

Lange, Hermann

## **Qualitätssicherung und Leistungsmessung in der Schule auf internationaler und nationaler Ebene**

*Oelkers, Jürgen [Hrsg.]: Zukunftsfragen der Bildung. Weinheim : Beltz 2001, S. 127-150. - (Zeitschrift für Pädagogik, Beiheft; 43)*

urn:nbn:de:0111-opus-79195



in Kooperation mit / in cooperation with:

# **BELTZ JUVENTA**

<http://www.juventa.de>

### **Nutzungsbedingungen / conditions of use**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

We grant a non-exclusive, non-transferable, individual and limited right to using this document.  
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.  
By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

**peDOCS**  
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)  
Mitglied der Leibniz-Gemeinschaft  
Informationszentrum (IZ) Bildung  
Schloßstr. 29, D-60486 Frankfurt am Main  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Zeitschrift für Pädagogik  
43. Beiheft



Zeitschrift für Pädagogik  
43. Beiheft

# Zukunftsfragen der Bildung

Herausgegeben von Jürgen Oelkers

Beltz Verlag · Weinheim und Basel

# Inhaltsverzeichnis

Vorwort .....	7
---------------	---

## **Teil I: Bildungsforschung und Legitimation**

<i>Jürgen Baumert</i> Vergleichende Leistungsmessung im Bildungsbereich. ....	13
<i>Helmut Fend</i> Bildungspolitische Optionen für die Zukunft des Bildungswesens. Erfahrungen aus der Qualitätsforschung .....	37
<i>Dietrich Benner</i> Bildung und Demokratie .....	49

## **Teil II: Bildungsökonomie**

<i>Manfred Weiß</i> Quasi-Märkte im Schulbereich. Eine ökonomische Analyse .....	69
<i>François Grin</i> On effectiveness and efficiency in education: Operationalizing the concepts .....	87
<i>Geoff Whitty/Sally Power</i> Devolution and Choice in Education: The research evidence to date .....	99

## **Teil III: Bildungspolitik und Lehrerbildung**

<i>Ernst Buschor</i> Evaluation als Teil der Zürcher Bildungspolitik .....	121
<i>Hermann Lange</i> Qualitätssicherung und Leistungsmessung in der Schule auf internationaler und nationaler Ebene .....	127
<i>Jürgen Oelkers</i> Welche Zukunft hat die Lehrerbildung? .....	151

#### **Teil IV: Neue Medien**

*Bernd Weidenmann*

Veränderungen des Lernens durch neue Medien. . . . . 167

*Renate Schulz-Zander*

Lernen mit neuen Medien in der Schule . . . . . 181

## **Qualitätssicherung und Leistungsmessung in der Schule auf internationaler und nationaler Ebene\***

In der gegenwärtigen schulpolitischen Diskussion spielen die Begriffe der Qualitäts- und Standardsicherung in einer scheinbar einfachen Analogie zu etablierten Vorgehensweisen der Wirtschaft eine große Rolle. Man fügt sich Modernisierungsforderungen, bei denen diejenigen, die sie erheben, nicht nur die Ziele zu kennen glauben, die es zu erreichen gilt, sondern auch über die ubiquitär einsetzbaren Mittel zu verfügen scheinen, die das Erreichen der Ziele möglich machen.

Indessen sind die Analogien nicht einfach zu begründen. »Qualitätssicherung« unterstellt, man weiß, was Qualität ist und muss sie deshalb nur noch sichern. Wer die breite Diskussion um Schulqualität kennt, weiß jedoch, dass die Verständigung auf das damit Gemeinte alles andere als leicht ist. Das hat seine Gründe nicht im pädagogischen Theorienstreit, sondern in der Vielschichtigkeit des zu lösenden Problems. »Standardsicherung« unterstellt die Standardisierbarkeit von Bildungsprozessen im Sinne normierter Produktionsverfahren und Ergebnisse. Wer dieses annimmt, wird alsbald auf das gewichtige Gegenargument treffen, »Bildung« sei ein individueller Prozess, der in der Freiheit des lernenden Subjekts zu gestalten und deshalb nicht normierbar sei. Dieses Argument gewinnt zunehmende Bedeutung in einer Situation, in der eine wachsende Heterogenität der Schülerschaft konstatiert wird und die Einräumung größerer Gestaltungsspielräume für die einzelne Schule zur Ermöglichung situationsgerechter Lösungen (»Schulautonomie«) zum Handlungsrepertoire der Bildungspolitik gehört.

Zugespitzt wird die Diskussion schließlich durch die unter dem Schlagwort »Entwickeln statt messen« geführte Auseinandersetzung, in der – neben der Behauptung mangelnder Notwendigkeit des Messens – auch das Argument mangelnder Messbarkeit von Schulqualität eine gewichtige Rolle spielt. Wer letzteres behauptet, wird freilich dem Entwicklungsstand der Wissenschaft nicht gerecht. Entwickeln verfolgt Ziele und will Wirkungen erzeugen. Die Ziele müssen formuliert und damit überprüfbar, die Wirkungen müssen geklärt, bewertbar und notfalls korrigierbar gemacht werden. Dies alles kann sich nicht auf die raunende Beschwörung eines pädagogischen Mysteriums beschränken. Es bedarf gesicherter Verfahrensweisen. Diese stehen teils bereits zur Verfügung, teils entwickeln sie sich im Zuge der eingeleiteten Untersuchungen.

\* Ergänzte Fassung eines Vortrages anlässlich der 3. Jahrestagung der Deutschen Gesellschaft für Evaluation e.V. »Evaluation – Reformmotor oder -bremse« am 23. 11. 2000 im Wissenschaftszentrum (WZB) in Berlin.

Damit ist zugleich gesagt: Die Entwicklung und der Einsatz der hierfür benötigten Untersuchungsinstrumente sind alles andere als einfach. Zu sichern ist, dass man klärt, was man messen will, dass man begründet, warum man gerade dies und nicht anderes misst, dass man tatsächlich misst, was man zu messen behauptet, und dass man Rechenschaft über die zur Bewertung herangezogenen Maßstäbe gibt. Man muss also – um einige Beispiele aus laufenden Projekten aufzugreifen – deutlich machen,

- was unter »Lesekompetenz«, »mathematischer Kompetenz« oder »naturwissenschaftlicher Kompetenz« verstanden werden soll und wie die zu erfassenden Konstrukte theoretisch expliziert werden;
- was »soziale Kompetenz« bedeutet und inwiefern es sich dabei um ein Wissen um soziale Sachverhalte, um die Fähigkeit zu Kooperation und Kommunikation oder um ein spezifisches Einfühlungsvermögen in die Lebenssituation anderer Menschen handelt;
- was »Problemlösekompetenz« heißt und wie sie sich von fachlichen Kompetenzen einerseits und allgemeiner Intelligenz andererseits abgrenzt;
- warum man gerade den untersuchten Aspekten vor anderen Aufmerksamkeit zuwendet;
- ob und warum man Zielerreichungsgrade anhand normativer Kriterien, Lernfortschritte innerhalb bestimmter Zeiträume aufgrund ipsativer Kriterien oder Leistungsunterschiede innerhalb von oder zwischen Lerngruppen im Rahmen von Vergleichen messen will.

Dies alles ist vorzuschicken, um deutlich zu machen, dass keine einfache Materie zu behandeln ist. Allzu grobe Vereinfachungen kann man sich nicht leisten. Dabei kann es gelegentlich auch systemrational sein, Ziele nur diffus zu beschreiben, sie nicht zu operationalisieren und die Zielerreichung damit auch nicht messbar zu machen, um sie in ihrem Symbolwert unangreifbar zu halten.

### *Zum Problem der »Qualitätssicherung«: Worauf kommt es an?*

Vor diesem Hintergrund sind zunächst einige Überlegungen zum Thema der »Schulqualität« anzustellen. Zu beginnen ist mit Unterscheidungen. Es geht um so unterschiedliche Dinge wie Strukturqualität, Prozessqualität und Ergebnisqualität. Die deutsche Diskussion um die richtige Bestimmung von Schulqualität konzentrierte sich bisher vor allem auf Strukturvergleiche und Aspekte der Prozessqualität, mindestens im Sinne der Klärung systematischer Zusammenhänge, nicht unbedingt auch im Sinne eines Monitoring, wieweit diese Merkmale im täglichen Handeln der Schulen tatsächlich realisiert werden. Prozessmerkmale sind die Merkmale »guter Schulen«, zu denen – ich lehne mich hierbei an Formulierungen von HELMUT FEND an – zum Beispiel folgende gehören: Eine klare Konzeption pädagogischer Leitideen, effiziente Führung in Fragen des Unterrichts und der Unterrichtspraxis, hohe Erwartungen an Lernfähigkeit und Leistungsbereitschaft von Schülerinnen und Schülern, eine sichere, ordentliche und ästhetische Schulumwelt, bestmögliche Zeit-



nutzung, häufige Beobachtung der Fortschritte der Schüler, positive Beziehungen zwischen Schule, Familien und Gemeinde (FEND 2000). EWALD TERHART nennt folgende innerschulische Faktoren: Akademisches Klima, hohes Verantwortungsgefühl innerhalb der Lehrerschaft, kompetente Schulleitung, Kooperation der Lehrkräfte (TERHART 2000a). Mit der Orientierung an diesen Prozessmerkmalen ist zugleich gesagt: »Qualität« entsteht auf der Ebene der konkreten Arbeit in der einzelnen Schule. Im Rahmen einer solchen Betrachtung wendet sich der Blick – dies ist ein wichtiges Ergebnis der Schulforschung – von der Betrachtung des Gesamtsystems auf die Analyse des Handelns der einzelnen Schule.

Die Betrachtung der Struktur- und Prozessqualität hat ihren eigenen Wert. Indessen werden Strukturen geschaffen und Prozesse initiiert, um Ergebnisse zu erreichen. Letztlich kommt es also auf die Ergebnisqualität an. Die Frage nach den »Ergebnissen« des Lernens deutscher Schülerinnen und Schüler, nicht zuletzt auch die Klärung ihrer Fachleistungen, spielte in der Diskussion der vergangenen Jahre um Schulqualität jedoch kaum eine Rolle. Dies erklärt viele Reaktionen auf die gegenwärtig verhältnismäßig große Dichte von Untersuchungen in diesem Bereich. Es macht zugleich die Unsicherheit von Schulen und Bildungsadministration verständlich, wie mit den Untersuchungsergebnissen umzugehen sein wird. Dabei ist einleuchtend, dass die genannten Prozessmerkmale auf einzelschulischer Ebene wichtige Voraussetzungen für Lernerfolg sind. Damit allein sind aber zureichende Bedingungen zur Sicherung der Ergebnisqualität des Lernens noch nicht formuliert. Der Hinweis, dass es – wo die Prozesse in der Schule in Ordnung sind – auch an guten Ergebnissen nicht fehlen könne, reicht jedenfalls nicht aus. Leistungen hängen nicht nur von dem Handeln der einzelnen Schule, sondern auch von der Lebenssituation ihrer Schülerinnen und Schüler, ihrer Umwelt und von übergreifenden Systembedingungen ab. Zudem geht es nicht allein um die Ergebnisse des Handelns einer einzelnen Schule, sondern auch um den Ertrag des Gesamtsystems als Summe des Handelns aller Schulen.

