aus dem PISA-Bericht aufnehmend, diesen Parameter bereits inhaltlich kritisiert. Zum Beispiel besitzen mehrere Bundesländer (darunter häufig solche mit niedrigem Wert) in nennenswertem Umfang Gesamtschulen, die zum Abitur führen, die aber bei diesen Rechnungen nicht berücksichtigt sind (PISA 2005, 262, Fußnote 5), und dass es zusätzlich auf den Expansionsgrad der Gymnasien ankommt (ebenda, 263), der ebenfalls nicht einbezogen wurde. Schon dort habe ich die Willkür in der Wahl der ESCS-Skala und der Einteilung in Quartile bemängelt. Der jetzt vorliegende Vergleich der drei PISA-Durchgänge bestätigt meine Skepsis voll und ganz.

Die Übersicht über die Gymnasialbeteiligung in jeder einzelnen Klasse (PISA 2008, 336) wäre wohl ergiebiger. Die durchweg niedrigeren Werte von Bayern i.V.m. seinen hohen Leistungspunkten sind Ausfluss eines bis vor kurzem noch intakten dreigliedrigen Schulsystems mit wenigen Schulabgängerinnen und -abgängern ohne Abschluss sowie geringerer Jugendarbeitslosigkeit und höchstem Leistungsniveau bei allen Schulformen. Mit seiner geringeren Migrationsquote hat es Bayern leichter als die anderen alten Bundesländer. Aber es hat sich auch bis vor kurzem erfolgreich dagegen gewehrt, dass die Hauptschule schlecht gemacht bzw. geredet wird.

Leider verbreitet sich in der deutschen Bildungsdiskussion zunehmend die Auffassung, dass der Mensch erst mit dem Abitur anfängt (so zugespitzt, würden sich natürlich alle distanzieren), und zwar mit dem Abitur des allgemeinbildenden Gymnasiums (obwohl etwa die Hälfte aller Hochschulzugangsberechtigungen auf anderem Weg erworben werden). Da ist auch der Duktus des PISA-Berichts verräterisch, wenn er (ebenda, ab S. 335) praktisch „Bildungsbeteiligung" mit „Gymnasialbeteiligung" gleichsetzt. Gewiss, diese beiden Parameter sind halt entsprechend definiert; und genau das ist zu kritisieren. Auch nicht-gymnasiale Bildung ist Bildung!

Den sozialen Gradienten könnte man einer ähnlichen innerdeutschen Analyse unterziehen und käme vermutlich zu ähnlichen Inkonsistenzen wie beim internationalen Vergleich und bei der rWG. Bei beiden Parametern ist der einstige Musterknabe „Brandenburg" trotz seiner Zweigliedrigkeit 2006 weit nach „vorne" gerückt.

Es ist amüsant zu lesen, wie sich der glühende Einheitsschulverfechter Christian Füller in der *taz* am 30.06.2009 an Erklärungen für diese Entwicklung abmüht, statt einfach festzustellen, dass die Parameter wertlos sind, auch wenn man den einen früher gern gegen Bayern in Stellung gebracht hat. Immerhin hat Füller dabei die Ausbreitung der Privatschulen in den Blick genommen. Vielleicht gelingt ihm da ja noch der Transfer zum Szenario der Einheitsschule.

4. BEISPIELE FÜR MÄNGEL BEI DER DATENERHEBUNG VON PISA

Das Image der Objektivität und Neutralität, das PISA so sehr pflegt, ist ein, allerdings nicht leicht zu durchschauender, Mythos. So sind bei den Erhebungen Unzulänglichkeiten noch und noch vorgekommen, manchmal sogar von PISA selbst vorgesehen, manchmal gegen die PISA-Absicht von Schulbehörden, -lehrerinnen und -lehrern, Schülerinnen und Schülern (S&S) fabriziert, manchmal ungeplant; z.T. repariert, z.T. nicht; z.T. veröffentlicht, z.T. nicht; z.T. selbst, z.T. von anderen, z.T. von niemandem bemerkt.

Zum Beispiel hat man sich in *Südtirol* nach 2003 für eine der besten PISA-Regionen der Welt gehalten, bis kritische Geister recherchiert haben, dass zahlreiche Schulen, vor allem Berufsschulen, mit einer schwächeren Klientel vom Test ausgeschlossen worden waren. Ich bin überzeugt, dass weltweit immer wieder solche Manipulationen auf Schul- und auf S&S-Ebene vorkommen.

In vielen Ländern, vor allem in *Entwicklungsländern*, geht ein großer Teil der 15-Jährigen (die und nur die von PISA untersucht werden) gar nicht (mehr) zur Schule und wird nicht erfasst. Aber auch in den *OECD-Ländern* Türkei, Mexiko und Portugal beträgt die Quote der erfassten S&S nur 54%, 58% und 86%, dagegen etwa in USA (trotz der vielen illegalen Migrantinnen und Migranten dort) genau 100,0%, in Schweden gar 102% und in der Toskana satte 108%.

Nach dem starken Absinken *Österreichs* von 2000 bis 2003, in den drei Inhaltsbereichen durchschnittlich um elf PISA-Punkte, ließ die Regierung die Ergebnisse von 2000 nachrechnen und erklärte schließlich, dass diese auf unzulässigen Stichproben beruhten und nach unten korrigiert werden müssten.

Als in den *Niederlanden* 2000 und in *Großbritannien* 2003 die vorgeschriebenen Teilnahmequoten von Schulen sowie von S&S nicht erreicht wurden, wurden diese Länder aus dem Ranking, ihre Daten jedoch nicht aus den Auswertungen ausgeschlossen. Die *USA* wiederum wurden gar nicht ausgeschlossen, obwohl ihre Teilnahmequoten noch viel niedriger lagen.

2000 haben in *Berlin* und in *Hamburg* im Gesamtschulbereich so viele Schulen ihre Mitarbeit verweigert, dass diese beiden Länder aus dem innerdeutschen Vergleich herausgenommen werden mussten.

Das schlechte Ergebnis von *Luxemburg* (446) im Jahr 2000 wurde damit erklärt, dass „Unterschiede [ein PISA-Euphemismus für „Fehler"] in der … Zuordnung der Testhefte nach Sprachgruppen" vorgekommen waren. Bei den nächsten Durchgängen wurde dieser Fehler vermieden. In der Folge stieg Luxemburg auf 493 und sank dann wieder leicht auf 490.

Starke Punktzahlveränderungen bei vielen Ländern innerhalb von drei Jahren oder gar die Berg- und Talfahrt von *Tschechien* von 498 über 516 bis 510 (alles noch viel gehäufter und extremer bei einem angepassten Einbezug der TIMSS-Ergebnisse von 1995 bis 2007) werden nicht als Anlass für Skepsis gegenüber der Statistik, sondern als Ausfluss und damit als Indikator einer entsprechenden Leistungsentwicklung genommen. Da wird sogar dem OECD-PISA-Verantwortlichen Andreas Schleicher widersprochen, wenn er den Anstieg der Deutschen von 2003 bis 2006 in Naturwissenschaften von 502 auf 516 Punkte mit der Unvergleichbarkeit der beiden Durchgänge relativiert. Hier pflichte ich Schleicher ausnahmsweise bei.

In *Südkorea* waren nur 40% der Teilnehmenden Mädchen, obwohl ihr Anteil an allen Jugendlichen 48% lautet. Diese Differenz beträgt zehn Standardabweichungen und ist damit ganz gewiss nicht zufällig.

Die Definition sowie der Ein- oder Ausschluss von fremdsprachlichen, behinderten oder legasthenischen Jugendlichen wird immer wieder unterschiedlich gehandhabt, was zu ausgeprägten Verzerrungen bei den Länderpunktzahlen führt. Solcherart begründete Ausschlüsse waren bis zu 5% statthaft. Obwohl 2003 die Länder *Dänemark, Kanada, Neuseeland, Spanien, USA* diese Grenze z.T. deutlich (bis zu 7,3%) überschritten, wurden deren Daten ohne Weiteres in die Auswertung mit aufgenommen.

Für die Kalibrierung der Aufgabenschwierigkeiten wurden aus jedem OECD-Land vorab 500 Jugendliche ausgewählt. Die Jugendlichen aus den *USA* haben dabei also etwa 1/1000 des Gewichts der Jugendlichen aus *Island*, und die Jugendlichen aus *Brasilien* haben gar kein Gewicht. – Hätte man die einbezogenen Jugendlichen aller OECD-Länder alle bevölkerungsproportional gewichtet, dann hätten alle Punktzahlen aufgrund dieser scheinbar geringfügigen Änderung jedes Mal ca. 10 Punkte besser ausgesehen (die deutschen in Mathematik also 500, 513, 514).

Im Vergleich zu diesen bekannt gewordenen Unzulänglichkeiten dürfte die Dunkelziffer weltweit viel, viel größer sein. – Solche Mängel sind bei einem Unternehmen wie PISA unvermeidlich; und das spricht, wie gesagt, gegen es.

5. WAS HAT DIE STUDIE DER BERTELSMANN-STIFTUNG ZUM SITZENBLEIBEN WIRKLICH GEZEIGT?

Nachdem das Instrument der Kassandra-Rufe zum deutschen Schulsystem in den letzten Jahren von der OECD gepachtet schien, hat sich im Sommer 2009 die Bertelsmann-Stiftung zu Wort gemeldet, deren Feld traditionell eher der Hochschulbereich ist, und die in ihrem Auftrag von Klaus Klemm erarbeitete Studie *Klassenwiederholungen – teuer und unwirksam* als wichtige Nachricht in der deutschen Öffentlichkeit publik gemacht. Der Titel suggeriert, dass die Kosten und die Wirksamkeit von Klassenwiederholungen untersucht worden seien. Tatsächlich stellt die Arbeit den Versuch einer naturgemäß groben und reduktionistischen Kostenberechnung dar, während für die behauptete Unwirksamkeit auf ein paar ältere untaugliche Arbeiten zurückgegriffen wird.

5.1. Wie ermittelt man die Kosten von Klassenwiederholungen?

Dass das Bildungssystem viel billiger wäre, wenn es keine „schlechten" Lernenden gäbe, ist eine Binsenweisheit. In viel größeren Klassen könnte in viel kürzerer Zeit das Nötige gelernt werden. Eigentlich bräuchte man gar keine Klassen (Schulgebäude!) mehr und kaum noch Lehrerinnen und Lehrer, weil die Kinder und Jugendlichen ja bequem zuhause an ihren Rechnern mit perfekter Software (z.B. Logo!) lernen könnten. Die Realität ist aber nun einmal nicht so; es gibt „schlechte" Lernende, und sie verursachen erhebliche Kosten, indem sie den Betrieb aufhalten, zusätzliche Maßnahmen verursachen, usw. bis hin zum frühen Ausscheiden aus dem Dienst einer manchen Lehrperson.

Selbstverständlich können „schlechte" Lernende den Unterricht auch bereichern. Das wären dann Leistungen, die von den Kosten wieder zu subtrahieren wären. Die Kostenstruktur ist jedenfalls viel komplexer, als sie wirkt, wenn man sie auf reale Zahlungen (hier: durch die öffentlichen Haushalte) reduziert. Das hat man in der Volkswirtschaftslehre schon lange erkannt, sieht sich aber mit der Bewertung nicht bezifferter Kosten und Leistungen und deren Zuordnung zu Verursacherinnen und Verursachern immer wieder vor erhebliche Probleme gestellt.

Zum Beispiel müssten in die Berechnungen die (zumindest eine Zeit lang ja vorhandenen) Vorteile der abgebenden Klasse, die der aufnehmenden Klasse (durch die Bereicherung) sowie die der Wiederholerinnen und Wiederholer (W&W) (durch die Chance des Neubeginns usw.) als Leistungen eingehen. — Natürlich weiß man zu wenig über diese Effekte, um sie bewerten zu können. Und genau das spricht überhaupt gegen eine solche Kostenrechnung, wie sie die Bertelsmann-Stiftung in Auftrag gegeben hat.

Wenn man sich aber einmal auf sie einlässt, dann stellt es durchaus eine wichtige Erkenntnis dar, dass die W&W nicht kostenneutral im System mitlaufen, sondern dass Klassenwiederholungen Kosten verursachen. Diese kann man *überschlägig* leicht ermitteln: Einmal Sitzenbleiben verlängert die durchschnittliche Schulzeit von ca. elf auf ca. zwölf Jahre, also um etwa 9%. Im letzten Jahrzehnt blieben vielleicht 30% aller S&S irgendwann einmal sitzen (Mehrfachfälle mehrfach gerechnet). Also wird die Gesamtschulzeit aller S&S durch das Instrument des Sitzenbleibens um 2,7% erhöht. Entsprechend geringer wären also die Personal- (und

verwandte) Kosten des Schulsystems, wenn es dieses Instrument nicht gäbe.

Mit Recht weist Klemm (13) darauf hin, dass diese Rechnung nicht nur in denjenigen Bundesländern so anzustellen ist, in denen die Zuweisung von Lehrerstellen (und entsprechender Mittel) pro S&S-Kopf erfolgt, sondern auch in denjenigen, in denen sie klassenweise erfolgt. Wohl ändert sich in einer Schule oft an der Klassenfrequenz eines Jahrgangs nichts, wenn W&W hinzukommen, aber manchmal, nämlich wenn die Höchstgrenze für die Klassenfrequenz überschritten wird, eben doch, und dann muss gleich eine ganze Klasse zusätzlich eingerichtet werden. Man wird diesen Fall in Zeugniskonferenzen tunlichst verhindern; aber das wird nicht immer gelingen. Jedoch auch wenn keine zusätzlichen Klassen eingerichtet werden müssen, entstehen durch W&W zusätzliche Ausgaben.

