

tengemeinschaft. Leistungsmessungen erscheinen nun im Weltmaßstab als Angelegenheit eines ‚exklusiven Clubs‘ reicher Industrienationen. Solche Messungen wären dann von den OECD-Staaten entweder explizit oder stillschweigend als selektive Instrumente eingesetzt, um sich in der Weltgesellschaft bildungsbezogene sowie letztlich wirtschaftliche Vorteile zu erobern.

Diese Perspektive stellt die vermeintlich neutralen Leistungsmessungen als semantische Konstrukte der Industrienationen dar, welche zudem handlungswirksame Konsequenzen haben. Erstens werden sich innerhalb einer nationalen Arena, auf der die Konstrukte als Leitideen verbreitet werden, entsprechende Modernisierungseliten ausmachen lassen, die sich selbst durch das Konstrukt dynamisieren – und außerdem auch die Akteurkonstellation in der nationalen Arena, insofern sich bezüglich der Tests Befürworter, Gegner, Zögerliche und Konservative unterscheiden lassen. Zweitens bewirkt die Semantik eine Dynamisierung von Nationen als Akteuren der *Weltgesellschaft*. Auf dieser supranationalen Arena lassen sich OECD-Staaten und gleichsam „der Rest der Welt“ unterscheiden, und auch hier lässt eine der möglichen Forschungsperspektiven nach der Dynamisierung der Akteurkonstellation fragen, die zweifelsohne mit den internationalen Leistungsmessungen gegeben ist.<sup>46</sup>

## HERMANN LANGE Qualitätssicherung in Schulen

Es ist immer gut, die Realitäten so in den Blick zu nehmen, wie sie wirklich sind. In der Bildungspolitik besteht dazu heute mehr denn je Anlaß. Vor ungefähr 30 Jahren hatte Heinrich Roth die empirische Wendung der pädagogischen Wissenschaften gefordert. Sie gilt es, so könnte man auch sagen, endlich zu vollziehen, und zwar auch in der Bildungspolitik und im Schulalltag. [...]

### Was ist neu an der „empirischen Wende“?

Methoden und Instrumente, die zur Klärung des Erreichens der jeweils verfolgten Ziele eingesetzt werden, sind nicht neu. Neu ist indessen, daß diese Instrumente und Methoden in breiter Form genutzt werden sollen, um die Ergebnisse des konkreten Handelns systematisch zu überprüfen. Dies wird große Anstrengungen erfordern, und zwar nicht nur weil es immer schwer

46 Für einzelne Nationen wäre zu erforschen, inwieweit sich eine solche Dynamisierung und Modernisierung nur auf die Ebene des „talk“ bezieht, insofern vielleicht an Leistungsmessungen teilgenommen und über Testergebnisse auch gesprochen wird, das Schulsystem von der Struktur her aber belassen bleibt. Diesen Verdacht gibt es, wie bereits erwähnt, für den Akteur Deutschland.

ist, einen vielfältigen Einflüssen und Interessen unterliegenden Großbereich überhaupt in Bewegung zu bringen. Vielmehr sprechen auch verbreitete Gewohnheiten des gegenwärtigen Wissenschaftsbetriebes dagegen. Zwar ist die Wendung der Erziehungswissenschaften zu einer (auch) empirisch-sozialwissenschaftlichen Orientierung seit langem verkündet, insgesamt aber noch kaum vollzogen. Die Frontlinie der Auseinandersetzung um Eignung und Nebenwirkungen der Instrumente verläuft deshalb heute bemerkenswerterweise nicht zwischen „A-Ländern“ (SPD-regiert) und „B-Ländern“ (CDU-regiert), sondern quer durch die Profession selbst. Insofern geht es nicht nur um die empirische Wende in der Bildungspolitik, sondern noch immer auch um die empirische Wende der pädagogischen Wissenschaften. Das Bemühen, Folgerungen daraus zu ziehen, stößt freilich auch auf die Schwierigkeit, dass die erziehungswissenschaftlichen Fakultäten vielfach nicht über die Stellen (Lehrstühle, Institute) verfügen, die sich kompetent mit den quantitativen Methoden empirischer Forschung befassen könnten. Daraus resultieren Defizite in der Lehre, die die Rezeption der Ergebnisse empirischer Untersuchungen durch die schulische Praxis erschweren.