Wer Ergebnisse bewerten will, muss wissen, was er erreichen wollte (Ziele) und was er unter den obwaltenden Umständen hätte erreichen können (Rahmenbedingungen). Ergebnisse können sich sowohl auf quantitative (Absolventenzahlen, Dauer von Ausbildungsgängen, Abbrecherquoten usw.) als auch auf qualitative Aspekte beziehen. Letztere zu bewerten ist nur möglich, wenn man zuvor formuliert, welche Lernziele man verfolgt. Das Zielsystem, vor dessen Hintergrund wir insoweit operieren, ist komplex und vielschichtig. Es gibt fachliche Zielsetzungen in unterschiedlichen Feldern. Es gibt kognitive und metakognitive, soziale und emotionale Lernziele. Man kann Lernziele in Bezug auf das Individuum (bestmögliche Förderung Einzelner) oder in Bezug auf die Gesamtheit der Individuen (bestmögliche Förderung aller) formulieren. Es gibt auch gesellschaftliche Ziele, die sich im Lernen von Individuen verwirklichen, aber darin nicht ihren letzten Grund finden: Die Sicherung der historischen Kontinuität der gesellschaftlichen Entwicklung, die Sicherung der Konkurrenzfähigkeit des eigenen Landes in einer globalen Weltordnung, die Sicherung einer demokratischen und toleranten Gesellschaft, die Sicherung gemeinsamer Wertvorstellungen usw.

Diese unterschiedlichen Zielsetzungen stehen in einem komplizierten Ver-

hältnis zueinander. Nicht alles lässt sich gleichzeitig gleich gut erreichen. Optimierungen zugunsten bestimmter Ziele können negative Konsequenzen für die Möglichkeit der Realisierung anderer Ziele haben. In anderen Fällen wird die Optimierung zugunsten eines Ziels auch die Erreichung anderer fördern. So kann es z.B. sein, dass die häufig behauptete Entgegensetzung von kognitiven und sozialen Lernzielen unzutreffend ist und dass zwischen beiden positive Rückkopplungszusammenhänge bestehen. Dennoch: Prinzipiell ist die Möglichkeit von Zielkonflikten nicht zu bestreiten. Dies ist freilich kein spezifisches Dilemma einer Evaluation. Es ist vielmehr ein Problem im Handeln selbst. Wer sich entscheidet, ein bestimmtes Ziel zu erreichen, muss zugleich klären, ob bzw. in welchem Umfang er dafür Einschränkungen bezüglich anderer Ziele in Kauf nimmt. Die Möglichkeit von Zielkonflikten wird im praktischen Handeln indessen häufig nicht bewusst oder nicht sorgfältig genug geklärt. Im Rahmen einer Evaluation ist eine derartige Problemverdrängung nicht möglich.

Zielklarheit und die damit verbundene Klärung von Bewertungsmaßstäben vorausgesetzt, kann man den Grad der Zielerreichung »messen« (auch wenn noch einmal zu betonen ist, dass das Messen und die Interpretation von Messergebnissen keine triviale Sache sind). Dabei gilt noch immer, was J.R. MISLEVY 1995 festgestellt hat: »... no single index of achievement can tell the full story and ... each has its own limitations ...« (MISLEVY 1995). Man kann nicht alles gewissermaßen in einem Aufwasch messen. Ebenso wenig ist es möglich, die Messergebnisse in einer einzigen Kennziffer auszudrücken. In unterschiedlichen Gegenstandsbereichen können sich unterschiedliche Leistungsniveaus zeigen. Man muss sich deshalb entscheiden, welchem Ausschnitt des Zielbündels man zum gegebenen Zeitpunkt seine Aufmerksamkeit zuwenden will, diese Auswahl begründen und klären, welche Skalen für die Ergebnisdarstellung gewählt werden sollen. Im Zuge von Evaluationen sind immer nur partielle Urteile über den Stand der Dinge möglich. Was in einer Hinsicht gilt, kann in anderer Hinsicht ganz anders sein. Aber es gibt auch Verknüpfungen: Wo eines nicht befriedigend ist, kann auch anderes im Argen liegen. Dabei kann sich der Blick zu unterschiedlichen Zeitpunkten unterschiedlichen Aspekten zuwenden. Was zu einem Zeitpunkt wichtig genommen wird, muss nicht zu jedem anderen Zeitpunkt gleichermaßen wichtig sein. Die Werteskala hat sich in der Vergangenheit verändert und wird dies auch in Zukunft tun. Dass man nicht alles gleichzeitig messen kann und sich zwangsläufig auf Teilaspekte beschränken muss, ist freilich kein Einwand gegen die Möglichkeit von Evaluationen schlechthin. Wem fehlt, was ihm wichtig ist, kann daraus nur die Forderung ableiten, ergänzende Untersuchungen einzuleiten. Die Folgerung, dass man – wenn nicht alles gleichzeitig zu machen ist – überhaupt nichts machen sollte, wäre jedenfalls nicht sinnvoll.

### *Welche Untersuchungen werden gegenwärtig durchgeführt?*

Zum gegenwärtigen Zeitpunkt wendet sich die Aufmerksamkeit insbesondere den Fachleistungen und dabei vor allem den Aspekten der Sprachkompetenz (Deutsch bzw. Muttersprache und Fremdsprache), der mathematischen Kompetenz und der naturwissenschaftlichen Kompetenz von Schülerinnen und

Schülern zu. Sie stehen im Mittelpunkt der großen internationalen und nationalen Studien wie TIMSS, PISA, DESI, PIRLS, IGLU, LAU, QuaSUM und MARKUS.

TIMSS (The Third International Mathematics and Science Study) wurde von der IEA (International Association for the Evaluation of Educational Achievement) durchgeführt. Gegenstand ist die mathematisch-naturwissenschaftliche Grundbildung. Die Untersuchungen beziehen sich auf drei »Populationen«: Schülerinnen und Schüler der Primarstufe, der Sekundarstufe I und der Sekundarstufe II (allgemeinbildendes und berufliches Schulwesen). In Deutschland sind nur die Populationen II und III, nicht auch die Grundschülerinnen und -schüler in die Untersuchung einbezogen worden (J. BAUMERT/R. LEHMANN u.a. 1997; J. BAUMERT/W. BOS/R. WATERMANN 1998; J. BAUMERT/W. BOS/R. WATERMANN 2000). An den vorausgegangenen Untersuchungen der IEA – der ersten und der zweiten Mathematikstudie (First and Second International Mathematics Study: FIMS 1964 und SIMS 1980 bis 1982) und den beiden Naturwissenschaftsstudien (First and Second International Science Study: FISS 1970/71 und SISS 1983/84) – war Deutschland nicht oder nur mit wenigen Bundesländern beteiligt. PISA (Programme for International Student Assessment) ist ein von der OECD (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung) initiiertes Projekt, dessen Gegenstand die vergleichende Untersuchung der Lesekompetenz, der mathematischen Kompetenz, des naturwissenschaftlichen Verständnisses und bestimmter Aspekte außerfachlicher Kompetenzen (»Cross curricular Competencies«<sup>1</sup>) von 15-jährigen Schülerinnen und Schülern ist. An der Untersuchung beteiligen sich mehr als 30 Staaten. PISA soll in »drei Wellen«, im Abstand von drei Jahren, durchgeführt werden, bei denen jeweils einer der genannten Kompetenzbereiche die Hauptkomponente bildet<sup>2</sup>. Der internationale Ergebnisbericht über die erste Welle der Untersuchung soll im vierten Quartal 2001 vorgelegt werden. Die Entscheidung über die Vergabe des nationalen Projektmanagements für die zweite Welle von PISA ist in der Kultusministerkonferenz unlängst erfolgt. Der entsprechende Bericht soll 2004 vorliegen. Für die erste Welle von PISA haben sich die Länder in der Kultusministerkonferenz darauf verständigt, die Stichprobe für bundesinterne Vergleiche zu erweitern (PISA-E). DESI (Deutsch-Englisch-Schülerleistungen-International) ist eine Untersuchung zur aktiven Beherrschung des Deutschen und zur Fremdsprachenkompetenz der Schülerinnen und Schüler an deutschen Schulen auf der 9. Jahrgangsstufe. Sie ist von der Kultusministerkonferenz ausgeschrieben worden und ergänzt PISA insofern, als es in dem letzteren Projekt nur um die Lesekompetenz im Deutschen geht. PIRLS (Progress in International Reading Literacy Study) bzw. IGLU (Internationale Grundschul-Lese-Untersuchung) ist eine internationale Studie zur Lesekompe-

- 1 In der »ersten Welle« von PISA sind dies im Rahmen der internationalen Untersuchung die motivationalen und kognitiven Voraussetzungen selbstregulierten Lernens (Lernstrategien, Lernbereitschaft, Interessen, Ausdauer, Kooperationsbereitschaft usw.). Im Rahmen der nationalen Ergänzung werden die Problemlösefähigkeit und Aspekte von Kooperation und Kommunikation einbezogen.
- 2 Nähere Einzelheiten zu dem Projekt sind im Internet zu finden: <http://www.mpib-berlin.mpg.de/PISA/>. Vgl. zur Rahmenkonzeption für diese Untersuchung auch: Deutsches PISA-Konsortium: Rahmenkonzeption.

tenz von Grundschülerinnen und Grundschulern, welche ebenfalls von der IEA durchgeführt wird. An ihr beteiligt sich Deutschland aufgrund eines Beschlusses der Kultusministerkonferenz. Der überwiegende Teil der Bundesländer hat sich darauf verständigt, mit der Lesestudie auch eine Studie zur mathematischen und naturwissenschaftlichen Kompetenz von Grundschulern zu verbinden und auf diese Weise die mangelnde Beteiligung deutscher Schulen an der TIMSS-Grundschulstudie zu kompensieren (IGLU-E).

Deutsch und erste Fremdsprache, vor allem aber Mathematik stehen auch im Zentrum regional begrenzter Untersuchungen, die von einzelnen Bundesländern durchgeführt werden. Zu ihnen gehören die »Lernausgangslagenuntersuchung« (LAU) in Hamburg (R.H. LEHMANN/R. PEEK/R. GÄNSFUSS 1997 und R.H. LEHMANN/R. PEEK/R. GÄNSFUSS 1999), die »Qualitätsuntersuchung zum Unterricht an Schulen in Mathematik« (QuaSUM) in Brandenburg (R.H. LEHMANN/R. PEEK/R. GÄNSFUSS/S. LUTKAT/S. MÜCKE/I. BARTH 2000) oder die »Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterricht, Schulkontext« (MARKUS)<sup>3</sup>.