Klemm setzt zu Recht in den Bundesländern mit klassenbezogener Lehrerstellenzuweisung pro W&W pauschal nur 50% der Kosten an, die in den Bundesländern mit kopfbezogener Zuweisung pro W&W anfallen, die er bei seinem Vorgehen wiederum sehr genau berechnen kann. – Der Prozentsatz von 50% ist extrem willkürlich. Außerdem liegt hier eine Argumentationslücke vor, indem nämlich nicht berücksichtigt wird (zumindest finde ich in der Arbeit keine Silbe dazu), dass in den abgebenden Klassen die Kosten entsprechend geringer werden. – Letztlich ist das aber egal; denn die ganze Rechnung ist voll von Annahmen, Schätzungen, Setzungen. Auch wenn statt des Werts von 50% für diese Kosten nur 0% angesetzt würde (was bestimmt zu niedrig wäre), käme man noch auf einen Gesamtbetrag von über 700 Millionen Euro.

In *mathematisch idealisierter* Form kann man sich die zahlenmäßige Wirkung des Sitzenbleibens so vorstellen, dass jede aufnehmende Klasse ihrerseits wieder W&W nach unten abgibt, ihr Umfang also (bei konstanter W&W-Quote!) über die gesamte Schulzeit hinweg konstant ist, und sich dann fragen, wo der W&W-Überschuss bleibt. Dieser entsteht im ersten Schuljahr, weil dort die Klassen keine W&W mehr nach unten abgeben können. Die ersten Klassen sind folglich alle um 2,6% größer, als sie wären, wenn das Instrument der Klassenwiederholung nicht existieren würde, bzw. es gibt entsprechend 2,6% mehr erste Klassen. Dieser Überhang zieht sich durch die ganze Schulzeit; jeder Jahrgang hat schließlich 2,6% S&S mehr, und in den Abgangsklassen sind dann jedes

Jahr 2,6% der S&S ein Jahr länger in der Schule gewesen als vorgesehen.

5.2. Worauf stützt sich die Behauptung von der fehlenden Wirksamkeit von Klassenwiederholungen?

Ihre Eignung als Sensationsmeldung für die Presse zieht die Studie einerseits aus dem scheinbar hohen Kostenbetrag von knapp einer Milliarde Euro (der viel weniger eindrucksvoll ist, wenn man ihn als unter 2% der Gesamtausgaben für die allgemein bildenden Schulen angibt) und andererseits in Verbindung damit aus der Behauptung, dass Klassenwiederholungen unwirksam seien.

Diese Behauptung widerspricht den positiven Erfahrungen ganzer Heerscharen von Lehrerinnen und Lehrern (vgl. a. Tietze / Roßbach 1998, 467) mit den auf einmal wieder möglichen Fortschritten in der abgebenden Klasse, mit der Bereicherung der aufnehmenden Klasse (wenigstens für ein Jahr, nicht nur, aber vor allem, in den Fächern, in denen die W&W nicht so schwach waren) und der Chance eines Neubeginns für einen Teil der W&W. Allerdings handelt es sich bei diesen Erfahrungen wieder nur um die Anhäufung von „opinions", die zwar von *den* Expertinnen und Experten für das Lehren und Lernen stammen, aber eben keine statistisch „belastbare" Forschungsergebnisse darstellen.

Wie belegt nun Klaus Klemm die Behauptung der Unwirksamkeit? Da es zu den Wirkungen auf Klassen verständlicherweise keine brauchbaren Untersuchungen gibt, beschränkt er sich auf die Zielgruppe der W&W selbst und gründet seine Behauptung auf eine Handvoll Arbeiten aus der Literatur (7, 2. und 3. Absatz).

Eine davon ist die Habilitationsschrift von Karlheinz Ingenkamp *Zur Problematik der Jahrgangsklasse* von 1967 in der 2. unveränderten Auflage von 1972. Diese bezieht sich auf eine Untersuchung von 1962 an Grundschulkindern im 6. Schuljahr in Berlin-Tempelhof, die wiederum mit einer entsprechenden Untersuchung von 1949 verglichen wird. Ingenkamp bedauert zwar, dass in seinem Buch nicht genug Raum für eine „ideologiekritische Analyse dieses Organisationssystems" (der Jahrgangsklassen) vorhanden ist (ebenda, 29, s.a. 290 u.a.), aber er nimmt über weite Strecken diese Kritik doch vor und plädiert vehement für eine „Integrierte Gesamtschule".

Dieses Plädoyer ist mehrfach widersprüchlich. Der Gegenstand seiner empirischen Forschung, die 6. Klasse, ist ja Teil einer Gesamtschule, nämlich der Berliner Grundschule von 1962, die die ersten sechs Schuljahre umfasste. Seine (negativen) Befunde legen doch gar nicht nahe, die Gesamtschule auf die nächsten Jahrgänge auszudehnen. Mit der Vereinheitlichung der Schulen wiederum fordert er zugleich eine ausgeprägte Differenzierung des Unterrichts, weil ja niemand mehr der Einheitsschule entrinnen kann. Man muss ihm zugute halten, dass er, anders als die heutigen Protagonistinnen und Protagonisten, wenigstens noch die organisatorischen Probleme eines solchen Systems erkennt (ebenda, 301).

Es ist klar, dass Klassenwiederholungen als Kulmination des von ihm kritisierten Systems erst recht in seine Schusslinie geraten. Zur damaligen Zeit wurden ja noch Intelligenztests durchgeführt, und es wundert nicht, dass die W&W bei diesem und bei fachspezifischeren Tests schlechter abschnitten als die „glatt Versetzten", erst recht, wenn die W&W mehrfach wiederholten (ebenda, 106, als Haupt-Argument von Klemm zitiert, und vor allem 275).

Wieso allerdings dieser Umstand gegen die Wirksamkeit von Klassenwiederholungen sprechen soll, erschließt sich mir nicht. Eine *Verringerung* des Leistungsrückstands ist doch auch eine (positive) Wirkung. Dass eine solche vorliegt, ist dermaßen plausibel, dass nicht die Frage, ob, sondern in welchem Ausmaß sie vorhanden ist, interessant gewesen wäre. Dieser Frage ist Ingenkamp allerdings nicht nachgegangen. Die Klassenwiederholung wäre wohl berechtigterweise dann als nicht wirksam zu bezeichnen, wenn die Verringerung des Leistungsrückstands zu schwach ausfiele. Dass der Leistungsrückstand aber dauerhaft auf Null zurückgehen muss, ehe man, konkludent gemäß Ingenkamp und Klemm, von Wirksamkeit sprechen können soll, ist jedoch nicht einzusehen.

Ganz in diesem Sinn spricht die Arbeit von Klemms nächsten Kronzeugen, Belser / Küsel (1976), (entgegen deren Tenor) sogar eher *für* den Erfolg von Klassenwiederholungen. Untersucht wurde die „Schullaufbahn von Volksschulabgängern an 26 [von 313] zufällig ausgewählten Schulen Hamburgs" (ebenda, 103) von 1963 bis 1966, also von Jugendlichen, die etwa 1948 bis 1952 geboren sind. Wohl ist „ganz allgemein zwar im Wiederholerjahr eine Leistungsverbesserung zu beobachten, aber schon im nächsten Schuljahr, in dem neue und höhere Anforderungen gestellt werden, sinken die Leistungen wieder ab" (ebenda, 105, zitiert von

Klemm, 7, als Beleg für die Unwirksamkeit von Klassenwiederholungen). – Wie weit die Leistungen „absinken", ist ein paar Seiten später ausgeführt: „Insgesamt erweisen sich dabei 75% aller zum Zeitpunkt des Sitzenbleibens ungenügenden Zensuren nach 3 Jahren als dauerhaft, mindestens auf ‚ausreichend' verbessert" (ebenda, 111). Die Leistungen sind also vielleicht nicht mehr gut oder befriedigend wie im ersten Jahr, aber eben dauerhaft ausreichend. Dieser Erfolg schlägt sich auch in der Quote der W&W nieder, die den Abschluss erreichen. Während die Hälfte der W&W die Schule mit dem Ende der Schulpflicht, also vor dem Ende der 8. Klasse, verlässt, geht die andere Hälfte ein Jahr länger zur Schule, und davon erreichen 86% (der nur einmal Sitzengebliebenen) den Abschluss. Dieses Faktum wird übrigens nur wenige Zeilen vor dem o.a. scheinbar kritischen Zitat auf Seite 105 mitgeteilt.

Unter den von Klemm zitierten Arbeiten ist (Belser / Küsel 1976) die einzige, die den *langfristigen Ertrag* des Klassenwiederholens kontrolliert. Auch wenn die Autorin und der Autor es nicht wahr haben wollen, hat sich in ihrer Untersuchung das Sitzenbleiben als eine erfolgreiche Maßnahme erwiesen, auch „schwächere" S&S zu einem Abschluss zu führen. – Man darf allerdings nicht außer Acht lassen, dass die Umstände von vor fast einem halben Jahrhundert sich nicht ohne Weiteres auf die heutigen Gegebenheiten übertragen lassen.

Auch beim Lexikon-Artikel von *Tietze / Roßbach* (1998) wird der Charakter des Zitats, das Klemm anführt, erheblich verändert, wenn man es fortsetzt. Trivialerweise schneiden die W&W bei Schulleistungstests nach einem Jahr schlechter ab als diejenigen (gleich-schwachen) S&S, die nicht wiederholen und also eine Klasse höher sind. Aber direkt danach folgt: „Werden Sitzenbleiber jedoch mit (leistungsschwachen) Schülern in der gleichen Klassenstufe verglichen (die nicht-versetzten Schüler sind dann mindestens ein Jahr älter), so zeigen sich (geringe) Leistungsdifferenzen zugunsten der Sitzenbleiber" (ebenda 467). – Na also.

Warum die W&W schlechter abschneiden als Diejenigen, die nicht wiederholen, bringen *Tillmann / Meier* (2001) auf den Punkt: „Zum einen sind Wiederholer im Durchschnitt mit weniger guten kognitiven Voraussetzungen ausgestattet …, zum zweiten wird ihnen aber auch die Befassung mit den anspruchsvolleren fachlichen Inhalten der nächsten Klassenstufe verwehrt" (475). Während bei TIMSS die 9. Schuljahre betrachten werden (mit gewissen Nachteilen), untersucht PISA die 15-Jährigen und handelt

sich dadurch andere Nachteile ein: Zum Beispiel wird, vor allem in Entwicklungsländern, nur eine nicht-repräsentative Stichprobe erfasst, nämlich die der *beschulten* 15-Jährigen. PISA kann zwar feststellen, dass in gewissen Populationen die Quote der W&W größer ist als in anderen oder dass die W&W weniger Punkte erreichen. Um aber Aussagen zur Wirksamkeit der Klassenwiederholung machen zu können, müsste PISA auch die 16-Jährigen untersuchen, und es müssten Kriterien festgelegt werden (vielleicht Punktedifferenzen gegenüber wohl bedachten Bezugspopulationen). Wie bei allen von PISA festgelegten Grenzen (z.B. bei den Kompetenzstufen u.v.a.) wäre das allerdings wieder eine subjektive Angelegenheit, die nur auf „opinions" beruhen würde.

Zusammenfassend lässt sich feststellen, dass die von Klemm herangezogene Literatur, *mit einer Ausnahme*, keine empirisch fundierten Aussagen über die Wirksamkeit von Klassenwiederholungen macht, auch wenn sie, inklusive dieser Ausnahme, sich zu Klassenwiederholungen (mehr oder weniger deutlich) kritisch äußert. Die Ausnahme (Belser / Küsel 1976) hat herausgefunden, dass unter den W&W, die 1965 in Hamburg ein Jahr länger zur Volksschule gingen, 86% den Abschluss schafften. Ob die Autorin und der Autor sich mit ihrer (sachten) Distanzierung vom „Sitzenbleiben" (im Zuge einer recht ausgewogenen Diskussion z.B. auf S. 113) dem Zeitgeist der universitären Pädagogik (nicht nur) der 1970er Jahre anpassten?

Man muss Klemm zugute halten, dass er Auswahl und Interpretation seiner Literatur i.W. komplett dem Diskussionsbeitrag von *Krohne / Tillmann* (Mitarbeiterin und Leiter der Bielefelder Laborschule) (2006) entnommen hat.

6. SCHLUSSBEMERKUNG

Mit PISA kann man messen, welche *PISA-Aufgaben* von wie vielen Jugendlichen gelöst werden. Mit den Zahlenkolonnen kann man allerlei Statistik treiben und dadurch auf manche Tendenz aufmerksam machen. So hat PISA durchaus seinen Nutzen. (Statt „PISA" kann man hier viele andere Projektnamen aus der empirischen Bildungsforschung einsetzen.) Vom Anspruch, mit dem PISA-Quader eine hinter diesen Zahlen stehende umfassende kognitive, soziale und kulturelle Realität abzubilden, ist man jedenfalls weit entfernt, und zwar nicht, weil man noch nicht gut genug ist, sondern weil dieser Anspruch prinzipiell nicht einzulösen ist. Es ist

mir darüber hinaus unbegreiflich, wie man mit Hilfe von PISA-Zahlen die Überlegenheit von Schulsystemtypen beweisen will. Jedenfalls ist die Behauptung, dass die Einheitsschule dem gegliederten Schulsystem überlegen sei, „just another opinion".