Neu ist auch eine Haltung, die das Dogma von der Unfehlbarkeit des Pädagogen aufhebt, solange er nur im rechten Geist handelt. Man muß zugeben lernen und zugeben können, daß auch gute Absichten fehlschlagen können, daß nicht jedes gute Ziel auch erreicht wird und daß die Gründe hierfür nicht bereits mit der heute vielfach üblichen pädagogischen Verelendungstheorie (gesellschaftliche Entwicklungen, fehlende Mittel etc.) abschließend beschrieben sind. Das Nickerreichen von Zielen, das bisher als Folge nicht veränderbarer Umstände allzu leicht aus dem Verantwortungsbereich der handelnden Personen herausgeschoben wird, als individuelles Scheitern hinter verschlossenen Türen verborgen bleibt, von der einzelnen Lehrkraft als persönlich belastendes Schicksal verarbeitet werden muß oder durch Gelingensgemeinschaften und Verbrüderungen (der Lehrer mit den Schülern, der Kollegen untereinander und mit den Schulleitungen) verschleiert wird, kann offen kommuniziert werden. Ehrlichkeit gegen sich selbst kann freilich schmerzlich sein. Sie ist ohne Frage risikoreich. Die Veränderungen werden Ängste und Widerstand auslösen. Sie können aber auch als Chance der (individuellen wie gemeinsamen) Befreiung und Erleichterung begriffen und als Möglichkeit genutzt werden, das Bildungswesen wieder in das Zentrum öffentlicher Aufmerksamkeit und öffentlichen Interesses zu rücken.

#### **Leistungsmessung und Leistungsvergleiche, national und international: TIMSS, LAU und PISA**

Gegenwärtig haben auf nationaler wie auf internationaler Ebene insbesondere vergleichende Untersuchungen zum Leistungsstand von Schülerinnen und Schülern Konjunktur. Das bekannteste und derzeit viel diskutierte Bei-

spiel einer internationalen Vergleichsuntersuchung ist die „*Dritte Internationale Mathematik- und Naturwissenschaftsstudie*“ (The Third International Mathematics and Science Study – TIMSS), die unter dem Dach der International Association for the Evaluation of Educational Achievement (IEA) durchgeführt wurde. Die Ergebnisse dieser Untersuchung dürfen als bekannt vorausgesetzt werden, so daß sie hier nicht im einzelnen referiert werden müssen (vgl. Baumert/Lehmann 1997; Baumert u.a. 1998).

Als ein wichtiges Beispiel für derartige Untersuchungen in Deutschland ist ferner die „Lernausgangslagenuntersuchung“ (LAU) zu nennen, mit deren Hilfe in Hamburg flächendeckend der Leistungsstand der Schülerinnen und Schüler am Beginn der Sekundarstufe I, d.h. letztlich am Ende der Grundschule, erfaßt wurde. Untersucht wurden dabei Kenntnisse, Fertigkeiten und Fähigkeiten in den Bereichen Sprachverständnis, Leseverständnis, Rechtschreibung, Mathematik und Informationsentnahme aus Karten, Tabellen und Diagrammen, die Fähigkeit zu freier und kreativer Text- und Bildgestaltung unabhängig von der Rechtschreibkompetenz, die Fähigkeit zu schlußfolgerndem Denken, insbesondere auch die Fähigkeit, komplexe Beziehungen in neuartigen Situationen wahrzunehmen, und zwar weitgehend unabhängig von dem Grad der Beherrschung der deutschen Sprache. Erfaßt wurden zugleich die unterschiedlichen familiären Hintergründe und Schülermerkmale wie Staatsangehörigkeit, Schullaufbahn, Besuch einer VorschulkLASse etc. sowie die Einstellungen der Kinder zum schulischen Lernen.

Die *Ergebnisse* der Untersuchung lassen sich – sehr vereinfacht – in folgenden Thesen zusammenfassen (vgl. Lehmann u.a. 1996):

(1.) Es gibt deutliche regionale Unterschiede im Leistungsniveau der Grundschulen. Ein großer Teil der beobachteten Varianz erklärt sich noch immer aus sozistrukturellen Merkmalen (insbesondere dem durchschnittlichen Buchbestand in den Elternhäusern, dem Schulabschluß der Eltern und dem Ausländeranteil), die das Umfeld einer Schule bestimmen. Dies gilt nicht in gleicher Weise für die Fähigkeit zum nicht sprachgebundenen schlußfolgernden Denken und schulbezogene Einstellungen von Schülerinnen und Schülern.

(2.) Vor diesem Hintergrund gibt es Schulen, die – gemessen an den Bedingungen ihres sozialen Umfeldes – erwartungswidrig gut, und solche, die erwartungswidrig schlecht abschneiden.

(3.) Bildet man Durchschnittswerte für alle Hamburger Schulen, so deuten sich möglicherweise auch für deren Leistungsniveau insgesamt Probleme im Verhältnis zu dem Niveau der Schulen in anderen Ländern (auch in anderen Großstädten) an.