Während TIMSS im Rahmen der IEA von der Wissenschaft in Gang gesetzt worden ist und Bildungspolitiker und Bildungsverwaltung (auch wenn diese Finanzier der Untersuchung war) gewissermaßen unvorbereitet traf, ist die Bildungspolitik bei den genannten anderen Studien Auftraggeber und an der Projektvorbereitung und -durchführung selbst intensiv beteiligt. Inwieweit sie damit wirklich Herr des Geschehens ist, wird zu diskutieren sein. Der Entscheidung der Kultusministerkonferenz über DESI ging eine Begutachtung der Projektvorschläge durch die Deutsche Forschungsgemeinschaft (DFG) voraus. Sie diene einerseits der Unterstützung der politischen Entscheidungsfindung; andererseits kann sie Folgerungen in Richtung einer verstärkten Lehr-Lernforschung im Bereich der Deutsch- und der Fremdsprachendidaktik anbahnen. Folgerungen sind von der DFG aus TIMSS durch Einrichtung eines Schwerpunktprogramms »Die Bildungsqualität von Schule im mathematisch-naturwissenschaftlichen Unterricht« und die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK) durch ein Modellversuchsprogramm zur »Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts« bereits gezogen worden.

Die Untersuchungen befördern einander. PISA hat aus der TIMSS-Debatte gelernt. Im Vergleich zu TIMSS ist PISA das umfassendere und auch theoretisch besser fundierte Programm. DESI wendet sich, wie auch MARKUS, der Unterrichtsbeobachtung zu. Unterrichtsbeobachtung spielte in der ersten Welle von PISA keine Rolle, wird aber als nationale Ergänzung der zweiten Welle von PISA in Deutschland einen wichtigen Platz einnehmen. Eine wechselseitige Qualifizierung von Untersuchungen ist ein wichtiger und erwünschter Effekt im Gang der Entwicklung. Er wird umso wirkungsvoller sein, je offener und fundierter die Fachdiskussion nicht nur in der Wissenschaft, sondern auch in und mit den Schulen geführt werden kann. Zur Unterstützung dieser Diskussion hat die KMK die Herausgabe einer wissenschaftlichen Veröffentlichung angeregt und finanziert, in der Experten die Ansätze und die mögliche Tragweite von Schulleistungsuntersuchungen erörtern (F.E. WEINERT 2001).

3 Nähere Informationen unter: <http://www.rhrk.uni-kl.de/~zentrum/markus/markus.html>.

Die genannten Studien sind in der Regel als Querschnittuntersuchungen angelegt. Sie sollen den Leistungsstand der Schülerinnen und Schüler zu einem bestimmten Zeitpunkt erfassen. Dies ist bei PISA und DESI die neunte Jahrgangsstufe, d.h. grundsätzlich das Ende der Pflichtschulzeit in allen beteiligten Ländern<sup>4</sup>. LAU will nicht nur Leistungsstände punktuell, sondern auch Lernfortschritte innerhalb bestimmter Zeiträume messen. Zu diesem Zweck wird derselbe Altersjahrgang jeweils im Abstand von zwei Jahren (Anfang der fünften, der siebten, der neunten und der elften Klassenstufe) erfasst. DESI soll den Leistungsstand der Schülerinnen und Schüler in der neunten Klasse zu zwei Messzeitpunkten, am Anfang und am Ende des Schuljahres, und damit ebenfalls Lernfortschritte klären sowie der Frage nachgehen, welche Gründe für unterschiedliche Entwicklungen ursächlich sein könnten. Ein ähnliches Vorgehen ist in der zweiten Welle von PISA für den Mathematikunterricht geplant. Auch in TIMSS wurden Aspekte des Lernfortschritts in bestimmten Schuljahren untersucht (J. BAUMERT/R. LEHMANN u.a. 1997, S. 141ff; J. BAUMERT/R. WATERMANN 2000). Diese unterschiedlichen Ansätze nehmen den Leistungsbegriff in doppelter Perspektive auf: Es geht um den zu einem bestimmten Zeitpunkt erreichten Leistungsstand. Es geht aber auch um den innerhalb eines definierten Zeitraums erreichten Leistungszuwachs (»Leistung = Arbeit in der Zeit.«). »Leistung« ist dabei nur vor dem Hintergrund der unterschiedlichen Umfeld- und Ausgangsbedingungen der in die Untersuchung einbezogenen Teilnehmer angemessen zu bewerten.

### *Gesichtspunkte für die Auswahl von Untersuchungsbereichen*

Dass vor allem der Leistungsstand von Schülerinnen und Schülern in den Sprachen, in der Mathematik und in den Naturwissenschaften zum Gegenstand von Untersuchungen gemacht wird, hat seine Gründe. Sie sind nicht mit dem Argument zu erledigen, hier werde das Falsche und zudem rasch veraltendes Wissen abgefragt. Es geht nicht um Wissen, sondern um grundlegende Schlüsselqualifikationen, die Voraussetzung vielfältiger Bildungsprozesse ebenso wie der Teilhabe am gesellschaftlichen Leben sind. Wer kein ausreichendes Leseverständnis erwirbt, dem wird auch das selbstständige Lernen in Schule und Beruf nicht gelingen. Wer kein ausreichendes mathematisches Verständnis hat, bekommt nicht nur eine schlechte Note in Mathematik. Ihm fehlt vielmehr ein entschei-

4 In den aktuellen Beratungen der Kultusministerkonferenz wird die Frage der Schulleistungen für die verschiedenen Schulstufen auf unterschiedliche Weise thematisiert. Für die Sekundarstufe I spielen die Schulleistungsstudien, von denen im vorliegenden Text die Rede ist, eine besondere Rolle. Für die gymnasiale Oberstufe hat die Kultusministerkonferenz für die Bereiche Deutsch, Englisch und Mathematik Expertisen von Fachdidaktikern eingeholt, die derzeit Gegenstand der Beratung sind. Sie schließen an die Beratungen der sog. Oberstufenkommission von 1995 an und sollen zu gegebener Zeit veröffentlicht werden. Für die Grundschule beginnt die Diskussion mit dem Projekt PIRLS. Für die berufliche Bildung sind, soweit erkennbar, vergleichbare Ansätze kaum vorhanden. Die Diskussion um »Qualität« und Modernisierung berufsschulischer Angebote muss unter Berücksichtigung der für diesen Bereich spezifischen Kontextbedingungen (Lernortkooperation, Zusammenhänge zwischen staatlicher Verantwortung und Rolle der Wirtschaft usw.) geführt werden.

dender Zugang zum Verständnis der Welt, wenn es z.B. darum geht, komplexe Sachverhalte an Modellen rational zu interpretieren (nicht nur in den Natur- und Ingenieurwissenschaften: Steuertarif, Rentenformel) oder heuristische Strategien zu entwickeln (Induktion, Analogiebildung, Analysis-Synthesis-Prozeduren, Denken in Extremen usw.). Ebenso geht es in den Naturwissenschaften nicht um Faktenwissen, sondern um das Verständnis naturwissenschaftlicher Konzepte und Prozesse sowie um epistemologische Überzeugungen. In einer zunehmend »internationaler« werdenden Welt, deren Lingua franca das Englische ist, wird schließlich Fremdsprachenkompetenz gleichermaßen zur unerlässlichen Voraussetzung für Informationsgewinnung und Lernen (Internet!) wie für die Möglichkeit der Teilhabe im beruflichen und im privaten Bereich.

Ein wichtiger Effekt der derzeitigen Diskussion ist es, sich auf die Bedeutung des Lernens gerade in diesen Bereichen zu besinnen und sie in das Zentrum schulischer Anstrengungen zu rücken, und zwar im Grunde unabhängig von und vor den spezifischen Ergebnissen der eingeleiteten bzw. durchgeführten Untersuchungen. Damit soll nicht gesagt sein, dass es nur auf diese Aspekte ankommt. Weder sind damit die fachlichen Lernziele von Schule ausgeschöpft, noch beschränken sich die Lernziele auf diese fachlichen Aspekte. Natürlich geht es auch um Kritik- und Urteilsfähigkeit, um die Solidarität mit Schwächeren, um interkulturelle Kompetenz. Gerade diese letzteren Aspekte haben große Bedeutung für die Formulierung schulischer Bildungsziele. Dies kann freilich nicht heißen, dass man sich um die Leistungen von Schülerinnen und Schülern im Deutschen, im Englischen oder in der Mathematik nicht oder allenfalls nachrangig zu kümmern hätte.

Untersuchungen können allerdings durch die Einengung der Aufmerksamkeit auf bestimmte Ausschnitte des Fächerspektrums und Bereiche mit primär kognitiver Rationalität und scheinbar einfacher Messbarkeit eine hohe definitorische Kraft für die Formulierung von Lernprogrammen entfalten. Gegenwärtig erleben wir z.B. in Deutschland eine Diskussion, die sich auf Aspekte mathematischer und naturwissenschaftlicher Kompetenz konzentriert. DESI wird – so ist zu hoffen – die Diskussion ausweiten und die Aufmerksamkeit auch auf die Kommunikationsfähigkeit und Aspekte sprachlicher Kompetenz (Deutsch und Fremdsprache) lenken. Eine ähnlich fundierte Beschäftigung mit dem Problem historischer und politischer Bildung<sup>5</sup> oder mit der Situation des ästhetisch-expressiven Bereichs ist dagegen nicht zu registrieren. Man kann dies beklagen, aber man kann die beklagte Einseitigkeit des Vorgehens nicht mit dem Verzicht auf Untersuchungen schlechthin quittieren. Die öffentliche Frage nach dem Stand der Dinge in wichtigen Bereichen ist legitim. Die Formulierung des Bildungsauftrags von Schulen ist eine öffentliche Angelegenheit und letztlich Gegenstand politischer Diskussion und Entscheidung. Man muss nur dafür sorgen, dass alles zur Sprache kommt, was wichtig ist, und nach ge-

5 Eine international vergleichende Untersuchung der politischen Bildung Jugendlicher ist Gegenstand des von der IEA initiierten Projekts Civic Education, dessen Ergebnisse in Kürze veröffentlicht werden (J. TORNEY-PURTA/R.H. LEHMANN/H. OSWALD/W. SCHULZ 2001). Detailliertere Informationen zur ersten Phase dieser Untersuchung auch unter: <http://www.wam.umd.edu/~iea/>.

eigneten Formen der Klärung suchen, ob man erreicht, was man erreichen will. Bloße Spekulation reicht in keinem dieser Bereiche aus, weder im Bereich der Fachleistungen noch im Bereich überfachlicher Lernziele.

### *Welche Akzeptanz finden Vergleichsuntersuchungen?*

Der Meinungsstreit über die Sinnhaftigkeit breit angelegter Vergleichsuntersuchungen war in Deutschland lange Zeit bekanntlich besonders heftig. Inzwischen hat sich die Szene deutlich beruhigt. Die Hauptuntersuchung für PISA ist im Sommer 2000 weitgehend problemlos über die Bühne gegangen. Dies ist umso bemerkenswerter, als Deutschland sich nicht nur an der internationalen Untersuchung beteiligt. Die Kultusministerkonferenz hat vielmehr beschlossen, die Untersuchung durch eine entsprechende Ausweitung der Stichprobe so anzulegen, dass auch die Situation innerhalb der einzelnen Bundesländer hinreichend präzise abgebildet wird und Vergleiche zwischen Bundesländern möglich sind. Aus diesem Grund sind in Deutschland insgesamt etwa 1.600 Schulen und fast 60.000 Schülerinnen und Schüler in die Stichprobe einbezogen worden. Der Managementaufwand für die Durchführung einer so breit angelegten Untersuchung ist enorm. Dass die Dinge bisher reibungslos verlaufen sind, ist auch ein Verdienst der hohen Professionalität, mit der diese Untersuchung von dem verantwortlichen wissenschaftlichen Konsortium unter Federführung des Max-Planck-Instituts für Bildungsforschung in Berlin betrieben wird.