LITERATUR

- *Belser, H. / Küsel, G. (1976): Zum Sitzenbleiber-Problem an Volksschulen. In: Biermann, R. (Hrsg.): Schulische Selektion in der Diskussion. Bad Heilbrunn, 101-115.*
- *Bender, P. (2007): Was sagen uns PISA & Co, wenn wir uns auf sie einlassen? In: Jahnke / Meyerhöfer, 281-337.*
- *Ingenkamp, K. (1972): Zur Problematik der Jahrgangsklasse. 2. Auflage. Weinheim.*
- *Jahnke, Th. / Meyerhöfer, W. (Hrsg.) (2007): PISA & Co. Kritik eines Programms. 2. Auflage. Hildesheim / Berlin.*
- *Klemm, K. (2009): Klassenwiederholungen – teuer und unwirksam. Gütersloh.*
- *Krohne, J. / Tillmann, K.-J. (2006): „Sitzenbleiben" – eine tradierte Praxis auf dem Prüfstand. In: SchulVerwaltung Spezial 4/2006, 6-9.*
- *Prenzel, M. / Baumert, J. / Blum, W. / Lehmann R. / Leutner, D. / Neubrand M. / Pekrun, R. / Rost, J. / Schiefele, U. (PISA-Konsortium Deutschland) (Hrsg.) (2005): PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche? Münster u.a. (zit. als PISA 2005).*
- *Prenzel, M., Artelt,C. / Baumert, J./ Blum, W. / Hammann, M. / Klieme, E. / Pekrun, R. (PISA-Konsortium Deutschland) (Hrsg.) (2007): PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie. Münster u.a. (zit. als PISA 2007).*

- *Prenzel, M. / Artelt, C. / Baumert, J. / Blum, W. / Hammann, M. / Klieme, E. / Pekrun, R. (PISA-Konsortium Deutschland) (Hrsg.) (2008): PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich. Münster u.a. (zit. als PISA 2008).*
- *Tietze, W. / Rossbach, H.-G. (1998): Sitzenbleiben. In: Rost, D.H. (Hrsg.): Handwörterbuch Pädagogische Psychologie. Weinheim, 465-469.*
- *Tillmann, K.-J. / Meier, U. (2001): Schule, Familie und Freunde – Erfahrungen von Schülerinnen und Schülern in Deutschland. In: Baumert, J. / Klieme, E. / Neubrand, M. / Prenzel, M. / Schiefele, U. / Schneider, W. / Stanat, P. / Tillmann, K.-J. / Weiß, M. (Deutsches PISA-Konsortium) (Hrsg.): PISA 2000. Basiskompetenzen von Schülerinnen & Schülern im internationalen Vergleich. Opladen, 468-509.*
- *Wartha, S. (2009): Zur Entwicklung des Bruchzahlbegriffs — Didaktische Analyse und empirische Befunde. In: Journal für Mathematik-Didaktik 30, 55-79.*
- *Wuttke, J. (2007): Die Insignifikanz signifikanter Unterschiede. Der Genauigkeitsanspruch von PISA ist illusorisch. In: Jahnke / Meyerhöfer, 99-246.*

<< zurück zum Inhaltsverzeichnis

BEITRAG DER BILDUNGSÖKONOMIE ZUR SICHERUNG DER QUALITÄT VON SCHULE

Manfred Weiß

1. EINLEITUNG

Die von einigen Kritikern (z.B. Becker 2000) schon totgesagte Bildungsökonomie erfährt gegenwärtig einen bemerkenswerten Aufschwung. Das betrifft ihre Position innerhalb der Wirtschaftswissenschaften ebenso wie ihre Funktion in der Politikberatung. Als Gründe dafür lassen sich anführen:

- die verschärften Knappheitsprobleme im öffentlichen Sektor, die auch im Bildungsbereich für eine Sensibilisierung für Effizienzfragen gesorgt haben,
- die wichtiger gewordene strategische Bedeutung von Bildung für die Bewältigung wirtschaftlicher und gesellschaftlicher Probleme,
- die „realistische Wendung" innerhalb der Bildungsökonomie: die thematische Akzentverschiebung von theoretischen Erkenntnisinteressen der Wirtschaftswissenschaften zu praktischen Steuerungsproblemen des politisch-administrativen Systems,
- die verbesserte Institutionalisierung der Bildungsökonomie im Wissenschaftssystem,
- die bildungsökonomisch fundierten Aktivitäten der OECD.

Sichtbarer Ausdruck des Bedeutungszuwachses der Disziplin ist die „Explosion bildungsökonomischer Arbeiten" (Hanushek/Welch 2006, xix). Sie manifestiert sich vor allem in einer beträchtlichen Erweiterung des empirischen Forschungsstandes. Eine wachsende Zahl von Wirtschaftswissenschaftlern hat den Bildungsbereich als Forschungsgegenstand (wieder-)entdeckt. Auch in Deutschland hat diese Entwicklung der Bildungsökonomie Schubkraft verliehen. Dass sie inzwischen international anschlussfähig geworden ist, wird eindrucksvoll durch die Publikationsaktivitäten von Nachwuchswissenschaftlern in hochrangigen englischsprachigen Fachzeitschriften und ihre Kooperation mit prominenten US-amerikanischen Bildungsökonomen belegt. Auch am Aufbau internationaler Netzwerke in der Bildungsökonomie sind deutsche Wirtschaftswissenschaftler maßgeblich beteiligt. Das von der Europäischen Kommission initiierte und finanzierte *European Network on Economics of Education* wird vom Ifo-Institut in München koordiniert. Diesem Netzwerk gehören rund 350 Wissenschaftler an. Eine steigende Mitgliederzahl hat auch der 1975 auf Initiative Friedrich Eddings gegründete „Ausschuss Bildungsökonomie" im Verein für Socialpolitik zu verzeichnen.

Im Zuge ihrer programmatischen Neuausrichtung hat die Bildungsökonomie in letzter Zeit in verstärktem Maße Fragen der *Qualität der Schulbildung* in den Blick genommen. Auslöser dafür ist nicht nur die hohe öffentliche Aufmerksamkeit, die diese Thematik im Gefolge von TIMSS und insbesondere PISA erfahren hat, sondern auch spektakuläre Befunde neuerer bildungsökonomischer Studien zum Wirtschaftswachstum. Sie identifizieren die über Schülerleistungen erfasste Humankapitalqualität als besonders wichtigen Einflussfaktor wirtschaftlichen Wachstums (Hanushek/Kimko 2000; Hanushek/Wößmann 2006).

Die von der Bildungsökonomie zur Qualitätsthematik im Schulbereich bearbeiteten Fragestellungen haben eine größere gemeinsame thematische Schnittmenge mit den Forschungsprogrammen anderer Bildungswissenschaften, insbesondere der Erziehungswissenschaft, entstehen lassen. Ihre Legitimation bezieht die Arbeit verschiedener Disziplinen an einem gemeinsamen Forschungsgegenstand dadurch, dass Theorie- und Methodenwettbewerb eine wichtige Voraussetzung für Erkenntnisfortschritt sind. Welchen spezifischen Forschungsbeitrag leistet dabei die Bildungsökonomie, worin besteht der „Mehrwert" ihrer Mitwirkung?

Bei der Erforschung der *Bedingungsfaktoren von Schulerfolg* gilt das Hauptaugenmerk der Bildungsökonomie zentralen bildungspolitischen Gestaltungsparametern: Ressourcen und Institutionen (im Sinne von verhaltenssteuernden Regelsystemen). In den Forschungsprogrammen der anderen Bildungswissenschaften finden solche „distalen Oberflächenvariablen" nur wenig Beachtung. Die Bildungsökonomie bringt zudem neuartige theoretische Zugangswege in die Bildungsforschung ein, z.B. Erklärungsansätze der *Neuen Institutionenökonomie*[1] bei ihren Analysen institutioneller Einflussfaktoren. Zugleich hat sie deren Methodenarsenal um Verfahren aus der Ökonometrie erweitert. Auf wachsendes Interesse der anderen Bildungswissenschaften stoßen vor allem Verfahren, die die Schätzung kausaler Effekte auch bei Vorliegen von Daten aus Beobachtungsstudien erlauben (vgl. die Übersicht bei Schneider et al. 2007). Alleinstellungsmerkmal der Bildungsökonomie ist die Bereitstellung von Effizienzinformationen durch Zusammenführung von Wirksamkeits- und Kostendaten. Dies qualifiziert sie in besonderer Weise zur Unterstützung von Entscheidungen unter verschärften Knappheitsbedingungen.

Eine Würdigung des Beitrags der Bildungsökonomie zur Schulqualitätsforschung und zur wissenschaftlichen Fundierung bildungspolitischen Steuerungshandelns kann allerdings nicht darüber hinwegsehen, dass ihre Analysen „distaler Oberflächenvariablen" und die diesen Analysen zugrunde liegenden Hypothesen der Spezifik des Untersuchungsgegenstands kaum angemessen Rechnung tragen. Mit anspruchsvollen Analysemethoden – dem „Markenzeichen" der neuen Bildungsökonomie – lässt sich das Manko einer unzureichenden Beachtung der Eigengesetzlichkeiten pädagogischer Prozesse (vage Technologien, starker Einfluss „externer Mitproduzenten" auf das Leistungsergebnis) nicht kompensieren. Hinzu kommt eine ausgeprägte Selbstreferenzialität der Bildungsökonomie. Eine Auseinandersetzung mit konkurrierenden Hypothesen und konträren Forschungsergebnissen anderer Disziplinen findet kaum statt. Ein Grund dafür dürfte u.a. darin zu suchen sein, dass die meisten Wirtschaftswissenschaftler die Bildungsökonomie nicht als Teildisziplin der Bildungsforschung begreifen und vielfach eine institutionelle Basis für eine Kooperation mit anderen Bildungswissenschaften fehlt. In Deutschland wird bildungsökonomische Forschung fast nur noch in wirtschaftswissenschaftlichen Fachbereichen von Universitäten und in Einrichtungen der außeruniversitären Wirtschaftsforschung betrieben. In Bildungsforschungsinstituten wie dem *Max-Planck-Institut für Bildungsforschung* und dem *Deutschen Institut für Internationale Pädagogische Forschung*,

in denen die Bildungsökonomie in der Vergangenheit eine bedeutende Rolle spielte, ist sie nicht mal mehr rudimentär vertreten. Für ein produktives Zusammenwirken verschiedener Disziplinen fehlt in Deutschland derzeit eine institutionelle Basis. In den Empfehlungen des Wissenschaftsrats zum Ausbau der Bildungsforschung (Wissenschaftsrat 2001, 54 ff.) war dies einmal anders vorgesehen.

2. DER FORSCHUNGSANSATZ DER BILDUNGSÖKONOMIE

Bei der empirischen Erforschung der Leistungswirksamkeit von Ressourcen und Institutionen bedient sich die Bildungsökonomie vorrangig des Konzepts der *Bildungsproduktionsfunktion.* Die untersuchten Variablenzusammenhänge werden darin als Input-Output-Beziehungen modelliert; unterstellt wird ein *direkter* Einfluss von Ressourcen und institutionellen Kontextbedingungen auf Lernergebnisse. Theoretisch zu begründen ist indes nur ein indirekter, über „unterrichtsnahe" prozessuale Bedingungsfaktoren vermittelter Einfluss: die angebotenen Lerngelegenheiten, die Qualität der Instruktion und die Nutzung der Lerngelegenheiten durch die Schüler (z.B. Fend 1998). In dem Produktionsfunktions-Ansatz der Bildungsökonomie bleibt dies als „black box" ausgeblendet. Wird z.B. ein Zusammenhang zwischen bestimmten institutionellen Kontextmerkmalen und Schülerleistungen ermittelt, dann wird dies als Ergebnis des verhaltenssteuernden Einflusses von Anreizstrukturen gesehen; das Verhalten der schulischen Akteure selber interessiert nicht. Eine solche Sichtweise kann, wie später noch zu zeigen sein wird, folgenschwere Fehlinterpretationen nach sich ziehen, weil sie nicht zwischen pädagogisch erwünschtem und unerwünschtem Verhalten differenziert. Ein Untersuchungsansatz, der die Mehrebenenstruktur im Schulbereich unberücksichtigt lässt und die Verknüpfung zwischen politischer und unterrichtlicher Handlungsebene ausblendet, ist für die Fundierung politischen Steuerungshandelns nur eingeschränkt tauglich (vgl. dazu auch Raudenbush 2009).

3. UNTERSUCHUNGEN ZUR RESSOURCENWIRKSAMKEIT

Mithilfe des Produktionsfunktionsansatzes ist in der Vergangenheit von der Bildungsökonomie vor allem der Einfluss real vorfindbarer Unterschiede in der finanziellen, personellen und materiellen Ressourcenausstattung von Schulen und Schulsystemen auf Schülerleistungen untersucht worden. Das daraus gezogene Resümee (z.B. Hanushek 1997) fällt für die Bildungspolitik ernüchternd aus: Zwischen Schulressourcen und

Schülerleistungen zeigt sich kein enger und konsistenter Zusammenhang. Dieses Ergebnis problematisiert die weit verbreitete Praxis des „Mehr desselben". Der breiten Öffentlichkeit ist es von den Medien in der vereinfachenden These vermittelt worden, gute Schulen seien keine Frage des Geldes, was die Popularität der Bildungsökonomie in der Erziehungspraxis nicht gerade befördert hat. War der Aussagegehalt älterer Studien zur Ressourcenwirksamkeit noch aufgrund methodischer Mängel stark eingeschränkt, so trifft das nicht mehr für die mit hohen Analysestandards arbeitende neue Generation von Wirkungsstudien zu. Angreifbar bleiben sie dagegen aufgrund des bereits angesprochenen Theoriedefizits: Die Frage, wie Ressourcen wirksam werden (können), beantworten sie nicht. Auch sehen sie sich – wie ganz allgemein die empirische Schuleffektivitätsforschung – dem Einwand ausgesetzt, Schulqualität auf ein enges Spektrum messbarer Bildungsergebnisse zu reduzieren. Die bildungspolitische Relevanz der Studien wird vielfach überschätzt. Was man ihnen attestieren kann, ist eine gewisse Sensibilisierung für die Grenzen einer ausschließlich auf Ressourcenvermehrung fixierten Politik der Qualitätsverbesserung im Schulbereich. Für die Mittelbereitstellung hatte dies bislang keine negativen Folgen: In nahezu sämtlichen OECD-Staaten sind die realen Ausgaben je Schüler über Jahre hinweg gestiegen (OECD 2008, 243). Das könnte sich in Zukunft ändern. Qualitätsverbesserungen werden dann unter Umständen nur noch über eine veränderte faktorielle Ausgabenpolitik, die „Binnenoptimierung" von Ressourcen, möglich sein. Die Bildungsökonomie liefert dazu vereinzelt empirisch fundierte Handlungsempfehlungen (z.B. Levacic et al. 2005). Dem wachsenden Bedarf der Bildungspolitik an solchen Informationen soll in PISA 2009 mit einem eigenen Themenband zur Kosten-Wirksamkeit unterschiedlicher Handlungsoptionen entsprochen werden.