In der Auseinandersetzung um die Untersuchungen haben sich unterschiedliche Argumentationsmuster herausgebildet. Neben der Behauptung mangelnder Messbarkeit von Schulqualität und der Sorge vor den Risiken und Nebenwirkungen derartiger Untersuchungen, auf die an anderer Stelle kurz eingegangen wird, spielen insbesondere die Befürchtung einer mit den Untersuchungen intendierten grundsätzlichen Änderung der Steuerungsphilosophie im Schulwesen einerseits wie die Behauptung mangelnder Nützlichkeit andererseits eine besondere Rolle. Die erstere Befürchtung interpretiert Vergleichsuntersuchungen als Ausdruck eines pädagogisch schädlichen Konkurrenzdenkens, hinter dem letztlich die Vorstellung der Etablierung von Bildungsmärkten und eines Bildungswettbewerbs, wenn nicht gar einer Privatisierung von Bildungseinrichtungen stehe, wie sie im angelsächsischen Bereich in den beiden letzten Jahrzehnten als zentrale Innovationsstrategie verfolgt werden. Bezweifelt wird, dass Markt- und Wettbewerbsmodelle im Bildungswesen geeignet seien, über die Qualifizierung einzelner Schulen mit günstigen Rahmenbedingungen hinaus Innovation und Leistungssteigerung des Gesamtsystems zu bewirken. Befürchtet wird vielmehr, dass am Ende ein »Ranking« steht, welches Schulen mit ungünstigeren sozialen Kontextbedingungen einem unfairen Vergleich aussetzt, der negative Rückwirkungen auf ihre Motivation und ihre Leistungsfähigkeit zur Folge haben könnte. Das zweite, mehr pragmatische Argumentationsmuster der Kritiker von Vergleichsuntersuchungen bezieht sich darauf, dass nicht zu erkennen sei, wie der erhoffte Nutzen der Untersuchungen eintreten könne. Untersuchungen seien zumal angesichts des mit ihnen verbundenen Aufwandes nur zu rechtfertigen, wenn sie unmittelbar zur Verbesserung der Situation der einzelnen Schule bei-

tragen. Bezweifelt wird, dass die Bildungsverwaltungen in der Lage seien, die hierzu notwendigen Schritte systematisch zu entwickeln und sich den dafür erforderlichen Aufwand (intellektuell wie finanziell) zu leisten.

Wettbewerbsmodelle lassen sich für bestimmte Sektoren des Bildungswesens – Fort- und Weiterbildung Erwachsener, möglicherweise auch für die Hochschulen – sehr gut denken. Das Handeln der öffentlichen Träger und Finanziers beschränkt sich hier darauf, Rahmenbedingungen zu schaffen, innerhalb derer die Individuen ihre eigenen Bildungsinteressen verfolgen. Entsprechend begrenzt ist die Notwendigkeit, Bildungsziele normativ festzulegen. Qualitätssicherung in diesen Bereichen kann auf Deregulierung und Erhöhung der Handlungsfreiheit für die einzelnen Institutionen und ihre Träger setzen und das Handeln zentraler Instanzen auf den Verbraucherschutz und die Sicherung von Markttransparenz konzentrieren. Dies schließt eigene Anstrengungen der beteiligten Institutionen und Träger um Qualitätssicherung – auch im Interesse der Verbesserung ihrer Wettbewerbsposition – nicht aus.

Anders ist es dort, wo eine öffentliche Bildungspflicht begründet worden ist, der ein Bildungsrecht korrespondiert, und allenfalls eine begrenzte Freiheit der individuellen Standortwahl besteht, wie dies im allgemeinbildenden Schulwesen der Fall ist. Hier bezieht sich die Verantwortung des Staates sowohl darauf, Bildungsziele normativ zu klären, als auch qualitative Standards im Sinne gleicher Bildungs- und Entwicklungschancen für alle Kinder und Jugendlichen, unabhängig von der ökonomischen Lage und dem kulturellen oder sozialen Status der Herkunftsfamilien, institutionen- und trägerübergreifend zu sichern. Weder kann das Handeln den einzelnen Schulen im Sinne vollständiger Autonomie überlassen bleiben, noch ließe sich ein solches »reines Marktmodell« unter den in Deutschland gegebenen Bedingungen verwirklichen. Es fehlen elementare Voraussetzungen für eine sinnvolle Anwendung privatwirtschaftlich-ökonomischer Modelle auf den Schulbereich wie Freiheit der Preisbildung, Kundenautonomie, Haftung usw. (E. TERHART 2000b). Dies ist politisch gewollt, verfassungsrechtlich abgesichert (Art. 7 Abs. 1 GG) und auch im Zuge einer Weiterentwicklung struktureller Rahmenbedingungen allenfalls begrenzt zu verändern. Dies gilt umso mehr, als Untersuchungen darauf hindeuten, dass Marktmodelle zwar die Leistung einzelner Schulen verbessern können, damit zugleich aber die Leistungsdisparität innerhalb eines Systems vergrößern, ohne die Gesamtleistung des Systems im Sinne durchschnittlich besserer Lernerfolge für alle Schülerinnen und Schüler zu steigern. Soweit einzelne Schulen ihre Position verbessern, geschieht dies offenbar auch dadurch, dass sie die Zusammensetzung ihrer Schülerschaft verändern. Diese hat unter dem Aspekt »mitproduzierender Klienten« für die Ergebnisse einer Schule große Bedeutung (M. WEISS 1999; H. FEND 2000).

Auch vor diesem Hintergrund ist die Einräumung größerer Handlungsfreiheit für die einzelnen Institutionen und Träger mit dem Ziel, Bewegung in das Gesamtsystem zu bringen und situationsangemessene Lösungen von Problemen zu ermöglichen, denkbar und nötig. Das Bemühen um Profilbildung ist gewollt und trägt zu einem Wettbewerb zwischen Schulen bei, den es faktisch natürlich gibt. Öffentliches Handeln kann sich aber nicht darauf beschränken, das Geschehen zu beobachten. Es muss sich vielmehr sowohl bezüglich des Gesamtsystems als auch bezüglich der einzelnen beteiligten Institutionen der Er-



gebnisse des Handelns vergewissern und daraus Folgerungen für die bestmögliche Gestaltung der Lernbedingungen im Interesse aller Kinder und Jugendlichen ziehen. Dezentral und zentral ansetzende Vorgehensweisen müssen sorgsam ausbalanciert sein (H. FEND 2000). Gerade deshalb und nicht aus einem unangemessenen Markt- und Wettbewerbsdenken heraus ist neben der auf die einzelne Schule bezogenen Evaluation (»schulinterne« oder »schuleigene Evaluation«) auch eine Evaluation auf Systemebene unverzichtbar. Hierzu tragen sowohl die derzeit laufenden internationalen Vergleichsstudien als auch die spezifischen Untersuchungen in den einzelnen Bundesländern bei. Wer die Autonomie der einzelnen Schule fordert, müsste mithin auch für zentrale Verfahren der Evaluation eintreten. Die Ablehnung dieser Konsequenz führt in Wahrheit ebenfalls zu einem, freilich imperfekten, marktwirtschaftlichen Modell. Es ist imperfekt insofern, als es die wettbewerbstheoretischen Voraussetzungen und Konsequenzen dieses Modells nicht ausbuchstabiert und offen lässt, wo Anreizmechanismen liegen könnten, die das erhoffte Verhalten für alle – nicht nur für einige! – Akteure wahrscheinlich machen, wie auf diejenigen zu reagieren ist, die dem angenommenen Handlungsmuster nicht folgen und wie die Konsequenz einer zunehmenden Leistungsdisparität des Gesamtsystems vermieden werden kann.

Natürlich hängen die von Schülerinnen und Schülern erreichten Leistungen immer auch von Randbedingungen ab. Zu ihnen gehören die organisatorischen und institutionellen Regelungen in den einzelnen Bildungssystemen (z.B. die Frage der Ausstattung von Schulen, die Qualifizierung der Lehrkräfte oder die unterschiedlichen Ausprägungen lernförderlicher Milieus in Schulen) wie insbesondere auch die sozialen und kulturellen Kontexte, in denen Kinder und Jugendliche aufwachsen und lernen (der ökonomische Status der Herkunftsfamilien, ihr soziales und kulturelles Kapital, »Peers«, an denen sie sich orientieren usw.). PISA, DESI, PIRLS, LAU, MARKUS, QuaSUM: Sie alle verwenden einen hohen Aufwand auf die Klärung dieser Aspekte. Faire Vergleiche setzen ein sorgfältiges Vorgehen gerade auch in dieser Hinsicht voraus. Nur wenn dies berücksichtigt wird, lassen sich sinnvolle Schlussfolgerungen ableiten.

Es bleibt damit der pragmatische Einwand gegen Untersuchungen, der den damit erreichbaren Nutzen in Frage stellt. Dieses Argument kann sich darauf berufen, dass die Frage, wie die Ergebnisse von Untersuchungen in eine Veränderung der pädagogischen Praxis umgesetzt werden sollen, noch kaum systematisch beantwortet ist. Hier haben wir ganz unbestreitbar noch Klärungsbedarf. Dieses Argument steht deshalb im Mittelpunkt der folgenden Überlegungen.

### *Welches Ziel haben Vergleiche?*

Vergleichsstudien haben zunächst Unterschiede aufgedeckt und werden dies auch künftig tun. Sinn von Vergleichen ist es, diesen Unterschieden systematisch nachzugehen, die Ursachen für Unterschiede zu klären und zu fragen, welche Konsequenzen daraus zu ziehen sind. Es geht also weder um eine »Bildungsolympiade« noch um einen »Bildungswettbewerb«. Es geht um Lernen

voneinander und um Entwicklung. Dabei sind internationale Vergleiche deswegen wichtig, weil nur so Systeme mit hinreichend stark ausgeprägten Unterschieden in die Betrachtung einbezogen werden, die eine genauere Analyse des Beitrags einzelner Faktoren zu den jeweils erreichten Ergebnissen ermöglichen. Bei lediglich intranationalen Vergleichen wäre dies angesichts der bei allen Unterschieden doch immer auch gegebenen strukturellen Ähnlichkeiten der beteiligten Schulen nur begrenzt möglich. Vergleiche sagen dabei noch nichts über normative Zielsetzungen. Unterschiede können aufgrund unterschiedlicher Bildungskonzepte gewollt sein. Vergleiche können aber Anlass sein, über die eigenen Zielsetzungen erneut nachzudenken und sie gegebenenfalls zu revidieren.

Von besonderer Bedeutung für internationale Vergleichsuntersuchungen ist in diesem Zusammenhang, nach welchen Prinzipien die Aufgabenauswahl für die Untersuchungen erfolgt. Hierfür gibt es bisher zwei unterschiedliche Vorgehensweisen: Die Sicherung curricularer Validität auf Lehrplan- und – soweit möglich – Unterrichtsebene einerseits und einen normativ-didaktischen Ansatz andererseits. Im ersteren Falle versucht man, Stoffe zu identifizieren, die in möglichst vielen Ländern Gegenstand von Lehrplan und Unterricht sind, und dabei einen Kompromiss zwischen dem kleinsten gemeinsamen Nenner und einer möglichst breiten Erfassung einschlägiger Stoffgebiete zu finden. In TIMSS wurde dieses Vorgehen zur Sicherung interkultureller Fairness dahin gehend erweitert, dass nach Durchführung der Hauptuntersuchung eine Test-Curriculum Matching Analysis anhand einer national optimierten Testversion durchgeführt und die Stabilität der Rangreihen kontrolliert wurde (J. BAUMERT/O. KÖLLER/M. LEHRKE/J. BROCKMANN 2000; E. KLIEME 2000; J. BAUMERT/W. BOS/R. WATERMANN 2000). Der didaktisch-normative Ansatz unterstellt, dass bestimmte Grundqualifikationen gewissermaßen Teil einer kulturellen Aphabetisierung sind, deren Universalisierung und Unverzichtbarkeit Ausdruck einer Modernisierung ist, der sich alle beteiligten Länder in gleicher Weise zu stellen haben. Dieser Ansatz schließt erkennbar an die insbesondere im angelsächsischen Bereich geführte Literacy-Diskussion an, bei der der Gedanke der Universalisierung von Grundqualifikationen mit dem modernisierungstheoretischen Argument neuer und infolge der sich beschleunigenden Entwicklung zur Wissensgesellschaft steigender Qualifikationsanforderungen (derzeit insbesondere im mathematisch-naturwissenschaftlichen Bereich) verknüpft wird (E. KLIEME/J. BAUMERT/O. KÖLLER/W. BOS 2000). PISA legt einen normativ-didaktischen Ansatz sowohl für die Untersuchung der Lesekompetenz<sup>6</sup> als auch für die Klärung mathematischer<sup>7</sup> und naturwissenschaftlicher<sup>8</sup> Kompetenz zugrunde. Im

6 »Reading Literacy« wird in der Rahmenkonzeption definiert als die Fähigkeit, »geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potenzial weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen« (Rahmenkonzeption, S. 24).

7 »Mathematical Literacy« wird folgendermaßen definiert: »Mathematische Grundbildung ist die Fähigkeit einer Person, die Rolle zu erkennen und zu verstehen, 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 dieser Person als konstruktivem, engagiertem und reflektiertem Bürger entspricht« (Rahmenkonzeption, S. 47).

8 »Scientific Literacy« wird folgendermaßen definiert: »Naturwissenschaftliche Grundbil-

Rahmen der nationalen Ergänzungen werden bei PISA in Deutschland jedoch auch Aspekte curricularer Validität berücksichtigt.

Gründe für Unterschiede im Leistungsniveau der Schülerinnen und Schüler und Schulen in den beteiligten Ländern können auf unterschiedlichen Ebenen liegen. Diese sind wiederum für die Art von Interventionen wichtig.

- Von Bedeutung ist zunächst die Individualebene: Kinder und Jugendliche haben unterschiedliche Potenziale an Intelligenz und Begabung. Ebenso können die Lebenssituationen, in denen sie aufwachsen und lernen (der ökonomische Status oder die Bildungsnähe bzw. Bildungsferne ihrer Familie, die sprachlichen Kompetenz bei Migrantenkinder usw.), und die damit verbundenen Chancen, ihr Potenzial zu entfalten, unterschiedlich sein. Die Lernausgangslagenuntersuchung in Hamburg hat gezeigt, dass soziale Kontextbedingungen noch immer einen hohen Erklärungswert für die Varianz der Schulleistungen am Ende der Grundschule haben.
- Von Bedeutung ist zweitens die Ebene der einzelnen Schule: Auch bei gleichen sozialen Kontextbedingungen kann es erhebliche Unterschiede des Leistungsniveaus geben, wenn man einzelne Schulen oder auch Klassen innerhalb der Schulen miteinander vergleicht. Auch dies hat die Lernausgangslagenuntersuchung bestätigt. Es gibt sowohl erwartungswidrig leistungsstarke als auch erwartungswidrig leistungsschwache Schulen und es gibt große Leistungsunterschiede zwischen Klassen derselben Schule. Dies deutet auf Ursachenfaktoren im Handeln der Schulen und der Lehrkräfte hin: auf unterschiedlich qualifizierten Unterricht oder auf unterschiedliche Ausprägungen der oben angesprochenen Prozessmerkmale »guter Schulen«.
- Schließlich kommt es drittens auf die Systemebene an: Unterschiede, die sich in länderübergreifenden Vergleichen zeigen, können in den spezifischen Traditionen von Schule und Unterricht, in den durch die Lehrerbildung vermittelten Qualifizierungen, in normativen Vorgaben wie Lehrplänen und Stundentafeln, in Ausstattungsstandards oder in strukturellen Gegebenheiten des Schulsystems begründet liegen. Vermutlich werden auch kulturelle Unterschiede in der spezifischen Wertschätzung bestimmter Lernbereiche eine Rolle spielen. Eine in der Diskussion immer wieder gestellte Frage ist auch die nach der Bedeutung der spezifischen Selektivität eines Bildungssystems für das durchschnittliche Leistungsniveau wie für Spitzenleistungen.

Entsprechend unterschiedlich setzen auch Vergleichsuntersuchungen an. Sie zielen teils auf ein »System Monitoring«, teils auf konkretes Wissen über einzelne Schulen, teils dienen sie der grundlegenden Klärung von Zusammenhängen, über die man entweder noch nicht ausreichend Bescheid weiß oder bei denen es darum geht, gesichert erscheinende Aussagen erneut zu überprüfen bzw. zu klären, ob die Zusammenhänge in allen Ländern, die sich an der Un-

dung ist die Fähigkeit, naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen, die die natürliche Welt und die durch menschliches Handeln an ihr vorgenommenen Veränderungen betreffen« (Rahmenkonzeption, S. 66).

tersuchung beteiligen, gleichermaßen straff sind. Je nach Fragestellung und Ergebnis der Untersuchungen werden auch die Folgerungen und die angesprochenen Akteure unterschiedlich sein. Generell gilt, dass selbst solche Untersuchungen, die einen hohen Aufwand auf die Klärung von Ursachen für Unterschiede verwenden, in der Regel nicht schon »Rezepte« für die Lösung von Problemen liefern können. Häufig zeigen die Untersuchungen nur Problemzonen auf, in denen intensivere Ursachenforschung betrieben werden muss. Dabei wird der praktische Nutzen derartiger Untersuchungen umso größer und umso schneller erreichbar sein, je intensiver sich diese Untersuchungen auch einer spezifischen Unterrichtsforschung zuwenden. Im Rahmen von TIMSS erfolgte z.B. eine Unterrichtsforschung durch vergleichende Videostudien. Auch für DESI, MARKUS und die zweite Welle von PISA ist eine Unterrichtsforschung projektiert. Gut geeignet hierfür ist ein mehrperspektivisches Verfahren unter Einschluss von Videostudien und einer Befragung von Lehrkräften und Schülern. Eine derartige Forschung ist mit erheblichem Aufwand verbunden. Sie kann auf Widerstände stoßen, weil sie in den Schutzraum von Lehrerinnen und Lehrern einzudringen scheint. Indessen muss man das Dilemma überwinden, welches darin besteht, dass die zielgerichtete Bemühung um Verbesserung des Unterrichts eingeklagt, zugleich aber das Instrumentarium, das dazu beitragen könnte, abgelehnt wird. Dies alles muss man berücksichtigen, wenn man die Untersuchungen richtig bewerten und ihren möglichen Ertrag wie auch Risiken und Nebenwirkungen zutreffend einschätzen will. In der öffentlichen Diskussion geht hier leider vieles durcheinander.

*Unterschiedlicher »Auflösungsgrad« der gegenwärtig in Deutschland betriebenen Untersuchungen*

Entsprechend den genannten Unterscheidungen setzen die verschiedenen Untersuchungen, von denen hier die Rede ist, auf unterschiedlichen Ebenen an. Sie haben damit auch eine unterschiedliche Tiefenschärfe, die für die Verwendung des erreichten Wissens von Bedeutung ist.

- Internationale Vergleichsuntersuchungen wie TIMSS, PISA und PIRLS dienen dem Systemmonitoring und der Klärung grundsätzlicher Zusammenhänge, nicht der Steuerung der Einzelschule. Sie operieren mit Stichproben. Eine Stichprobenuntersuchung auf nationaler Ebene ist auch DESI. Auf regionaler Ebene ist QuaSUM diesem Untersuchungstyp zuzurechnen. Beabsichtigt sind primär Reaktionen auf Systemebene, und zwar in fachlicher wie in allgemein bildungspolitischer Hinsicht. Angesprochene Akteure sind insbesondere die Bildungsverwaltungen, aber auch die Institutionen der Wissenschaft. Die einzelne Schule wird sich dagegen nur im Ausnahmefall als Akteur angesprochen sehen. Es geht um Konsequenzen für die Curriculumsrevision und die Lehrplanarbeit, um die Entwicklung der Fachdidaktiken sowie die Entwicklung schlüssiger Modelle der Lehrgangs- und Unterrichtsplanung, um die Entwicklung von Programmen der Lehr-Lern-Forschung und um Nachwuchsförderung, um die Entwicklung der Lehreraus- und -fortbildung, um die Diagnose und die Möglichkeit des Ausgleichs

sozialer Unterschiede oder auch um die Notwendigkeit, bestimmte Schülergruppen in besonderer Weise zu fördern. Ergebnisse auf dieser Ebene versprechen die größte Nachhaltigkeit der Veränderungen. Aber sie fordern auch ihre Zeit. Insofern wäre es unrealistisch und schädlich, falsche Erwartungen über kurzfristig handgreifliche Veränderungen zu wecken.