4. UNTERSUCHUNGEN ZUR WIRKSAMKEIT VON INSTITUTIONEN

Die insgesamt wenig ergiebigen Befunde zur Wirksamkeit schulischer Ressourcen sind in der bildungsökonomischen Forschung in der letzten Zeit zum Anlass genommen worden, das Augenmerk stärker auf andere Strategien der Qualitäts- und Effizienzverbesserung zu richten: die als Anreizstrukturen wirkenden *institutionellen Rahmenbedingungen* des Schulsystems. „Aus ökonomischer Sicht versprechen solche institutionellen Rahmenbedingungen den größten Erfolg, die für alle Beteiligten Anreize schaffen, die Lernleistungen der Schüler zu erhöhen: […] Rege-

lungen und Regulierungen des Schulsystems, die explizite oder implizite Belohnungen und Sanktionen für unterschiedliches Verhalten der Akteure erzeugen" (Wößmann 2005a, 19). Als besonders leistungsfördernd gelten Dezentralisierung und Schulautonomie, extern gesetzte Standards und zentrale Abschlussprüfungen sowie Wettbewerbselemente. Die empirische Untersuchung der Wirksamkeit dieser Maßnahmen wird durch den Zugang zu Datensätzen aus internationalen Schulleistungsstudien begünstigt. Sie erfüllen die Voraussetzung einer für das Auffinden von Effekten hinreichenden Varianz der institutionellen Faktoren, wie sie im nationalen Kontext meist nicht gegeben ist. Dieser Vorteil wird freilich mit den bekannten Problemen des internationalen Vergleichs erkauft (vgl. z.B. Klieme/Stanat 2002). Sie manifestieren sich, wie die nachstehende Übersicht verdeutlicht, augenfällig in Unterschieden in den Effektschätzungen für einzelne Variablen, die sowohl die Effektstärke als auch die Effektrichtung betreffen. Im Bemühen um generalisierbare Aussagen berechnete Durchschnittseffekte („Metaeffekte") können sich dann als höchst problematisch erweisen, wenn damit – ungeachtet der jeweiligen Kontextbedingungen – nationale Politikempfehlungen begründet werden.

Schulische Faktoren und ihr Einfluss auf die Lesekompetenz in verschiedenen Ländern (Zuwachs an Punkten bei Veränderung der Prädiktoren um eine Standardabweichung)

Länder	Schulklima[1]	Arbeitshaltung und Stimmung der Lehrer[1]	Lehrerautonomie[2]	Schulautonomie[2]	Schüler-Lehrer-Verhältnis[3]	Schuldisziplin[3]	Leistungsdruck[3]
Australien	8	7	-6	14	8	19	-6
Kanada	4	2	2	18	18	10	-3
Finnland	-3	8	6	-4	11	6	-13
Deutschland	4	1	-11	-8	32	1	0
Schweiz	-8	4	-2	9	11	10	-11
Vereinigtes Königreich	17	5	-2	18	19	15	-8
Meta-Effekt	4	3	-2	4	16	10	3

[1] Einschätzung des Engagements des Lehrkörpers durch Schulleiter
[2] Summierung von Entscheidungsmöglichkeiten
[3] nach Einschätzung der Schüler
Quelle: Fend 2004, Seite 32

Eine institutionelle Rahmenbedingung, über deren Qualitätsrelevanz breiter Konsens besteht, stellt der *Autonomiegrad der Schulen* dar. Die Bildungsökonomie bezieht dazu eine differenziertere Position als die Bildungspolitik, indem sie – gestützt auf Erkenntnisse der *Neuen Institutionenökonomie* – bereits in ihrer theoretischen Argumentation auf eine Ambivalenz hinweist: Größere Handlungsautonomie erlaubt auf der einen Seite die leistungsfördernde Nutzung des in der größeren „Geschehensnähe" liegenden Informationsvorteils der schulischen Akteure; auf der anderen Seite begünstigt sie opportunistisches, von Eigennutzmotiven geleitetes Handeln. Welches Verhalten sich letztlich durchsetzt, hängt einmal von der Bedeutung einzelner Handlungsfelder für die Verfolgung individueller Wohlfahrtsziele ab, zum anderen von den jeweiligen verhaltenssteuernden institutionellen Rahmenbedingungen. Auswertungen des internationalen Datensatzes aus PISA 2000 durch Wößmann (zusammenfassend 2007) verweisen auf die besondere Bedeutung *externer Abschlussprüfungen*. Positive Autonomieeffekte zeigen sich nur in Verbindung mit solchen Prüfungen. Fehlt diese Bedingung, dann geht ein hoher Autonomiegrad meist mit niedrigeren Schülerleistungen einher. Wößmann sieht darin die These bestätigt, dass die schulischen Akteure ihre Autonomie nur dann zur Leistungsförderung der Schüler statt zum eigenen Vorteil nutzen, wenn die Schulen durch externe Leistungsprüfungen zur Rechenschaft gezogen werden. Deshalb sollte eine effiziente Bildungspolitik „Zentralprüfungen mit Schulautonomie verbinden, sie sollte Standards extern vorgeben und überprüfen *und es gleichzeitig den Schulen überlassen, wie sie diese Standards erreichen wollen"* (Wößmann 2005b, 166, Hervorhebung M. W.). Dieser „Vertrauensvorschuss" kontrastiert auffällig mit der die Lehrerschaft unter Generalverdacht stellenden Opportunismusthese. Übersehen wird, dass das Verhaltensrepertoire der schulischen Akteure weit größer ist, als einfache ökonomische Erklärungsansätze unterstellen, und dass sich darunter Verhaltensreaktionen befinden, die unerwünschte Wirkungen hervorbringen. Durch die internationale Bildungsforschung sind solche Wirkungen hinreichend dokumentiert (vgl. zusammenfassend Bellmann/Weiß 2009). In der bildungsökonomischen Forschung wird ihnen bislang kaum Beachtung geschenkt. Dieses Manko verweist wiederum auf theoretisch- konzeptionelle Grenzen des Produktionsfunktionsansatzes. Die auf dieser Basis gewonnenen Informationen aus statistischen Zusammenhangsanalysen reichen für die Formulierung bildungspolitischer Empfehlungen nicht aus. Benötigt werden Informationen darüber, wie externe Steuerungsvorgaben von den Schulen verarbeitet werden, welche Faktoren zum Gelingen oder Misslingen erweiterter Handlungskompetenz auf Schulebene beitragen. Dazu sind die Forschungsprogramme der anderen Bildungswissenschaften zu befragen. Und darin finden sich kaum Hinweise auf substanzielle Autonomieerträge im Leistungsbereich (Schümer/Weiß 2008).

Wie von Dezentralisierung und Autonomie, so werden auch von *Wettbewerb* nachhaltige Qualitäts- und Effizienzverbesserungen im Schulbereich erwartet. In einer ganzen Reihe von Ländern, insbesondere im angelsächsischen Raum, wurden Steuerungssysteme im Bildungsbereich etabliert, die Wettbewerbselemente als konstitutiven Bestandteil beinhalten: Schulwahlfreiheit, die Stärkung der Konkurrenz durch private Bildungsangebote und Formen nachfrageorientierter Finanzierung der Schulen (Pro-Kopf-Zuweisungen, Bildungsgutscheine). Für vieler Ökonomen steht die Wirksamkeit von Wettbewerb im Schulbereich außer Frage: „Die Nutzen stiftenden Wirkungen von Wettbewerb sind in anderen Handlungsfeldern so gut dokumentiert, dass es kaum vorstellbar ist, mehr Wettbewerb sei für Schulen nicht vorteilhaft" (Hanushek/Wößmann 2007, 70). Die empirische Evidenz fällt indes weit weniger eindeutig aus, als überzeugte Wettbewerbsapologeten mit ihrer selektiven Forschungsauswahl glauben machen wollen.

Die bislang wohl umfangreichste Dokumentation von US-amerikanischen Studien zur Effizienzwirksamkeit von Wettbewerb im Schulbereich stammt von Belfield und Levin (2002). Die von ihnen vorgenommene statistische Auswertung der in diesen Studien berichteten Effektschätzungen („Meta-Analyse") zeigt insgesamt einen positiven, aber geringen Effekt von Wettbewerb auf Schülerleistungen. Bis zu zwei Drittel der in den Einzelstudien berichteten Effektschätzungen sind nicht signifikant. Auch die aus anderen Ländern vorliegenden Forschungsergebnisse fallen widersprüchlich aus und legen eine eher zurückhaltende Einschätzung des leistungsfördernden Potenzials von Wettbewerb im Schulbereich nahe (Oelkers 2007; Weiß 2009). Die Erwartungen an Wettbewerb dämpft auch die Auswertung der Daten aus PISA 2006 durch die OECD (2007): Ein positiver Effekt konkurrierender Schulen auf die Leistungen der 15-Jährigen lässt sich nicht belegen (OECD 2007, 309). Wettbewerb ist offensichtlich kein uneingeschränkt einsetzbares Universalmodell. Diese Schlussfolgerung drängt sich noch mehr auf, wenn die für andere Qualitätsdimensionen empirisch hinreichend belegten Dysfunktionalitäten der Wettbewerbssteuerung im Schulbereich berücksichtigt werden: zunehmende Transaktionskosten und Leistungsdisparitäten sowie eine Verstärkung sozialer Segregation. Immerhin machen zum Teil auch bildungsökonomische Studien auf diese Effekte aufmerksam (z.B. Andersen/Serritzlew 2007; Böhlmark/Lindahl 2007).

Die bescheidene Erfolgsbilanz der Wettbewerbssteuerung im Schulbereich nährt Zweifel an der Belastbarkeit der dem Wettbewerbsmodell zugrunde liegenden Verhaltensprämissen, die einem deterministischen Verständnis von Anreizstrukturen folgen. Das Damoklesschwert „Klientenverlust", so die zentrale Modellannahme, sorge für eine besondere Anstrengungsbereitschaft, um mit hohen Leistungsstandards im Wettbewerb bestehen zu können. Wettbewerb diszipliniere, indem er die Schulen zwinge, effizient zu arbeiten und leistungssteigernde Innovationen einzuführen. Die Empirie versagt auch hier den Modellannahmen weitgehend die Gefolgschaft (vgl. dazu ausführlich Weiß 2009). Die Bandbreite dokumentierter anbieterseitiger „Verhaltensanomalien" reicht von Selektionsstrategien, die die Schulen gezielt zur „Optimierung der Schülerpopulation" entwickeln, über die Umverteilung von Unterrichtszeit zugunsten von Testfächern und getesteten Inhalten bis hin zu Manipulationen von Testergebnissen. Solche Verhaltensreaktionen sind ebenso wenig intendiert wie von Wettbewerb erzwungenes Handeln, das mit dem professionellen Selbstverständnis von Lehrern unvereinbar ist. Wettbewerb zwingt sie in die Rolle eines Dienstleisters des Elternwillens.

Auch die Nachfrager wollen sich in der Realität nicht so verhalten, wie es das theoretische Wettbewerbsmodell vorsieht: dass Leistungsunterschiede zwischen Schulen massive Wanderungsbewegungen auslösen. Zu den empirisch vielfach belegten Auffälligkeiten zählt etwa, dass viele Eltern von ihrem Recht auf freie Schulwahl häufig keinen Gebrauch machen, dass sie sich bei ihren Entscheidungen oftmals von pragmatischen Gesichtspunkten (z.B. der Wohnortnähe) leiten lassen, schlechte Leistungswerte von Schulen ihre Wechselbereitschaft nicht unbedingt beflügeln und bisweilen von ihnen gezielt „inferiore" Anbieter gewählt werden, um sonst nicht erreichbare Berechtigungen für ihre Kinder zu erlangen, eine Option, die im Wettbewerbsmodell überhaupt nicht vorgesehen ist.