- Flächendeckende Evaluationen wie LAU oder MARKUS erbringen ebenfalls ein Steuerungswissen auf Systemebene. Sie zielen vor allem aber auf ein Wissen über die einzelnen Schulen. Ein solches Wissen lässt sich selbst für die in Stichprobenuntersuchungen einbezogenen Schulen mit den zuvor genannten Projekten nicht erreichen, da die durch die Art der Stichproben bedingten Schätzfehler auf einzelschulischer Ebene aus statistischen Gründen beträchtlich sein können. Wer durch Untersuchungen dieses zweiten Typs als Akteur angesprochen wird, hängt von den getroffenen Vereinbarungen über die Ergebnisrückmeldung ab. Die Ergebnisrückmeldung wird in jedem Falle an die einzelne Schule gehen. Sie kann daneben aber auch an die Schulaufsicht gerichtet sein. Klassenbezogene Rückmeldungen werden oft nur an die betroffenen Lehrkräfte gehen. Ob die angesprochenen Akteure aufgrund der Ergebnisrückmeldungen auch wirklich agieren werden und wie wirksam die Aktionen sein werden, ist eine Frage, die weiter zu klären sein wird. Zu klären ist insbesondere, welche Art der Darstellung von Ergebnissen notwendig ist und welche Hilfestellung Schulen und Bildungsverwaltungen benötigen, um Ergebnisse »lesen« und aus ihnen Schlussfolgerungen ziehen zu können.
- Als drittes sind schuleigene oder schulinterne Evaluationen zu nennen, die sich nur auf die Situation einer einzelnen Schule beziehen. Sie können von Mitgliedern der betreffenden Schule oder in deren Auftrag von externen Evaluatoren betrieben werden und müssen sich nicht auf die Erhebung schulinterner Informationen und Sichtweisen beschränken. Schuleigene Evaluationen dienen der Reflexion des eigenen pädagogischen Handelns vor dem Hintergrund der diesem zugrunde liegenden spezifischen Zielsetzungen und Erwartungen. Diese sind in den schulgesetzlichen Vorgaben, den allgemein geltenden Bildungs- und Lehrplänen und den jeweiligen Schulprogrammen formuliert. Schuleigene Evaluationen sind ein unverzichtbares – und angesichts der Tatsache, dass die derzeitigen Großuntersuchungen wie PISA, DESI, LAU, QuaSUM und MARKUS nur Ausschnitte betreffen und allenfalls begrenzt wiederholbar sind – das wichtigste Element des Bemühens um eine Verbesserung der Lernchancen von Kindern und Jugendlichen. Aber auch für schuleigene Evaluationen spielen die Ergebnisse breiter angelegter Untersuchungen und das Urteil Außenstehender eine wichtige Rolle. Sie liefern Informationen darüber, was andere unter vergleichbaren Bedingungen erreichen, und bieten eine Außensicht z.B. in Form der Befragung von weiterführenden Schulen oder im Urteil von Absolventinnen und Absolventen aus der Rückschau. Schuleigene Evaluationen bedürfen des Blicks von und nach außen und damit der Ergänzung durch externe Evaluationen. Ohne diese wären die Ergebnisse kaum einzuordnen und zu bewerten. Auch schuleigene Evaluationen setzen Professionalität und ein hohes methodenkritisches Bewusstsein voraus. Sie sind keine einfache und schon gar keine bequeme Sache. Das Instrumentarium hierfür

und die Fähigkeit, es sachgerecht einzusetzen, entwickeln sich erst allmählich (G. KEMPFERT/H. ROLFF 1999; LEITFADEN 2000).

Von diesen Untersuchungsansätzen zu unterscheiden sind im übrigen standardisierte Test- oder Prüfungsverfahren, die der Zertifizierung von Individualleistungen dienen. Derartige Verfahren sind z.B. in den USA verbreitet, in Deutschland aber eher selten vertreten. Zu ihnen gehören schriftliche Prüfungen im Rahmen des Zentralabiturs, jedenfalls soweit die Auswertungsprozeduren festgelegt sind. Zentrale Prüfungen können hohe Bedeutung für die Sicherung der Vergleichbarkeit von Zertifizierungen und die Sicherung von Minimalstandards haben. Inwieweit sie auch der Qualitätssicherung im Sinne der Gewährleistung eines hohen Durchschnittsniveaus dienen und welche Nebenfolgen – beispielsweise im Sinne eines exzessiven »Teaching to the Test« oder der Behinderung curricularer und didaktischer Innovationen – mit ihnen verbunden sind, ist dagegen eine offene Frage. Auch hier könnten freilich nationale und internationale Vergleichsuntersuchungen etwa im Hinblick auf die Wirkungen unterschiedlicher Prüfungssysteme innerhalb Deutschlands oder im Vergleich Deutschlands mit seinen europäischen Nachbarn Klärungen herbeiführen und fruchtlose Glaubenskriege auf einer rein spekulativen Basis vermeiden helfen.

#### *Welche Ergebnisse liegen bereits vor?*

Der hier zu gebende Bericht kann keinen vollständigen Überblick über bereits vorliegende Erkenntnisse aus Untersuchungen liefern. Er muss sich mit einigen Andeutungen begnügen, die lediglich eine gewisse Anschaulichkeit vermitteln sollen. Einige Hinweise hierzu wurden auch bereits in den vorausgegangenen Bemerkungen gegeben. Für weitere Einzelheiten ist auf vorliegende Publikationen insbesondere zu TIMSS, LAU und QuaSUM zu verweisen. Die folgenden Hinweise beschränken sich im Übrigen verständlicherweise auf Stichprobenuntersuchungen und flächendeckende Untersuchungen im Sinne der obigen Einteilung. Für eine Übersicht zu Ergebnissen schulinterner oder schuleigener Evaluationsansätze – auch mit Hilfe externer Evaluateure – wäre eine eigene Untersuchung erforderlich.

- Eine wichtige Erkenntnis aus der ersten Stufe der Hamburger Lernausgangslagenuntersuchung ist z.B. die Tatsache, dass Leistungsunterschiede zwischen Grundschulen noch immer vor allem mit unterschiedlichen soziostrukturellen Rahmenbedingungen zu erklären sind. Die Hoffnung, die Unterschiede durch zielgruppenorientierte Förderprogramme in den Schulen kompensieren zu können, hat sich bisher offenbar nicht oder nur begrenzt erfüllt. Wir müssen also neu darüber nachdenken, wie man soziale Benachteiligungen im Bildungswesen ausgleichen kann. Von Bedeutung ist dabei nicht zuletzt die Situation jener Kinder, die in Familien nichtdeutscher Umgangssprache aufwachsen. Eine der Konsequenzen aus derartigen Überlegungen kann es sein, dass Förderangebote nicht erst und nur in der Schule einsetzen dürfen. Allerdings ist es eine offene Frage, welche Wirkungen die

in jüngerer Zeit eingeleiteten Maßnahmen zur gezielten Förderung sprachlicher und mathematischer Kompetenz in den Grundschulen haben werden. Möglicherweise ist deren Laufzeit noch zu kurz, um nachhaltige Effekte erwarten zu können (P. MAY 2001).

- Aus den Ergebnissen der zweiten Stufe der Lernausgangslagenuntersuchung in Hamburg läßt sich entnehmen, dass die Überlappung der Leistungsbereiche der unterschiedlichen Schulformen auch in einem gegliederten Schulwesen beträchtlich ist und dass diese Überlappung im Bereich der Mathematik stärker ausgeprägt ist als im Bereich der sprachlichen Kompetenz. Zugleich zeigt die Untersuchung, dass der Lernzuwachs in den beiden ersten Jahren der Sekundarstufe bei leistungsschwächeren Schülern größer ist als bei leistungsstärkeren. Offenbar konzentrieren sich die Bemühungen der Lehrkräfte mindestens in dieser Phase auf die Förderung der leistungsschwächeren Schüler. Es wäre freilich falsch, aus diesen Erkenntnissen unmittelbar auf die Notwendigkeit einer größeren Selektivität im Übergang von der Grundschule in die Sekundarstufe I zu schließen. Wichtiger könnte es sein, Modelle binnendifferenzierenden Unterrichts zu entwickeln, die dem festgestellten Leistungsspektrum besser gerecht werden, und Durchlässigkeit zu sichern.
- Die Befunde von TIMSS belegen, dass Staaten, die sich durch einen weniger restriktiven Zugang zu vorakademischen Bildungsgängen auszeichnen, kein Absinken des Leistungsniveaus zu verzeichnen haben. Das Wissen im Bereich der mathematisch-naturwissenschaftlichen Grundbildung ist in den Ländern am höchsten, in denen ein hoher Prozentsatz eines Geburtsjahrganges bis zum letzten Jahr der Sekundarstufe II im Schulsystem verbleibt. Auch die Befürchtung der Skeptiker, dass die zunehmende Zahl von Jugendlichen einer Alterskohorte, die eine allgemeine Hochschulreife anstreben, zu einer Beeinträchtigung möglicher Spitzenleistungen führt, wird durch die Untersuchungen nicht belegt. Es deuten sich im Gegenteil positive Auswirkungen einer verbreiterten Basis für die Erzielung von Spitzenleistungen an.
- TIMSS hat zudem gezeigt, dass deutsche Schülerinnen und Schüler insbesondere Schwierigkeiten haben, den mathematischen Gehalt von alltagsweltlichen Situationen zu erkennen und mit Hilfe mathematischer Modelle aufzuschlüsseln sowie die eingeübten (formalen) Rechenoperationen anzuwenden, wenn es darauf ankommt. Das mathematische Wissen deutscher Schülerinnen und Schüler ist offenbar träges Wissen. Es gibt Grund zur Annahme, dass die besseren Erfolge etwa der Holländer ihren Grund in der dort praktizierten didaktischen Konzeption der »Realistic Mathematics« haben. Aus vergleichenden Unterrichtsbeobachtungen kann man auch entnehmen, dass der deutsche Mathematikunterricht in seinem fragend-entwickelnden Charakter den Schülerinnen und Schülern anders als der japanische Mathematikunterricht wenig Spielraum für das Aufspüren eigener Lösungswege und damit für mathematische Kreativität läßt.

Die Interpretation dieser Ergebnisse und die Ableitung von Schlussfolgerungen und Maßnahmen mit dem Ziel einer Verbesserung der Situation ist keineswegs einfach. Im Rahmen des notwendigen mehrbenenanalytischen Vorgehens ist

jeweils genau zu fragen, welchen spezifischen Erklärungsbeitrag für Unterschiede Faktoren auf den verschiedenen Ebenen liefern, von denen oben die Rede war. Weder ist die Interpretation eines unbefriedigenden Abschneidens nach einem simplen Verfallsparadigma möglich, noch kann sich die Reaktion auf die in der Vergangenheit übliche Vermehrung des »Input« (Vermehrung der Zahl der Stunden, Senkung von Klassenfrequenzen usw.) beschränken. TIMSS besagt zunächst nicht, dass wir mehr und strengeren, sondern dass wir einen anderen Mathematikunterricht brauchen, der sich nicht auf die syntaktische Funktion der Mathematik beschränkt, sondern auch die semantische und die pragmatische Funktion mathematischer Symbolverarbeitung zu ihrem Recht kommen lässt. Die bisher vorliegenden Untersuchungen deuten darauf hin, dass die unterschiedliche Ressourcenausstattung und Stundentafelunterschiede (sofern es sich nicht um wirklich gravierende Sprünge handelt) keinen oder nur einen geringen Erklärungswert für Leistungsunterschiede haben. Es kommt offenbar weniger darauf an, wieviele Stunden gegeben werden als vielmehr darauf, was in den Stunden geschieht. Erst dann – dann aber sehr wohl – kann die Zeit ein Faktor sein. Bei alledem ist es gegenwärtig allemal ratsam, Schlussfolgerungen nur zurückhaltend zu formulieren. Wir stehen am Anfang einer Entwicklung, die einen erheblichen Wissenszuwachs erwarten lässt. Die derzeit laufenden Untersuchungen werden aufgrund der sich entwickelnden Untersuchungskonzepte und der verfügbaren Mengen von Vergleichsdaten Zusammenhänge besser klären können, als dies bisher der Fall war. Sie werden aber auch an der einen oder anderen Stelle scheinbar gesicherte Schlussfolgerungen in Frage stellen und viele neue Fragen und Probleme aufwerfen.

### *Wie werden Untersuchungsergebnisse »umgesetzt«?*

Zunehmende Informationen über Leistungsunterschiede und Wirkungsgrade erzeugen Verunsicherung und Abwehr. Das Wissen über Leistungsergebnisse und Leistungsvoraussetzungen von Schulen und Schulsystemen wächst offenbar schneller als das Wissen darüber, was man mit diesem Wissen anfangen kann und wie dieses in reale Verbesserungen der Situation auf den verschiedenen hier angesprochenen Ebenen »umzusetzen« ist (E. TERHART 2000c). Dies ist kein Hinweis auf die strategische Unschlüssigkeit des gewählten Vorgehens. Das Schulsystem kann sich dem Denken in empirisch kontrollierten Wirkungskategorien ebenso wenig entziehen wie andere Bereiche. Dies gilt für Schulen und Bildungspolitik gleichermaßen. Deutlich ist aber auch, dass es einen hohen Klärungsbedarf gibt, wie der Prozess zu denken ist, in dem aus Erkenntnissen eine neue Wirklichkeit wird.

Insbesondere die groß angelegten nationalen und internationalen Vergleichsuntersuchungen sollen, wie man zu sagen pflegt, »Steuerungswissen« vermitteln. Damit ist freilich die Frage noch nicht beantwortet, wie man sich eigentlich die »Steuerung« des Bildungswesens vorzustellen hat. Die Debatte hierüber wird in der Bildungsforschung, wie KLAUS KLEMM es jüngst formuliert hat, »erst neuerdings mehr als randständig geführt« (KLEMM 2000). Die Metapher »Steuerung« suggeriert ein eher mechanistisches Bild der Prozesse, um die es hier geht, so als ob jemand sich an das Steuer des Vehikels »Bildungswesens«



sen« setzen und durch einfaches Drehen an einem Lenkrad oder durch Druck auf das Gaspedal bestimmte Effekte erreichen könnte. Indessen sollte aus der verwaltungs- und politikwissenschaftlichen Implementations- und Policyforschung bekannt sein, dass die Steuerung eines komplexen sozialen Systems nach dem einfachen Muster von Regelung und Ausführung nicht funktioniert. In der Organisationsliteratur wird deshalb auch gelegentlich davon gesprochen, dass man komplexe soziale Systeme nicht »steuern«, sondern allenfalls »segeln« könne (J.G. MARCH 1980). Dieses Bild macht deutlich, dass – wer ein Ziel erreichen will – nur begrenzt Herr der Antriebskräfte ist, dass er sehr genau Wind und Wetter beachten und gelegentlich auch kreuzen muss. Für das Bildungswesen gilt zudem, dass wir es nicht nur mit einem einzelnen Schiff, sondern mit einer ganzen Flotte zu tun haben, die auch noch in unterschiedlichen Gewässern operiert. Selbst wenn es einen tüchtigen Flottenchef gibt, müssen die Kapitäne und Mannschaften an Bord der einzelnen Schiffe ihre eigene Professionalität einsetzen: Auch sie müssen die Seekarten studieren, die Informationen der Meteorologen abhören, Strömungen berechnen und darauf ihre eigenen Entscheidungen gründen. Wenn wir dieses Bild benutzen, so ist Adressat der Erkenntnisse aus Vergleichsuntersuchungen nicht allein die Bildungsverwaltung. Weder erkennt sie unbedingt immer die Anlässe für Untersuchungen, noch ist sie ohne weiteres kompetent, deren Fragestellungen selbst sachkundig zu formulieren. Ebenso wenig ist sie allein Herr über die durch die Untersuchungsergebnisse ausgelösten Steuerungsimpulse. Sie ist auf die Seekarten und die Wettervorhersagen angewiesen, die ihr die Wissenschaft liefert, sie muss Rücken- und Gegenwinde des öffentlichen Diskurses beachten und sie muss darauf vertrauen, dass die Ergebnisse der Untersuchungen auf den verschiedensten Ebenen auch unabhängig davon rezipiert werden, ob und inwieweit sie systematisch in förmliche Programme umgesetzt werden. Die Idee des großen Plans, der sich um PISA, DESI und PIRLS gruppiert, kann hier so wenig funktionieren wie in anderen gesellschaftlichen Bereichen. Ein solcher Plan wäre notwendigerweise von nicht handhabbarer Komplexität und alsbald durch nicht einkalkulierte Umstände überholt.

Zu erinnern ist dabei an den bereits oben erwähnten Sachverhalt, dass die Studien zu Schulleistungen weder schon die »Rezepte« für Interventionen liefern noch bereits alle Zusammenhänge aufklären, an die die »Rezeptentwicklung« anzuschließen hätte. Allenfalls ansatzweise geklärt sind z.B. die Strukturen und Entwicklungsverläufe des Wissenserwerbs, die Zusammenhänge zwischen fachlichem und überfachlichem Lernen oder zwischen Emotion, Motivation und Interesse, die Voraussetzungen des Transfers von schulisch erworbenem Wissen in außerschulische Anwendungssituationen, die Wirkungen unterschiedlicher Unterrichtsformen aufeinander und ihre Interaktion mit Schülermerkmalen (R. PEKRUN 2000). Die Untersuchungsergebnisse müssen deshalb Impulse geben, die zu Aktivitäten an den verschiedensten Stellen führen: Zu weiterführenden und ergänzenden Forschungen, zur Entwicklung neuer Unterrichtsmodelle und Lehrpläne, zu einem veränderten Handeln von Schulen und Bildungsverwaltung, zur Qualifizierung der Evaluationsansätze auf den verschiedenen Ebenen, zur Veränderung der Aus- und Fortbildung von Lehrkräften usw. Angesprochen sind damit viele Akteure. Bildungspolitik und Bildungsverwaltung dürfen in diesem Prozess freilich nicht nur ein Spielball

der Elemente sein. Sie haben eine eigene Rolle zu spielen und dafür zu sorgen, dass die Entwicklungen, wo immer dies nötig ist, angestoßen und vorangetrieben werden, dass Folgerungen gezogen werden, wo dies nicht in ausreichender Weise im Wege der professionellen Selbstorganisation geschieht und dass Fehlentwicklungen und falsche Akzentsetzungen stets auch korrigiert werden. Aber sie müssen auch die Grenzen ihrer Möglichkeiten erkennen und das Einverständnis begrenzter Einflussmöglichkeiten auf Entwicklungen aushalten, deren Anfang und Ende sie nicht genau übersehen. Bildungspolitik und Bildungsverwaltung erhalten durch Untersuchungen wichtige Basisdaten für die Bildungsplanung. Aber sie können das Geschehen innerhalb des Klassenzimmers oder auch die Reaktion des Wissenschaftssystems auf Untersuchungsergebnisse kaum unmittelbar beeinflussen. Es geht deshalb auch darum, Rezeption zu fördern und Evolution zu ermöglichen, nicht aber im klassischen Schema bürokratischer Systeme nach dem Muster von Anordnung und Ausführung zu agieren. Werden die Beteiligten die notwendige Beharrlichkeit und den erforderlichen langen Atem aufbringen? Oder werden wir am Ende doch nur das Strohfeuer einer politischen Kampagne erleben, bei der es um symbolische Akte geht: Eine Stunde im Fach X mehr in der Stundentafel hilft manchmal und beruhigt immer!? Die Rolle, die Bildungsverwaltung und Bildungspolitik zu spielen haben, ist nicht leicht, gerade wenn es sich, um das Bild noch einmal aufzugreifen, nicht um »Steuern«, sondern um »Segeln« handelt.

Klärungsbedarf besteht auch hinsichtlich möglicher Reaktionsweisen auf der Ebene der einzelnen Schule. Auch darauf wurde bereits hingewiesen. Welche Art der Darstellung von Ergebnissen ist notwendig und welche Hilfestellung benötigen Schulen, um Ergebnisse »lesen« und aus ihnen Schlussfolgerungen ziehen zu können, und wie nachhaltig wird die Wirkung derartiger Berichte sein? Welche »Übersetzungsarbeit« müssen die Fachdidaktiken in den einzelnen Disziplinen leisten und werden sie dazu bereit und in der Lage sein? Gibt es Zusammenhänge zwischen dem je spezifischen Stand der Diskussion über Schulentwicklung und Schulprogramm einerseits und der Fähigkeit, Schulrückmeldungen zu lesen und aus ihnen Folgerungen zu ziehen, andererseits? Werden alle Schulen fähig sein, wirksame Vorgehensweisen zu wählen oder muss man befürchten, dass zwar einige es können, andere aber nicht? Wird dies zu zusätzlichen Disparitäten im Gesamtsystem führen und was müsste geschehen, um dies zu verhindern? Wie kann insbesondere eine Schule mit »erwartungswidrig schlechten« Ergebnissen klären, was eine Schule, die unter gleichen sozialen Umfeldbedingungen bessere Ergebnisse erreicht, anders macht? Sie bedürfte vermutlich externer Unterstützung und des Rückgriffs auf Systemwissen, das nicht notwendigerweise mit der jeweiligen Untersuchung selbst schon geliefert wird. Wer könnte die notwendige Unterstützung geben? Sollten es unabhängige Experten sein oder kann und muss auch die Schulaufsicht eine Rolle spielen? Wäre sie dazu in der Lage bzw. wie könnte sie dafür qualifiziert werden? Lassen sich aus den Untersuchungsinstrumenten der großen Untersuchungen auch »Tools« entwickeln, mit deren Hilfe Schulen, die nicht Bestandteil der Untersuchungsstichproben sind, aber selbst größere Klarheit über ihre Situation erhalten möchten, Erkenntnisse gewinnen können?

Hier liegt ein Feld der Bildungsforschung, die sich der Klärung der Steuerungs- und Transferprobleme mehr als bisher annehmen muss. Es darf nicht

bei der »randständigen Diskussion« der Probleme bleiben, von der KLAUS KLEMM gesprochen hat. Gesucht ist ein neuer Forschungstyp, der Analyse und Intervention längsschnittlich verknüpft, der aber nicht einfach als angewandte oder entwicklungsorientierte bzw. als Aktions- oder Transferforschung zu bezeichnen ist (E. TERHART 2000c). Denn es geht nicht allein darum, Konzepte und Programme zu transferieren, sondern durch diesen Transfer bei den Beteiligten »Transformationen« auszulösen und weiterzuführen sowie dabei insbesondere auch die komplizierten Beziehungen zwischen der einzelnen Schule als handelnder Einheit und zugleich als Subsystem eines übergreifenden Zusammenhanges in den Blick zu nehmen. Ein Anfang ist dadurch gemacht, dass die Deutsche Forschungsgemeinschaft die Klärung zweckmäßiger Formen der Ergebnismeldung und der Gelingensbedingungen von Rezeptionsprozessen in ihr Schwerpunktprogramm zur Verbesserung der Qualität des mathematisch-naturwissenschaftlichen Unterrichts einbezogen hat.