5. RESÜMEE

„Die Bildungsökonomie lebt, und weil sie lebt, ändert sie sich". Mit diesen Worten kommentierte Friedrich Edding, der Nestor der Bildungsökonomie im deutschsprachigen Raum, Anfang der 1970er-Jahre Entwicklungen in der sich thematisch ausdifferenzierenden Disziplin. Besser denn je beschreiben diese Worte aktuelle Entwicklungen in der Bildungsökonomie. Sie sind durch verschiedene Trends charakterisiert: eine Stärkung der

Position der Disziplin in den Wirtschaftswissenschaften und in der Politikberatung, einen Bedeutungszuwachs von Steuerungsproblemen des politisch-administrativen Systems im Forschungsprogramm, ein wachsendes Interesse an den Wirkungen und Determinanten der Qualität der Schulbildung, eine verstärkte empirische Forschungsorientierung mit hohen Analysestandards und die intensive Nutzung von Datensätzen aus internationalen Schulleistungsstudien. Innerhalb der Qualitätsthematik hat sich in den letzten Jahren eine deutliche Verlagerung des bildungsökonomischen Forschungsinteresses von Ressourcen zu Institutionen vollzogen. Ihnen wird eine Überlegenheit gegenüber ressourcenbezogenen Strategien der Qualitätsverbesserung attestiert. Die empirische Basis dafür liefern „Produktionsfunktionsschätzungen". Die ihnen zugrunde liegende Input-Output-Logik erlaubt nur eine unterkomplexe Erfassung von Wirkungszusammenhängen. Die Belastbarkeit der Forschungsergebnisse ist dadurch eingeschränkt. Die Beschäftigung der Bildungsökonomie mit Fragen der Bildungsqualität im Schulbereich hat eine größere gemeinsame thematische Schnittmenge mit den Forschungsprogrammen anderer Bildungswissenschaften entstehen lassen. Doch bleiben Synergiepotenziale wegen der vorherrschenden disziplinären Forschungsorganisation weitgehend ungenutzt. Mit der Rückkehr der Bildungsökonomie in einen kooperativen Forschungsverbund mit anderen Bildungswissenschaften wird die berechtigte Erwartung einer nachhaltigen Steigerung der wissenschaftlichen und politischen Relevanz ihrer Forschungen zur Qualität der Schulbildung verknüpft. Derzeit finden vor allem ihre empirischen Arbeiten zur Funktionalität von Bildung für die wirtschaftliche Entwicklung von Volkswirtschaften hohe Aufmerksamkeit. Damit sorgt die Bildungsökonomie für ein Meinungsklima, das sich positiv auf die Ressourcenmobilisierungsfähigkeit des Bildungswesens auswirkt. Immerhin – das sollten auch ihre Kritiker konzedieren – leistet sie damit vielleicht einen wichtigen indirekten Beitrag zur Sicherung der Qualität von Schule.

1| *Die Neue Institutionenökonomie beschäftigt sich mit den Auswirkungen von Institutionen (z. B. Verfügungsrechten, Verträgen, Organisationsstrukturen, Märkten) auf menschliches Verhalten. Sie untersucht insbesondere Möglichkeiten der effizienten Gestaltung von Institutionen. Die der Neuen Institutionenökonomie zurechenbaren Ansätze (Theorie der Verfügungsrechte, Transaktionskostentheorie und Principal-Agent-Theorie) sind durch weitgehend übereinstimmende Annahmen zum menschlichen Verhalten gekennzeichnet: individuelle Nutzenmaximierung und begrenzte Rationalität des Handelns (vgl. Picot / Dietl / Franck 1999, 54 ff.).*

LITERATUR

- *Andersen, S.C. / Serritzlew, S. (2007): The unintended effects of private school competition. In: Journal of Public Administration Research and Theory 17 (2), 335-356.*
- *Becker, E. (2000): Von der Zukunftsinvestition zur Effektivitätskontrolle des Bildungssystems. In: Radtke, F.-O. / Weiß, M. (Hg.): Schulautonomie, Wohlfahrtsstaat und Chancengleichheit. Opladen, 95-114.*
- *Belfield, C. / Levin, H. M. (2002): The effects of competition between schools on educational outcomes: A review for the United States. In: Review of Educational Research 72 (2), 279-341.*
- *Bellman, J. / Weiß, M. (2009): Risiken und Nebenwirkungen Neuer Steuerung im Schulsystem. In: Zeitschrift für Pädagogik 55, H2, 286-308.*
- *Böhlmark, A. / Lindahl, M. (2007): The impact of school choice on pupil achievement, segregation and costs: Swedish evidence. Bonn (IZA Discussion Paper 2786).*
- *Fend, H. (1998): Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lernleistung. Weinheim.*
- *Fend, H. (2004): Was stimmt mit den deutschen Bildungssystemen nicht? Wege zur Erklärung von Leistungsunterschieden zwischen Bildungssystemen. In: Schümer, G. / Tillmann, K.-J. / Weiß, M. (Hrsg.): Die Institution Schule und die Lebenswelt der Schüler. Wiesbaden, 15-38.*
- *Hanushek, E. A. (1997): Assessing the effects of school resources on student performance. An update. In: Educational Evaluation and Policy Analysis 19(2), 141-164.*
- *Hanushek, E. A. / Kimko, D. D. (2000): Schooling, labor force quality, and the growth of nations. In: American Economic Review 90 (5), 1184-1208.*

- *Hanushek, E. A. / Welch, F. (Eds.) (2006): Handbook of the economics of education. 2 Vols. Amsterdam.*
- *Hanushek, E. A. / Wößmann, L. (2007): The role of education quality in economic growth. Washington (World Bank Policy Research Working Paper 4122).*
- *Klieme, E. / Stanat, P. (2002): Zur Aussagekraft internationaler Schulleistungsvergleiche – Befunde und Erklärungsansätze am Beispiel von PISA. In: Bildung und Erziehung, 55, H. 1, 25-45.*
- *Levacic, R. et al. (2005): Estimating the relationship between school resources and pupil attainment at key stage 3. London (Institute of Education).*
- *OECD (2007): PISA 2006 – Naturwissenschaftliche Kompetenzen für die Welt von morgen. Paris.*
- *OECD (2008): Bildung auf einen Blick 2008. Paris.*
- *Oelkers, J. (2007): Expertise Bildungsgutscheine und freie Schulwahl. Bern.*
- *Picot, A. / Dietl, H. / Franck, E. (1999): Organisation. Eine ökonomische Perspektive. 2. Auflage. Stuttgart.*
- *Raudenbush, S.W. (2009): The Brown legacy and the O`Connor challenge: Transforming schools in the images of children`s potential. In: Educational Researcher 38 (3), 169-180.*
- *Schneider, B. et al. (2007): Estimating causal effects. Washington (AERA).*
- *Schümer, G. / Weiß. M (2008): Bildungsökonomie und Qualität der Schulbildung. Kommentar zur bildungsökonomischen Auswertung von Daten aus internationalen Schulleistungsstudien. Frankfurt am Main (Max-Traeger-Stiftung).*
- *Weiß, M. (2009): Schule und Wettbewerb. In: SchulVerwaltung Hessen u. Rheinland-Pfalz, 14, H. 2, 34-36 u. H. 3, 69-71.*

- *Wissenschaftsrat (Hg.) (2001): Empfehlungen zur künftigen Struktur der Lehrerbildung. Köln.*
- *Wößmann, L. (2005a): Leistungsfördernde Anreize für das Schulsystem. In: ifo Schnelldienst 58, Nr. 19, 18-27.*
- *Wößmann, L. (2005b): Ursachenkomplexe der PISA-Ergebnisse: Untersuchungen auf Basis internationaler Mikrodaten. In: Tertium Comparationis. Journal für International und Interkulturell Vergleichende Erziehungswissenschaft Vol. 11(2), 152-176.*
- *Wößmann, L. (2007): Extern überprüfte Standards, Schulautonomie und Wettbewerb: Chancen für das deutsche Schulsystem. In: Recht der Jugend und des Bildungswesens 55, H. 1, 64-83.*

<< zurück zum Inhaltsverzeichnis

PÄDAGOGISCHE EMPIRIE AUS BILDUNGSPHILOSOPHISCHER SICHT

Volker Ladenthin

Von hoher Aktualität und politischer Beliebtheit ist derzeit die *empirische* Bildungsforschung – spätestens seit TIMSS und PISA. Ich möchte diese Entwicklung[1] aus Sicht der Bildungs*philosophie* der Pädagogik kommentieren. Nun könnte diese Themenstellung zur Vermutung verleiten, als beschäftige sich die Bildungs*philosophie* mit den *philosophischen* Problemen der Pädagogik – während sich die anderen Teilgebiete der Pädagogik mit den *wirklichen* Problemen beschäftigen. Eine solche Annahme ist unbegründet – schon vom Begriff her.

1. VOM PRAKTISCHEN NUTZEN DER BILDUNGSPHILOSOPHIE

Die Aufgabe der *Bildungs*philosophie kann nur sein, analog zu der Aufgabenbestimmung der Philosophie über Bildung nachzudenken. Um nun nicht in den historischen Diskurs darüber zu geraten, was denn nun die Aufgabe der Philosophie sei, beziehe ich mich auf eine Position, die das bestimmt, was Philosophie zumindest *auch* ist:

Als Aufgabe der Philosophie hat Kant die Beantwortung der Fragen bestimmt: Was kann ich wissen? Was soll ich tun? Und: Was darf ich hoffen?

Ganz analog kann man die Aufgaben der Bildungsphilosophie mit diesen drei Fragen bestimmen, mit drei Fragen, die nicht zu umgehen sind: Was kann ich pädagogisch wissen? Wie gelange ich zu handlungsrelevanten pädagogischen Normen? Was sind die Voraussetzungen und der Sinn des pädagogischen Wissens und Planens?

Diese drei Fragen gehören zusammen. Die Antworten auf diese drei Fragen bedingen einander. Das hängt mit einer Besonderheit der Pädagogik zusammen: Pädagogik ist immer Praxis. Alles Forschen in der Pädagogik (oder Erziehungswissenschaft)[2] zielt letztendlich auf die Frage nach der individuell oder in sozialen Systemen gelingenden Bildungspraxis in Familie und Schule. (Wenn ein Teil der Erziehungswissenschaft sich dieser Fragestellung entzieht – dann sind damit nicht die Fragen suspendiert; sie werden nur von dieser Art Erziehungswissenschaft nicht bearbeitet. Dazu unten mehr.)

Erziehungswissenschaft – und damit auch Bildungsphilosophie – ist letztlich eine praktische Wissenschaft. Es ist die Praxis, in der aus jemandem, wie er zufällig ist, jemand wird, der er sein *soll*. Diese Praxis ist keine großmütige Umsetzung einer bloßen Idee, kein psychologisches Konstrukt, kein ökonomischer Wettbewerbsvorteil oder gar nur ein Projekt der Moderne. Die pädagogische Praxis ist vielmehr dem Menschen von Beginn seiner Bewusstwerdung an aufgegeben – diese Praxis ist die Antwort auf die *Bildsamkeit* des Menschen. Der Mensch ist von Natur aus nicht festgelegt – und so muss er sich fragen, wie er sich festlegen *soll*. Pädagogik hat es immer mit Sollenssätzen zu tun: Pädagogen sollen etwas tun oder nicht tun; die Adressaten ihres Bemühens sollen etwas lernen oder nicht lernen.

In der Pädagogik geht es immer um dieses Sollen – deshalb ist sie auch *Praxis* im aristotelischen Sinne.[3] In dieser speziellen, d.h. von anderen Praxen zu unterscheidenden Praxis geht es um die Bildung des Menschen. Bildung zielt auf den ganzen Menschen, auf seine Totalität. So ist zu formulieren, dass „alle pädagogischen Maßnahmen (...) erst ihre Begründung von einer Auffassung der Bildung (erhalten), die als *letzter Bezugspunkt* pädagogischen Tuns diese Maßnahmen als sinnvoll für das Leben des Menschen auszuweisen hat."[4] Bildung ist so – nun in alternativen Formulierungen - der Vorgang der Personalisation der Person, der Menschwerdung des Menschen, des Vernünftigwerdens des Verstands, letztlich die Selbstbestimmung des Menschen.[5] Eine solche Begriffsfestlegung schließt andere Humanwissenschaften – Biologie, Soziologie, Psychologie usw. – gerade nicht aus der Bestimmung dessen aus, was der Mensch ist und was er tun soll, ordnet sie aber im Hinblick auf den Anspruch gelingenden Lebens als Aspekte der umfassenderen Bildungsaufgabe unter. Die Frage der Menschwerdung des Menschen kann demnach weder ausschließlich durch die Biologie (Gesundheit, Physiologie), durch die Soziologie (Sozialisation), durch die Psychologie (Kompetenzen) usw. beantwortet werden, noch kann umgekehrt die pädagogische Frage ohne die Beachtung des in den anderen Wissenschaften – zum Beispiel – der Biologie, der Soziologie, der Psychologie usw. – bereitgestellten Wissens zu beantworten versucht werden.

Mit „Bildung" ist gemeint, was zwar in verschiedenen Kulturen mit unterschiedlichen Worten belegt wird, aber immer gleich konstituiert ist:[6] die anthropologische Notwendigkeit, erst noch lernen zu müssen, eigenständig angemessen zu Handeln. Diese Fähigkeit, „auf eigenen Beinen zu stehen", ist durchaus unterschiedlich bezeichnet: Mündigkeit, Bildung, die Griechen nannten es „phrónesis". Diese Fähigkeit musste aber und muss zu allen Zeiten in allen Kulturen herausgebildet werden. Bildung ist universal – wie die Aufgabe sittlich zu handeln oder sich um die Gesellschaft zu sorgen. Keine Kultur, in der dies nicht stattfände. Das Ziel allen pädagogischen Forschens und Handelns war und ist die Bildung des Menschen. Alles, was nicht der Bildung dient, gehört nicht in die Pädagogik. Da sollte man sehr streng sein.

2. PÄDAGOGIK – WAS IST DAS?

Die Pädagogik betrifft jene Handlungen, die die Selbstwerdung eines anderen Menschen praktisch ermöglichen (man nennt dies: „konstitutive Prinzipien"). *Wie* man pädagogisch handeln soll, wird durch „regulative Prinzipien" bestimmt.[7]

Mit dem Begriff der Praxis ist der Begriff des Handelns verbunden: Das pädagogisch begründete Sollen richtet sich ja nicht auf Gedankenspiele, sondern auf Handlungen, die jemand – sei es der Pädagoge, sei es der Adressat – ausführen soll. Die Fragen der Bildungsphilosophie lauten demnach: Was kann ich über diese pädagogische Praxis wissen? Wie soll ich diese pädagogische Praxis gestalten? Welchen Sinn hat diese Praxis?