Die Notwendigkeit, empirische Bildungsforschung in Deutschland zu verstärken, führt naturgemäß auch zu der Frage, ob die Voraussetzungen hierfür im deutschen Wissenschaftssystem ausreichend gegeben sind. Dies gilt für die Lehr-Lernforschung aus psychologischer Perspektive ebenso wie für Unterrichtsforschung, fachdidaktische Forschung oder vergleichende Evaluationsstudien auf der Basis quantitativer sozialwissenschaftlicher Forschungsansätze. Die von HEINRICH ROTH 1962 propagierte »realistische Wende« hat sich offenbar gegenüber der starken Tradition geisteswissenschaftlich-hermeneutisch (und reformpädagogisch) ausgerichteter Erziehungswissenschaften in Deutschland noch nicht ausreichend durchsetzen können. Dies wird eine Ursache dafür sein, dass die hier genannten Studien bisher letztlich von einem verhältnismäßig kleinen Kreis von Wissenschaftlern getragen werden, die in unterschiedlicher Zusammensetzung die wissenschaftlichen Konsortien dieser Untersuchungen bilden. Um die Entwicklung voranzubringen, ist mithin auch eine Nachwuchsförderung nötig. Sie muss Wissenschaftler qualifizieren, die die künftigen Untersuchungen tragen und vor allem auch in der Lehrerbildung Qualifikationen vermitteln können, welche ihrerseits unerlässliche Voraussetzung für die Rezeption der Untersuchungsergebnisse ist. Damit soll eine geisteswissenschaftlich ausgerichtete Pädagogik nicht verdrängt, wohl aber durch sozialwissenschaftlich orientierte Methoden ergänzt werden. Beide zusammen erst ergeben das vollständige Bild.

### *Risiken und Nebenwirkungen*

Zu sprechen wäre ferner von den Risiken und Nebenwirkungen eines Vorgehens, wie es hier skizziert wurde. Auch davon war schon in Andeutungen die Rede: Von der hohen definitorischen Kraft der Untersuchungen, die zu einer gefährlichen Verengung von Bildungszielen führen kann, von den Gefahren eines »Teaching to the Test«, von der Behinderung von Innovationen durch das Festhalten am testsicher Bewährten, von möglichen Fehlreaktionen seitens der Schulen und der Bildungsverwaltungen, von ungerechten Vorwürfen gegen Schulen, die wegen ungünstiger Umfeldbedingungen Konkurrenz Nachteile gegenüber anderen haben. Dies alles will sorgfältig bedacht sein, wenn nicht aus

dem Bemühen um Qualitäts- und Standardsicherung Unvernunft entspringen soll. Gefeit gegen Fehlentwicklungen sind wir nicht. Die Hände in den Schoß zu legen, wo Handlungsbedarf besteht, kann dennoch nicht das Mittel der Wahl sein.

Dass Leistungsvergleiche Unterschiede aufdecken werden, deren Ursache insbesondere auch in den unterschiedlichen Lebensbedingungen von Kindern und Jugendlichen und den unterschiedlichen sozialen Kontextbedingungen der Organisation schulischen Lernens liegt, ist ebenfalls schon gesagt worden. Ziel aller Bemühungen muss es deshalb sein, eine erneute Anstrengung zum Ausgleich sozialer Unterschiede und Benachteiligungen zu unternehmen. Es ist zu hoffen, dass wir auf diesem Wege dem bildungspolitisch fundamentalen Ziel der Sicherung von Chancengleichheit einen großen Schritt näherkommen. Dennoch muss man realistisch sehen, dass die festgestellten Unterschiede jedenfalls kurz- und mittelfristig nur in Grenzen auszugleichen sein werden. Auch mit dieser Erkenntnis müssen wir leben. Dies ist vielleicht am schwersten auszuhalten.

Zu sprechen wäre schließlich von der Form der Präsentation von Untersuchungsergebnissen und ihrer Rezeption durch Medien und Öffentlichkeit. Blicke es bei der oft üblichen Sensationsberichterstattung, bei der am Ende doch nur das »Ranking« (international wie national) und nicht der differenzierte Befund zählt, wäre wenig gewonnen. Triumphgeschrei der einen und Rechtfertigungsversuche der anderen Seite könnten das Interesse an der nötigen Entwicklung und die dazu erforderliche Kraftanstrengung blockieren, ehe sie sich richtig entfalten. Was besagte freilich die kleine Genugtuung im innerdeutschen Vergleich, wenn sich herausstellen sollte, dass wir alle miteinander unsere Probleme weniger gut gelöst haben als Holland, die Schweiz oder die skandinavischen Länder? Die differenzierte Berichterstattung anlässlich der gerade jetzt erfolgten Vorstellung der detaillierten Ergebnisse von TIMSS durch das Max-Planck-Institut für Bildungsforschung lässt allerdings auch insoweit durchaus hoffen.

### *Vergleichsuntersuchungen und kein Ende?*

Natürlich wird die derzeitige Welle von Vergleichsuntersuchungen mit hohem Aufmerksamkeitswert in der Öffentlichkeit irgendwann durch eine andere Welle abgelöst werden. Zu hoffen ist indessen, dass das, was jetzt als »empirische Wende« der Bildungspolitik in Gang gekommen zu sein scheint, auch dann gewissermaßen im Untergrund weiterwirkt und die Standards künftigen professionellen Handelns bestimmt. Käme es so, dann hätte sich gelohnt, was jetzt geschieht.

### *Literatur*

BAUMERT, J./BOS, W./LEHMANN, R. (Hrsg.): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn Bd. 1. (Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit) sowie Bd. 2 (Mathematische und Physikalische Kompetenzen am Ende der gymnasialen Oberstufe). Opladen 2000.

- BAUMERT, J./BOS, W./WATERMANN, R.: TIMSS/III. Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich. Zusammenfassung deskriptiver Ergebnisse. Berlin 1998.
- BAUMERT, J./BOS, W./WATERMANN, R.: Fachleistungen im voruniversitären Mathematik- und Physikunterricht im internationalen Vergleich. In: J. BAUMERT/W. BOS/R. LEHMANN (Hrsg.): TIMMS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2: Mathematische und Physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen 2000, S. 129–180.
- BAUMERT, J./KÖLLER, O./LEHRKE, M./BROCKMANN, J.: Untersuchungsgegenstand, allgemeine Fragestellungen, Entwicklung der Untersuchungsinstrumente und technische Grundlagen der Studie. In: J. BAUMERT/W. BOS/R. LEHMANN (Hrsg.): TIMMS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2: Mathematische und Physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen 2000, S. 27–56.
- BAUMERT, J./LEHMANN, R./LEHRKE, M./SCHMITZ, B./CLAUSEN, M./HOSENFELD, I./KÖLLER, O./NEUBRAND, J.: TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen 1997.
- BAUMERT, J./WATERMANN, R.: Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. In: J. BAUMERT/W. BOS/R. LEHMANN (Hrsg.): TIMMS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2: Mathematische und Physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen 2000, S. 317–372.
- DEUTSCHES PISA-KONZORTIUM (Hrsg.): Schülerleistungen im internationalen Vergleich. Eine neue Rahmenkonzeption für die Erfassung von Wissen und Fähigkeiten. (Übersetzung der englischen bzw. französischen Originalfassung). Berlin 2000 (Rahmenkonzeption).
- FEND, H.: Qualität und Qualitätssicherung im Bildungswesen. Wohlfahrtsstaatliche Modelle und Marktmodelle. In: Zeitschrift für Pädagogik, 41. Beiheft 2000, S. 55–72.
- KEMPFERT, G./ROLFF, H. G.: Pädagogische Qualitätsentwicklung. Ein Arbeitsbuch für Schule und Unterricht. Weinheim und Basel 1999.
- KLEMM, K.: Internationaler Bildungswettbewerb. Stellungnahme zu Fragen der Arbeitsgruppe »Künftige Struktur der Lehrerbildung« des Wissenschaftsrates. November 2000 (unveröffentlichtes Manuskript).
- KLIEME, E.: Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In: J. BAUMERT/W. BOS/R. LEHMANN (Hrsg.): TIMMS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2: Mathematische und Physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen 2000, S. 57–128.
- KLIEME, E./BAUMERT, J./KÖLLER, O./BOS, W.: Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In: J. BAUMERT/W. BOS/R. LEHMANN (Hrsg.): TIMMS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen 2000, S. 58–133.
- LEHMANN, R. H./PEEK, R./GÄNSFUSS, R.: Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996. Hamburg 1997.
- LEHMANN, R. H./PEEK, R./GÄNSFUSS, R.: Aspekte der Lernausgangslage und der Lernent-

- wicklung von Schülerinnen und Schülern an Hamburger Schulen – Klassenstufe 7 – Bericht über die Untersuchung im September 1998. Hamburg 1999.
- LEHMANN, R. H./PEEK, R./GÄNSFUSS, R./LUTKAT, S./MÜCKE, S./BARTH, I.: QuaSUM – Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. (In: Schulforschung in Brandenburg. Heft 1), Potsdam 2000.
- Leitfaden: Behörde für Schule, Jugend und Berufsbildung Hamburg: Schulinterne Evaluation. Ein Leitfaden zur Durchführung. Hamburg 2000.
- MARCH, J. G.: The Technology of Foolishness. In: H.J. LEVITT/L.R. PONDY/D.M. BOJE (Hrsg.): Readings in Managerial Psychology. Chicago (University of Chicago Press), <sup>3</sup>1980, S. 576.
- MAY, P.: Lernförderlichkeit im schriftsprachlichen Unterricht. Ergebnisse der Evaluation des Projekts »Lesen und Schreiben für alle« (PLUS). Hamburg 2001 (unveröffentlichtes Manuskript).
- MISLEVY, J. R.: What can we learn from international assessments? In: Educational Evaluation and Policy Analysis. 17 (1995), S. 419.
- PEKRUN, R.: Vergleichende Evaluationsstudien zu Schülerleistungen: Konsequenzen für zukünftige Bildungsforschung, Expertise für den PISA-Beirat der KMK. Januar 2001 (unveröffentlichtes Manuskript).
- ROTH, H.: Die realistische Wendung in der Pädagogischen Forschung. In: Neue Sammlung 2 (1962), S. 481–490.
- TERHART, E.: Zwischen Aufsicht und Autonomie. Geplanter und ungeplanter Wandel im Bildungsbereich. In: Neue Sammlung, 40 (2000), 1, S. 123–140 (TERHART 2000 a).
- TERHART, E.: Qualität und Qualitätssicherung im Schulsystem. Hintergründe – Konzepte – Probleme. Zeitschrift für Pädagogik 46 (2000), Heft 6, S. 809–829 (TERHART 2000 b).
- TERHART, E.: Wie können die Ergebnisse von vergleichenden Leistungsstudien systematisch zur Qualitätsverbesserung in Schulen genutzt werden? Expertise für den Beirat des PISA Projekts. November 2000 (unveröffentlichtes Manuskript; TERHART 2000 c).
- TORNEY-PURTA, J./LEHMANN, R.H./OSWALD, H./SCHULZ, W.: Citizenship and Education in Twenty-eight Countries: Civic Knowledge and Engagement at Age Fourteen. Amsterdam 2001 (im Druck).
- WEINERT, F. E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim und Basel 2001.
- WEISS, M.: Mehr Effizienz im Schulbereich durch dezentrale Ressourcenverantwortung und Wettbewerbssteuerung? In: Recht der Jugend und des Bildungswesens (RdJB) (1999) 4, S. 413–423.

*Anschrift des Autors*

Staatsrat Hermann Lange, Hamburger Straße 31, 22083 Hamburg