Eine der Voraussetzungen dieser Praxis ist, dass das Gegenüber in pädagogischen Prozessen nie Objekt sein darf (wie der Patient, der operiert wird), sondern bereits Subjekt ist. Die pädagogischen Interaktionspartner haben die *gleiche Würde* als Menschen. Sie haben beide einen eigenen, nur *in sich gründenden Willen* – die Philosophen nennen dies: Sie haben beide *Selbstbewusstsein*. Und beide sind ausschließlich verbunden über etwas, was beiden (und eigentlich allen) Menschen gemeinsam ist: *Die Vernunft*.

Etwas gilt ja nicht, weil es der Erzieher sagt, sondern es gilt, weil es richtig ist. Und zwar richtig sowohl für den Bildner als auch für den zu Bildenden. Und dieses gleichermaßen für alle Richtige ist in der Moderne, in unserer nachmetaphysischen Zeit, *ausschließlich* durch Verstand und Vernunft bestimmt. Auf sie hoffen wir. Was sich nicht vernünftig rechtfertigen lässt, darf auch in Bildungsprozessen nicht gelten. Man darf es nicht lehren. Man kann es auch nicht lehren – wenn denn lehren heißt, jemandem zur Einsicht verhelfen. Erzieher und Zögling sind also durch wechselseitige Einsicht verbunden – Einsicht kann man aber weder *bewirken* noch *erzwingen*. Man kann sie nur durch geduldige Argumentation adressatengemäß evident machen.

Die in der pädagogischen Praxis Involvierten (Erzieher und Zögling; Lehrer und Schüler, Eltern und Kind) stehen also nicht wie Ursache und Wirkung zueinander – wie etwa in der Medizin – sondern eben in einem besonderen und *nur* in der Pädagogik thematisiertem Verhältnis. *Dieses spezifische pädagogische Verhältnis besteht darin, dass der eine den anderen zu etwas anleiten will, was dieser nur selbst tun kann.* Wer je mit einem Kind Klavier oder Gitarre oder mit einem älteren Erwachsenen das Tanzen geübt hat, weiß, wie schwierig sich dieser theoretisch einfache Satz in Praxis umsetzen lässt. Und weil es so schwierig ist, dieses pädagogische Paradox auszuhalten, verfällt das *unprofessionelle* Denken häufig auf zwei Lösungen: Die Allmacht und die Ohnmacht.

Die *eine falsche Lösung* gibt dem Pädagogen alles Recht – und sieht den Adressaten letztlich als *Objekt* an, als ein Wesen, auf das man *einwirken* kann und bei dem *gleiche Ursachen auch gleiche Wirkungen* haben. Man *misst* die Güte der Einwirkungen an den sichtbaren Auswirkungen. Nur: Damit stellt sich dieser Ansatz selbst außerhalb des pädagogischen Diskurses, weil dieser nicht nach Wirkungen fragt, sondern nach Einsicht. Ob jemand aus Einsicht handelt, oder weil man ihn nötigt, kann man

aber nicht messen. Das heißt: Gerade das pädagogische Handlungsziel, Handeln *aus Einsicht*, ist nicht messbar. Es ist nicht einmal beobachtbar. Es ist eine *Einstellung* zum eigenen Handeln. Eine Haltung. Es ist die gültige Selbstbestimmung, die auszulösen sich die Pädagogik müht.

Ob jemand aber etwas einsieht, hängt davon ab, ob er etwas einsehen *will*. Diesen Willen darf man nicht erzwingen oder bezwingen – er macht nämlich die Würde des Menschen aus.[8] Ohne den vorausgesetzten eigenen Willen des anderen gäbe es keine Bildsamkeit, bräuchten wir nicht zu bilden, zu unterrichten, zu erziehen. Ohne den vorausgesetzten Willen des Anderen könnten wir sofort mit der Dressur beginnen – so wie man es ja auch mit Tieren macht, deren Willen man bei der Dressur nicht achtet oder sogar „bricht". *Outputdefinierte Handlungssteuerung* missachtet also das pädagogische Grundverhältnis, das auf Autonomie, nicht auf Output zielt. Es ist die eine Verkürzung des pädagogischen Paradoxes.

Die *andere Verkürzung des pädagogischen Paradoxes* gibt dem Zögling alles Recht: Kinder an die Macht, Jahrhundert des Kindes, Antipädagogik. In den so genannten „Selbstlernzentren" und der Umwidmung des Lehrers zum „Arrangeur" ist diese Fehlform noch aktuell. Ein sich Selbstüberlassen des zu Erziehenden missachtet nämlich ebenfalls das *pädagogische* Grundverhältnis, das auf Geltung, nicht auf Beliebigkeit zielt.

Beide Verkürzungen haben den Vorzug, einfach zu sein. Sie sind für den Laien unmittelbar verständlich. Sie sind phrasentauglich. Sie beflügeln zudem *Allmachtsphantasien* oder *entlasten* mit Hinweis auf Selbstverwirklichung von jeder *Verantwortung*. Nur: Pädagogisch sind beide Auffassungen nicht. Angemessenes pädagogisches Handeln setzt ein paradoxes Verhältnis voraus: Jemanden verbindlich zu etwas aufzufordern, was dieser nur selbst tun kann.

- Ein Gedicht interpretiert ja weder der Schüler richtig, der die Interpretation eines Lehrers nacherzählt und vorgefertigte Antworten ankreuzt, noch der, der zu einem Text sagt, was ihm gerade einfällt. Ein Gedicht interpretiert nur der Schüler angemessen, der sein Verständnis des Textes eigenständig unter dem Anspruch von Geltung zur Sprache bringt (– das meint der Terminus „verstehen").

- Moralisch zu handeln hat nicht der gelernt, der so handelt wie ein zufälliges oder erzwungenes Vorbild in seiner Umgebung. Gleiches gilt für den, der seinen Willen schon für ein moralisches Gebot hält. Moralisch handelt nur derjenige, der selbst Verantwortung wahrnehmen kann.

Diese Selbstbestimmung *kausal* bewirken zu wollen ist ein Widerspruch in sich selbst. *Pädagogisch* aber löst sich dieser Widerspruch im pädagogischen Paradox: Einsicht durch Dialog, *dia-logos*, der Gang durch die allen Menschen gemeinsame Vernunft. Einsicht durch vernünftigen Unterricht.

Wirkung nennt man, was ohne Zutun des Subjekts erfolgt. *Einsicht* ist Ergebnis eines Bemühens des Subjekts. Wenn also empirische Studien mit Kausalitätsvorstellungen arbeiten (Maßnahme → Wirkung), befinden sie sich außerhalb des pädagogischen Diskurses.

Das Ziel der wissenschaftlichen *Pädagogik* ist die Forschung zur gültigen Gestaltung einer genuin *pädagogischen* Praxis: Wie soll man *als Pädagoge* so handeln, dass der andere sich gültig selbst bestimmen lernt? Hierfür benennt die Pädagogik die regulativen Prinzipien. Es geht also in der Pädagogik immer um ein *Sollen*. Die pädagogische Praxis gestaltet sich durch Handeln; um handeln zu können, muss man die Wirklichkeit kennen, in der man handelt. Von daher richtet sich pädagogisches Denken von Beginn an nicht nur auf die Frage nach den Handlungsprinzipien sondern zugleich (1) auf die Frage nach den Handlungsvoraussetzungen und (2) immer auch auf die Geschicklichkeit, beides miteinander zu verbinden. Empirie ist *keine* Erfindung der Gegenwart. Im 7. Buch der Politik fragt Aristoteles nach dem richtigen Umgang mit den Kindern in der Polis. Und er schreibt: „Das springt ja in die Augen: Denn Gefühl, Willen und Begierde hat schon das neugeborene Kind….".[9] „Das springt ja in die Augen", eine treffliche Umschreibung der Empirie. Und natürlich hat Aristoteles den attischen Staat im Blick, wenn er seine Prinzipien entwickelt. Aber er entwickelt diese Prinzipien nicht *aus* der Empirie, sondern *für* die Wirklichkeit. Deswegen handelt er die Pädagogik in der „Politik" und der „Ethik" ab – und nicht in der Biologie. (Allerdings handelt Aristoteles die pädagogischen Fragen eben nicht in einem eigenen Buch ab – erst der Moderne gelingt die systematische Unterscheidung der Pädagogik von der Politik und der Ethik.)[10]

Alle Pädagogik ist *Praxis*, also die Vermittlung von universalen Prinzipien mit regionaler Wirklichkeitserfahrung. Selbst in einer arbeitsteiligen Gesellschaft kann dieser Sach-Zusammenhang nicht aufgehoben werden. Wer praktisch handelt, kann *Prinzipien* von *Fakten* zwar unterscheiden, aber nicht trennen; er muss beides in der *Handlungsnorm* verbinden. Für die pädagogische Forschung heißt dies: Ob eine empirische Untersuchung pädagogisch bedeutsam ist, lässt sich nicht aus ihr selbst ableiten, sondern bedarf immer erst des Bezugs zu den pädagogischen Prinzipien. Nur ein Denken, das Wirklichkeitswahrnehmung (also Empirie), prinzipiengeleitetes Denken und pädagogische Urteilskraft verbindet, ist pädagogisch. Handeln kann man nur in einer Wirklichkeit, die man zuvor erkannt hat. Sonst handelt man nicht, sondern tappst herum. Das war immer so.

3. DIE EMPIRISCHE WENDUNG?

Was sich historisch verändert hat, sind die Methoden der Wirklichkeitswahrnehmung. Zumindest im Bereich der Wissenschaft. Da kommt es zu einer Verfeinerung von Methoden; da kommt zur (vorherrschend) qualitativen die quantitative Empirie mit delikaten Methoden, z.B. der Statistik. Und darüber kann man sich, wie über jede Verfeinerung, nur freuen. Auf ein Problem, das die immer aufwendigeren Studien hervorrufen, will ich am Schluss hinweisen.

Die Empirie, also die Wirklichkeitswahrnehmung, wird in der Pädagogik an zwei Stellen thematisiert, die man unterscheiden muss.

3.1 Alltägliche Erfahrung

Zuallererst ist es die Wirklichkeitserfahrung in der Praxis. Der Erzieher muss sein prinzipielles Wissen in einer konkreten Situation anwenden – und dazu muss er die Situation erfassen: Was weiß der Adressat schon, was soll er wissen? Wie kann er mich verstehen? Wie viele Adressaten habe ich? Wie lange darf ich sprechen? Was kann man den Adressaten zumuten?

Der Pädagoge muss diese Wirklichkeit „mit allen Sinnen" aufnehmen; er muss sehen, er muss hören, was hinter seinem Rücken vor sich geht, ja, er muss sich sogar in den Schüler hineinversetzen – um dann adäquat auf die geplanten Ziele hin handeln zu können. Dies ist eine hermeneu-

tische Fähigkeit, die zudem Wissen über Handlungsprinzipien voraussetzt – und nun beides in Verbindung zu bringen weiß. Kant hat diese besondere Fähigkeit „Urteilskraft" genannt – und seine dritte Kritik beschäftigt sich mit dieser menschlichen Grundfähigkeit. Sie überdauert alle Zeiten.

Der Pädagoge vor Ort muss seine pädagogische Wahrnehmungsfähigkeit professionell schulen. Er muss die Handlungsprinzipien kennen. Anders als ein Fluglotse, der sich an feste Vorschriften halten kann, muss der Lehrende allerdings die Handlungs*normen* situativ selbst finden. Er kann sich nur an Prinzipien halten. Er braucht nicht *bestimmende Urteilskraft*, wie ein Fluglotse, sondern *reflektierende Urteilskraft* (Kant). Hierin müssen die jungen Kolleginnen und Kollegen geschult werden. Ich nenne dies die „praktische Empirie": Wahrnehmung als Handlungsbedingung.

3.2 Die theoretische Empirie

Der zweite Ort, an dem in der Pädagogik die Wirklichkeitswahrnehmung relevant wird, ist die theoretische Empirie. Sie widmet sich jenen theoretischen Fragen, die sich nicht durch Nachdenken lösen lassen: Wie viele Schülerinnen und Schüler gibt es? Wie viele Schüler haben das Abitur? Wie viele Kinder lesen gerne? Wie viele Schüler gibt es, die Nachhilfeunterricht brauchen? (Und natürlich gehören hierher die komplexen Fragen der empirischen Bildungsforschung.)

Die Empirie kann allerdings *von sich aus* nicht aktiv werden. *Sie bedarf der ihr **logisch** vorausliegenden pädagogischen Fragestellung, um überhaupt aktiv werden zu können.* Illustriert am letzten Beispiel: Bevor man misst, wie viele Schüler es gibt, die Nachhilfeunterricht brauchen, muss man theoretisch klären: Was ist Nachhilfe? Was heißt brauchen? (Juristisch, pädagogisch, ethisch?) Wie misst man einen Mangel, einen Bedarf? Misst man das Können oder auch das Wollen?

All diese Fragen kann man beantworten. Man muss sie vor dem Messen beantworten haben. Nur wie? Das ist eine ausschließlich theoretische Frage. Empirie ist also immer *theoretisch* fundiert, ist immer prinzipiengeleitet. *Empirie ist immer der Prinzipienforschung nachgeordnet.* Sie handelt immer im Auftrag der Prinzipienforschung. Sie ist demnach relativ.

Allein schon, dass sie sich als *pädagogische* Empirie, als Bildungsforschung bezeichnet, setzt ja voraus, dass sie weiß, was Pädagogik ist und was Bildung heißt. Keine Empirie kann von sich aus bestimmen, was pädagogisch ist oder was Bildung soll. Alle Empirie sammelt ausschließlich Daten im Auftrag der ihr stets logisch vorgängigen prinzipienwissenschaftlichen Forschung. Empirische Aussagen sind also immer relativ.

Da sich nun aber die Prinzipienforschung historisch entfaltet, *kann die empirische Forschung nicht ‚besser' (objektiver) sein, als die Prinzipienforschung.* So genannte „rein empirische Forschung" ist genau in das involviert, was auch die Prinzipienforschung umtreibt: Den Streit um die richtigen Begriffe und Prinzipien. Dieser Streit lässt sich also gar nicht durch die Empirie beenden. Empirische Forschung kann nicht den Streit um Prinzipien aushebeln oder auf ein neues Niveau heben. Sie ist vielmehr Folge und Vertreter dieses Streits um Begriffe und Prinzipien. Denn die Empirie lebt ja nur und ausschließlich von diesen Begriffen und Prinzipien. Empirische Forschung in der Pädagogik ist nichts anderes als der verlängerte Arm, das Werkzeug der Prinzipienwissenschaft.

Schon was sich als *pädagogische* Empirie bezeichnet, hängt von dem theoretischen Begriff der Pädagogik ab. Jeder Begriff entfaltet sich aber in einem nicht still-stellbaren Diskurs. Also führt die empirische Forschung genau jenen Streit durch, der auch die Frage nach dem, was Pädagogik ist, umtreibt. Wer Pädagogik als, wie Dietrich Benner es sagt – *Aufforderung zur Selbsttätigkeit* – versteht, kommt zu einer anderen Empirie als derjenige, der Pädagogik als *Beeinflussung zur Verhaltensänderung* beschreibt – wie etwa Wolfgang Brezinka.

Die jeweilige Empirie klärt nicht, wer von beiden Recht hat. Sie gibt nur Auskunft innerhalb eines Modells. Sie bezieht sich nur auf das ihr theoretisch vorausliegende Modell. Empirie in der Pädagogik ist im besten Fall eine kontrollierte Form von Relativismus.

Innerhalb eines Diskurses kann man empirische Fragen empirisch beantworten. Selbstverständlich. Wer den Nachhilfebedarf klären will, kann dies ja nur empirisch. Aber was er unter Nachhilfe versteht, kann er nicht empirisch klären, sondern dieses Wissen setzt er voraus. Einen Begriff, eine Prinzipien-Theorie kann man empirisch weder bestätigen noch widerlegen. (Man kann nur *Hypothesen* widerlegen oder bestätigen, also vorläufige Annahmen, ob etwas vorab Bestimmtes der Fall sei oder eben nicht.)

Theorien, Begriffe oder Konzepte stehen unter dem Anspruch von Wahrheit. Empirische Aussagen stehen nur in Bezug zu Theorien. Sie selbst können keinen Wahrheitsanspruch erheben. Sie können nur wahrhaftig sein, d.h. verlässlich erworben. Für jegliches Handeln aber brauchen wir Aussagen unter Wahrheitsanspruch. Da empirische Aussagen diesen Wahrheitsanspruch allein nicht einlösen können, haben sie für die Frage, wie wir handeln sollen, keine basale Bedeutung. Erst im Zusammenhang mit einer wahrheitsfähigen Theorie, von der sie definitiv abhängig sind, haben empirische Aussagen Bedeutung. Empirie ist keine Bedingung des begründeten Handelns, sondern eine Bedingtheit.

Ein Beispiel: Die PISA-Studie definiert Kompetenzen und misst, inwieweit sie bei den Probanden vorhanden sind. Wenn nun jemand, wie z.B. ein Lehrer vor Ort, Bildung begründet anders definiert und nun beurteilt, inwieweit diese Bildung vorhanden ist, kommt er vermutlich zu einem anderen Ergebnis. Wer von beiden hat denn nun Recht? Nun: Jeder für sich, weil ja jeder seine Kriterien angelegt hat und - hoffentlich methodisch korrekt - misst. Aber damit lässt sich nicht klären, ob der von PISA bestimmte Kompetenzbegriff oder der Bildungsbegriff des Fachlehrers besser dafür geeignet ist, Aussagen über den Grad von Selbstbestimmung zu machen. Der Diskurs um die leitenden Begriffe kann nicht mittels der Empirie geführt und geklärt werden. Denn es ist ein theoretischer Diskurs, nämlich ein Diskurs über die Kriterien, die man für das Messen anlegen will. Die empirisch gewonnenen Daten klären diesen Streit in keiner Weise. Ich wiederhole mich: Empirie bestätigt nur, was vorausgesetzt wurde. Mehr kann sie nicht.

Empirie kann nicht über die Wahrheit von prinzipiellen Aussagen entscheiden, *sie setzt vielmehr die Wahrheit der Prinzipien – also der Messkriterien – als bereits gültig voraus*. Damit verweist sie alle Verantwortung an die theoretische Pädagogik. Und da ist sie gut aufgehoben.

Empirie setzt also erst *nach* einer theoretischen Wahrheitsklärung ein und kann sie nicht beeinflussen. Sie folgt immer und notwendig den außerhalb von Empirie gesetzten Kriterien. Empirische Forschung ist also „diskursrelativ". Sie gilt ausschließlich unter den angenommenen Voraussetzungen des Diskurses. Legt man einen neuen Diskurs zu Grunde, ist nichts von der alten Empirie gültig. Das meint der Satz: Empirie ist diskursrelativ.

Wissenschaft zielt aber nicht auf einen Relativismus, sondern auf Wahrheit. Der Diskurs um die Wahrheit findet aber in der Pädagogik nicht und nie in der Empirie statt, sondern bei der Diskussion darum, was Kriterien sind. Was Prinzipien sind. Hier wird entschieden. In der Empirie wird nur exekutiert, was zuvor entschieden wurde.

4. IST EMPIRIE NICHT EVIDENT?

Aber was ist mit „evidenzbasierter pädagogischer Forschung"? Für den Bildungsphilosophen disqualifiziert sich dieses *Wort* selbst – zumal dann, wenn es etwas Neues bezeichnen soll.

Seit Beginn der menschlichen Überlieferung (man vergleiche das angeführte Zitat von Aristoteles) zeigt sich, dass es jegliches Denken darauf anlegt, *evidenz*basiert zu sein. „Evident" heißt (wie das Lexikon belehrt), „unmittelbar einsichtig", „nicht in Frage zu stellen", „unabweisbar": ein „unzweifelbare[s] Einsichtigsein eines Sachverhalts aus dem Sichtighaben seiner Gründe".[11] Niemand, der Geltungsansprüche durchsetzen will, wird mit Einsichten zu überzeugen versuchen, die nicht evident sind. Eine Aussage, eine Forderung, die von sich selbst sagt, sie wolle aber nicht evident sein, hebt sich ja selbst auf. Mit dem Wort aber suggeriert eine besondere Art von Erziehungswissenschaft, nur die eigene Forschung sei evident – die andere Forschung nicht. Das „wording" diffamiert also andere Erkenntnisarten. Und zwar *alle anderen* Erkenntnisarten. Denn es behauptet, dass nur es selbst „evident" sei.

Das Problem, das aber nun die Erkenntnistheorie seit 4000 Jahren umtreibt, ist die Frage, was denn nun evident ist. Und mit welchen Evidenzen beweist man, was denn „Evidenz" ist? Das Wort „Evidenz" ist doch nicht die *Lösung* des Erkenntnisproblems, sondern das Problem.[13] Wenn wir Evidenzen hätten, dann brauchten wir keine Wissenschaft mehr. Jeder Mensch will auf Evidentes zurückgreifen – nur besteht das Problem genau darin, herauszufinden, was denn als evident gelten kann. „Evidenzbasiert" ist, philosophisch betrachtet, ein Nichtbegriff. Nebenbei: Intellektuelle Redlichkeit fängt bei einer redlichen Terminologie an.

5. EMPIRIEBASIERT: VOM SEIN ZUM SOLLEN – WIE GEHT DAS?

Der zweite Teil des Kompositums ist noch rätselhafter: evidenz*basiert*. Basieren heißt: Etwas zur Grundlage haben. Wir kennen das aus dem Marxismus: Basis und Überbau. Evidenzbasiert soll also heißen, dass man die Evidenz als Basis hat. Ich habe bereits darauf hingewiesen, dass diese Formulierung eine bloße Tautologie ist.

Aus der Verwendung des Wortes „evidenzbasiert"[13] ergibt sich, dass die Nutzer unter „Evidenz" die Empirie verstehen, also eigentlich „empiriebasiertes Entscheiden" meinen. Nun war aber eingangs gezeigt worden, dass pädagogisches Handeln immer normativ ist. Die Frage ist, wie man aus empirischen Daten normative Entscheidungen ableiten soll? Wie man normative Entscheidungen auf empirischen Daten basieren lassen, also Daten diesen zu Grunde legen will? Mit einer vollzogenen Handlung kann man nicht begründen, ob diese Handlung auch künftig sein soll. In der Philosophie wird der Trugschluss, aus dem Sein ein Sollen abzuleiten, seit zweieinhalbtausend Jahren als naturalistischer Fehlschluss bezeichnet. Es ist einer der ganz klassischen Fehlschlüsse und Irrtümer, nämlich der, normative Entscheidungen würden aus empirischen Aussagen hervorgehen, lägen also in deren Daten begründet. Jedoch: Normative Entscheidungen trifft man *angesichts* empirischer Umstände, nicht aber *auf Grund* von empirischen Daten.

Wenn man erhebt, dass die durchschnittliche Klassengröße 32 Kinder beträgt, dann folgt daraus doch nicht, ob man den Zustand ändern oder beibehalten soll. Aber könnte man denn nicht herausfinden, welche Klassenstärke die beste ist? Man könnte dies, wenn man Einigkeit über das Kriterium hätte: Was ist gute Pädagogik? Aber wenn man Dietrich Benner befragte, bekäme man eine andere Antwort, als wenn man Wolfgang Brezinka befragte. Und dieser Streit lässt sich nicht per Dekret schlichten. Er ist es nämlich, der für den Fortschritt in der Geschichte sorgt: Die Suche nach der Wahrheit. Wer behauptet, die Wahrheit bereits gefunden zu haben und nur noch empirisch nachprüfen zu wollen, der beendet die Geschichte. Der setzt Dogmen an die Stelle von Wahrheitssuche.

Ich vermute, ich muss jetzt nicht mehr im einzelnen nachweisen, dass die angeblich „evidenzbasierte Forschung" genau so oder genau so wenig evident ist wie jene Forschung, die sich nicht mit dieser Tautologie schmückt: „Dem positivistischen Wissenschaftsbegriff geht eine normative Entscheidung darüber, was Ziel menschlichen Erkennens und Handelns ist, voraus, welche den Horizont der Erkenntnis wie auch den des Handelns radikal einschränkt und auf eine technizistische Weltanschauung verkürzt."[14] Auch „evidenzbasierte Forschung" ist demnach diskursrelativ. Wenn man die Kriterien als gültig ansieht, kommt man nachgängig zu quantitativen Aussagen. Und diese sind sicherlich nach bestem Wissen und unter strengen Auflagen erhoben. Das wissenschaftliche Ethos gilt selbstverständlich auch hier. Aber die Daten geben keine Auskunft über die Gültigkeit der Kriterien. Damit stellt sich aber sofort der Meta-Diskurs über die Gültigkeit der Kriterien. Alles wie gehabt.

Auch hier gilt: Die Politiker, die sich von angeblich evidenzbasierter Forschung leiten lassen, sind nicht besser beraten als jene, die offen ihre Zielvorstellungen ausweisen. Jene vollstrecken, genau wie diese, lediglich Zielvorstellungen, die vor aller Erfahrung getroffen wurden – so, wie man es immer macht. Neu ist nur das Wort.

Die Vorstellung, dass eine „wertneutrale" Erziehungswissenschaft durch „Systembeobachtung" der Politik „neutrale" aber handlungsrelevante Daten liefert (als neues Konzept der sich von der „Pädagogik" unterscheidenden „Erziehungswissenschaft" dargestellt bei Terhardt[15]), hat keine Lösung für das Problem, woher denn in einem solchen Modell die *pädagogische* Expertise herkommen soll. Die Politik kann *aus ihrem eigenen Paradigma* keine *pädagogische* Expertise ableiten[16]; sie ist vielmehr auf wissenschaftliche pädagogische Expertise angewiesen; andernfalls wird sie zu „Staatspädagogik". Nur in den geschlossen Gesellschaften der Antike oder in den totalitären Systemen der unmittelbaren Vergangenheit galt bisher ein Konzept, in dem die Politik auch die pädagogischen Ziele (für Universität, Schule und Familie) bestimmte und die Erziehungswissenschaft lediglich als Datenlieferant tätig war und Vermittlungstechnologie entwickelte: Staatspädagogik.[17] Vielleicht erklärt dieser verkürzte Wissenschaftsbegriff von einigen Vertretern der Erziehungswissenschaft, warum es zu einem massiven Verlust pädagogischer Expertise im bildungspolitischen Handeln gekommen ist: Eine nur noch beschreibende Erziehungswissenschaft entzieht sich der normativen Verpflichtung, überantwortet pädagogische Entscheidungen der Politik, die mittels eines ja

nur *politischen Diskurses* pädagogische Prozesse qua Definition nicht gestalten kann.

6. EMPIRIE ALS ZWANGSMITTEL

Wissenschaftliche Empirie ist bei Lehrern wenig beliebt. Warum eigentlich? Wenn sie so viel hält, wie sie verspricht, dann müsste sie doch hochwillkommen sein. Ist sie aber nicht. Zwei Gründe hierfür:

Die Empirie hat ihre mindere Anerkennung bei den Schulpraktikern einem besonderen Umstand zu verdanken. Es hat sich schnell herausgestellt, dass die quantitativen Erhebungen mehr sein wollen als neutrale Bestandsaufnahmen, die dann besseres Handeln ermöglichen. Es hat sich gezeigt, dass empirisch quantitative Erhebungen eine *Lenkungsfunktion* haben. Die Auftraggeber wollen mit den Untersuchungen etwas durchsetzen, und zwar ihr Programm. Es scheint, dass die Empirie vernünftige Begründungen (und damit Diskussionen um die Gültigkeit der Gründe) ersetzen soll. Die Daten werden erhoben, um gewissermaßen unter Umgehung von Argumentationen normativ zu wirken. Die Zahlen sollten für sich sprechen. Sie sind (angeblich) „ohne Alternative". Man will nicht (langwierig) überzeugen, sondern (kurzfristig) erzwingen.[18] Damit wandelt sich die Empirie von einem Verfahren zum Sammeln von Daten zu einem Instrument zur Lenkung von Prozessen. Das aber überfordert sie, stellt – philosophisch betrachtet – eine Erschleichung dar.

Wir haben gesehen: erst zusammen mit begrifflichen und prinzipiellen Vorentscheidungen haben Daten überhaupt eine Aussagekraft. Allein besagen sie gar nichts: Genau umgekehrt aber werden die Erhebungen nun angewandt. Manchmal sagen dies dann auch Vertreter der Empirie. Ulrich Lommel empfiehlt, dass „der Markt" oder die staatliche Aufsicht „schlechte Ausbildungsergebnisse mit *Eingriffen* in die Ressourcenausstattung *ahnde*". [19] Hier zeigt sich die Intention im Vokabular: Es geht um „Eingriffe", um das „Ahnden" von etwas. Empirische Standards z.B. sollen das Lehrpersonal zu „einem gemeinsamen Verständnis" „zwingen". Empirische Standards sollen falsche Lehrmethoden – z.B. die Vorlesung – „abstrafen". Hier haben wir alle Begriffe des New Public Managements vorliegen. *Die Empirie dient*, in Verbindung mit staatlicher Macht oder der Macht des Marktes, *dem „Ahnden", dem „Zwingen" und dem „Abstrafen"*.

Mag sein, dass man mit *Ahndung, Zwang* und *Strafen* technische Prozesse in Fabrikationsbetrieben optimieren kann: Im gesamten Bildungsbereich jedoch sind „Ahndung", „Zwang" und „Strafe" wenig überzeugend. Sie führen zur inneren Kündigung. Sie dokumentieren die Ohnmacht gegenüber dem Umstand, den anderen durch vernünftige Argumente nicht überzeugen zu können. *Ahndung, Zwang* und *Strafen* in Bildungsprozessen sind ein Dokument intellektueller Hilflosigkeit.

Und außerdem: Niemand, der seine Schüler lehren soll, sein eigenes Leben zu bestimmen, wird sich damit abfinden, dass er selbst gezwungen und abgestraft wird, wenn er nicht unmittelbar so pariert, wie sich das eine für fünf Jahre gewählte Verwaltung so denkt. Die Empirie spielt das Spiel längst getroffener Entscheidungen der Politik. Sie dient der „Akzeptanzbeschaffung".

Da aber leicht zu durchschauen ist, dass Daten nicht neutral erhoben werden, sondern lediglich das durchsetzen sollen, was längst entschieden ist, wehren sich die Menschen gegen diese Datenerhebungen. Niemand wird ehrlich antworten, wenn er weiß, dass seine Antwort ihn seinen Arbeitsplatz kosten kann. Insofern befindet sich die Empirie auch in einem tragischen Teufelskreis: Selbst dort, wo man sie als neutralen Datengeber nutzen könnte, kann sie diese Aufgabe nicht übernehmen: Man kann damit rechnen, dass die Daten nicht stimmen.

Kann ich das empirisch beweisen? Nein – und da sind wir bei einem weiteren Problem:

7. EMPIRIE ALS MACHT

Empirie führt zu ungleicher Machtverteilung: Empirie ist teuer und aufwändig. Man kann sie nicht schnell widerlegen – und zwar nicht, weil die Ergebnisse so gut wären, sondern weil die Methoden so teuer und aufwändig sind. Wenn man den Ergebnissen nicht traut, ist es ohne erhebliche Drittmittel nicht möglich, eine Kontrollstudie zu erstellen. Wer gewährt die Drittmittel…?

Empirie wird zur Macht, nicht auf Grund von Argumenten, sondern weil sie ein kostspieliges Verfahren ist, dem der Einzelne nichts entgegenzusetzen vermag. Für die Wahrheitsfindung ist derlei Machtgefälle nicht hilfreich.

Aber wohin geht die Ohnmacht? In die Resignation. Die innere Kündigung – das ist das, was die Schulen am wenigsten brauchen.

Wahrheit darf nicht an die zur Verfügung stehenden Geldmittel gebunden bleiben. Gut für eine Gesellschaft ist das nämlich nicht. Die Irrtümer dauern dann zu lange.

8. DOPPELTER FEHLER

Empirische Forschung in der Pädagogik ist keine Neuheit. Sie stellt die Pädagogik nicht vom Kopf auf die Füße.[20] Ihre Ergebnisse sind nicht auf eine neue Art evident. Sie „basiert" keine normativen Entscheidungen. Das alles ist nicht wahr: Pädagogik war und ist immer, und zwar in der geschilderten Weise, in doppelter Hinsicht empirisch: In der reflektierenden Urteilskraft der Handelnden und in der Datenerhebung innerhalb eines Diskurses.

Allerdings wurde diese doppelte Empirie nicht immer bildungsphilosophisch erinnert. Und immer dann, wenn diese Erinnerung ausblieb, kam es zu Selbstüberschätzungen der einen oder anderen Art.

1| Vgl. ergänzend meinen Aufsatz (2006): „Das Milieu muss besonders günstig gewesen sein." Über die Dignität von Praxis und die Vorläufigkeit von Geschichte. In: Montessori. Zeitschrift für Montessori-Pädagogik 44, 69-84.

2| Die folgende Begrifflichkeit nach: Böhm, W. (2005): Wörterbuch der Pädagogik, 16. Auflage. Stuttgart.

3| Vgl. Böhm, W.(1995): Theorie und Praxis: Eine Einführung in das theoretische Grundproblem, 2. Auflage. Würzburg.

4| Menze, C. (1983): [Art.] Bildung. In: Enzyklopädie Erziehungswissenschaft. Handbuch und Lexikon der Erziehung in 11. Bänden. Bd. 1: Theorien und Grundbegriffe der Erziehung und Bildung. Hg. v. Lenzen, D. / Mollenhauer K. Stuttgart, 350-356. Hier 350.

5| Vgl. Heitger, M. (2004): Selbstbestimmung als regulative Idee der Bildung. In: Ders.: Bildung als Selbstbestimmung. Hg. v. Böhm, V. W. / Ladenthin, V. Paderborn, 19-34.

6| Vgl. die Zusammenstellung in: Ladenthin, V. (2007): Philosophie der Bildung. Eine Reise von den Vorsokratikern bis zur Postmoderne. Bonn (Klassiker Denken Bd. 4).

7| Umfassend dargelegt in: Benner, D. (2005): Allgemeine Pädagogik. Eine systematisch-problemgeschichtliche Einführung in die Grundstruktur pädagogischen Denkens und Handelns. 5. korrigierte Auflage, Weinheim.

8| Vgl. Ladenthin, V. (2008): Das lernende Selbst zwischen Inhalt und Methode. Zum Umgang der Schule mit dem Selbst. In: Arnold, R. u. a. (Hg.): Lernen lebenslang – Ansichten und Einsichten. Baltmannsweiler, 52-74.

9| Im Original steht: Φανερον δε και τουτο, was dann auch soviel heißt wie: Dann/Auch ist dieses offenbar. Die Grundbedeutung des ersten gr. Wortes umfasst Folgendes: 1. leuchtend, sichtbar, in die Augen fallend, glänzend; 2. (übtr.) a.) ersichtlich, offenbar, augenscheinlich, offenkundig, deutlich, klar, einleuchtend b.) öffentlich, unverhohlen; c.) hervortretend, auffallend.

10| Vgl. Ladenthin, V. (2009): Ist Bildung notwendig? In: Heimbach-Steins, M. u. a. (Hrsg.): Bildung, Politik und Menschenrecht. Ein ethischer Diskurs. Bielefeld, 69-80.

11| Kleines Philosophisches Wörterbuch (1971). Hg. v. Müller, N. / Halder, A., unter Mitarbeit von Brockard, H. / Müller, S. / Welsch, W.. Freiburg i. Br., 80 [Art. Evidenz].

12| Vgl. Kulenkampff, A.(1974): [Art.:] Evidenz. In: Handbuch philosophischer Grundbegriffe. Hg. v. Krings, H./ Baumgartner, H. M. / Wild, C., Bd. II. München, 425-435 (Studienausgabe).

13| Vgl. „evidenzbasierte Erziehungswissenschaft: auf bestätigten Erfahrungen beruhende, in ihrer Qualität an hohen Standards der Prüfung orientierte, i.d.R. experimentell fundierte (…) Erwartung an die Erziehungswissenschaft. In der Übernahme von Kriterien für die Forschung, die in der Medizin entwickelt wurden…." Aus: Beltz. Lexikon Pädagogik (2007): Hg. v. Tenorth, H.-E. / Tippelt, R.. Weinheim/Basel, 224 [Stichwort „evidenzbasierte Erziehungswissenschaft].

14| Benner, D. (1991): Hauptströmungen der Erziehungswissenschaft. Eine Systematik traditioneller und moderner Theorien. 3. verbesserte Auflage. Weinheim, 191.

15| Vgl. Terhart, E. (2003): Erziehungswissenschaft zwischen Forschung und Politikberatung. In: Vierteljahrsschrift für wissenschaftliche Pädagogik 79, 74-90.

16| In knapper Form hat D. Benner den systematischen Grund für die notwendige und kategoriale Unterscheidung des politischen vom pädagogischen Diskurs noch einmal dargestellt in: Benner, D.: [Über] Wilhelm von Humboldt: Ideen zu einem Versuch, die Grenzen der Wirksamkeit des Staates zu bestimmen.

In: Böhm, W. u. a. (Hg.) (2009): Hauptwerke der Pädagogik. Paderborn, 202-204. „Eine erneute Aufmerksamkeit" finde dieser Text erst „seit den 1980er Jahren, in denen zwischen staatlicher und öffentlicher Erziehung (…) sowie wie widerstreitenden Abstimmungsproblemen von Ökonomie, Ethik, Pädagogik, Politik, Kunst und Religion wieder unterschieden (!) wird." (204).

17| *Vgl. Benner, D. / Sladek, H. (1996): Ist Staatspädagogik möglich? In: Vierteljahrsschrift für wissenschaftliche Pädagogik 72, 1-15: „Staatspädagogik ist nur als Ideologie möglich." (14).*

18| *Vgl. Ladenthin, V.(2003): Pluralismus im Schulsystem: Begründungen und Formen. Nebst einigen Überlegungen zum Konzept der „Selbständigen Schule" und zur Verwaltungsreform im Bereich der Schuladministration. In: Ladenthin, V. (Hrsg.) (2003): Religion und Bildung im Pluralismus. Münster, 82-102 (Münstersche Gespräche zur Pädagogik Bd. 19).*

19| *Vgl. Lommel, U. (2009): Ohne Alternative. In: Forschung & Lehre 16 H.7/09, 498-499.*

20| *Jungmann, W. / Huber, K. haben gezeigt, dass der Begriff der „Empirischen Wendung" der Pädagogik, der sich gerne auf Heinrich Roth beruft (vgl. Jungmann, W. / Huber, K. [Hg.] [2009]: Heinrich Roth. – ‚moderne Wissenschaft als Pädagogik. Weinheim/München, 28), sich nur in umfassender Perspektive auf ihn berufen kann: Roth hat explizit für ein Ineinander von Theorie und Empirie plädiert: „Es geht ja nicht nur um das Präsenthaben der Forschungsergebnisse aus den Wissenschaften vom Menschen, es geht um die Integrierung dieser Ergebnisse unter der spezifischen Fragehaltung der Pädagogik im Hinblick auf die dauernd wechselnden Aufgaben der unterrichtlichen und erzieherischen Praxis." (Aus: Roth, H. [1958]: Die Bedeutung der empirischen Forschung für die Pädagogik. In: Jungmann / Huber, a.a.O., 28-50, hier 34).*

<< zurück zum Inhaltsverzeichnis

HERAUSGEBER UND AUTOREN

Professor Dr. Jörg-Dieter Gauger
Stellvertretender Leiter der Hauptabteilung Wissenschaftliche Dienste der Konrad-Adenauer-Stiftung, Koordinator für Bildungs- und Kulturpolitik in der Hauptabteilung Politik und Beratung

Josef Kraus
OStD, Präsident des Deutschen Lehrerverbandes (DL)

Prof. Dr. Peter Bender
Didaktik der Mathematik, Fakultät für Elektrotechnik, Informatik und Mathematik, Universität Paderborn

Prof. Dr. Volker Ladenthin
Institut für Kommunikationswissenschaften, Abteilung für Bildungswissenschaft, Lehrstuhl für Historische und Systematische Erziehungswissenschaft, Rheinische Friedrich-Wilhelms-Universität Bonn

Prof. Dr. Dr. h.c. Rainer Lehmann
Abteilung Empirische Bildungsforschung und Methodenlehre, Philosophische Fakultät IV, Institut für Erziehungswissenschaften, Humboldt-Universität zu Berlin

Prof. Dr. Heinz-Elmar Tenorth
Historische Erziehungswissenschaft, Philosophische Fakultät IV, Institut für Erziehungswissenschaften, Humboldt-Universität zu Berlin

Prof. Dr. Manfred Weiß
Bildungsökonomie und Bildungsforschung, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Frankfurt am Main

ANSPRECHPARTNER IN DER KONRAD-ADENAUER-STIFTUNG

Prof. Dr. Jörg Dieter Gauger
Stellv. Leiter der Hauptabteilung Wissenschaftliche Dienste
53754 Sankt Augustin
Telefon: +49(0)-2241-2 46 23 02
E-Mail: joerg-dieter.gauger@kas.de