Inzwischen werden im Rahmen der OECD unter dem Kürzel *PISA* („*Programme for International Student Assessment*“) weitere Vergleichsuntersuchungen vorbereitet, an denen sich ungefähr 30 Staaten, darunter auch die Bundesrepublik Deutschland, beteiligen werden. Projektnehmer auf internationaler Ebene ist das Australian Council of Educational Re-

search (ACER)<sup>47</sup>. Die Durchführung der Untersuchungen in Deutschland ist einem Konsortium unter Führung des Max-Planck-Instituts für Bildungsforschung in Berlin übertragen worden. Gegenstand der Untersuchung ist ein Vergleich der Leistungen von 15-jährigen Schülerinnen und Schülern in den Bereichen Leseverständnis, Mathematik, Naturwissenschaften, wobei neben international vereinbarten Untersuchungsansätzen<sup>48</sup> bestimmte Ergänzungen in den einzelnen beteiligten Ländern (nationale Optionen) vorgesehen sind. Im Rahmen der nationalen Optionen soll in Deutschland z.B. geklärt werden, welche aussichtsreichen pädagogischen Interventionsmöglichkeiten zur Verbesserung von Leseverständnis und Lesekompetenz sich identifizieren lassen.

Dabei geht es sowohl um die *pädagogisch-psychologische Basis von Fördermaßnahmen* (bereichsspezifisches Vorwissen, motivationale Faktoren wie Aufmerksamkeit und thematisches Interesse, Arbeitsgedächtnis und Verfügbarkeit von Lese- und Lernstrategien) als auch um die Situation von Risikogruppen („schwache Leser“ und Leser nicht-deutscher Muttersprache), um außerschulische Stützsysteme beim Erwerb der Lesekompetenz (Leseanlässe und Lesengelegenheiten, häusliche Lesekultur und Regeln der Mediennutzung, Wertschätzung des Lesens in den jeweiligen peer-groups), um die Rolle des Lesens im Sprach- und Literaturunterricht sowie seine Bedeutung und Nutzung als Instrument selbständigen Wissenserwerbs. Untersucht werden sollen ferner *fächerübergreifende Kompetenzen* (cross curricular competencies), so z.B. die Fähigkeit, extracurriculare Projekte zu planen und zu steuern, allein oder gemeinsam mit anderen technische Probleme zu lösen oder Lösungsstrategien durch Analogiebildung auf neue Aufgabenstellungen zu übertragen. Unter dem Aspekt „soziale Kompetenz“ soll die Fähigkeit untersucht werden, soziale Informationen zu dekodieren und zu interpretieren sowie ein Repertoire variabler Handlungsentwürfe für gängige Interaktionssituationen zu entwickeln. Die Untersuchungen sollen sowohl die sozialen Kontexte wie auch Aspekte der Schulqualität etc. einschließen.

Die Felderprobung der erforderlichen Tests wird bis zum Sommer 1999 durchgeführt werden. Die Hauptuntersuchung soll im Jahre 2000 erfolgen, der Bericht im Jahre 2001 vorliegen. Inzwischen hat die Kultusministerkonferenz beschlossen, die Untersuchungen über die für den internationalen Vergleich erforderliche begrenzte Stichprobe hinaus so anzulegen, daß auch *Leistungsvergleiche innerhalb der Bundesrepublik* möglich werden<sup>49</sup>. Ebenfalls beschlossen wurde von der Kultusministerkonferenz, eine weitere Vergleichsuntersuchung zur Fremdsprachenkompetenz der Schülerinnen

47 Beteiligt sind ferner das Netherlands Institute of Educational Management (Cito), der Service de Pedagogie Experimentale, Universite de Liege (SPE), und Westat Inc., Rockville (USA), als Datenverarbeitungszentrum.

48 Die „Frameworks“ der geplanten Untersuchungen sind nach dem jeweils aktuellen Stand im Internet verfügbar (<http://www.pisa.oecd.org>).

49 Beschuß der Amtschefskonferenz der KMK vom 17./18. September 1998.

und Schüler an deutschen Schulen durchzuführen. Diese Untersuchung ist inzwischen ausgeschrieben worden.<sup>50</sup>

### **Was weiß man und was tut man, wenn man untersucht hat?**

Vorbereitung und Durchsetzung von empirischen Untersuchungen zum Leistungsstand von Schulen erfordern einen hohen Aufwand. Dieser steht in einem eigenartigen Gegensatz zu dem vergleichsweise geringen Aufwand, den man sich bisher noch leistet, um eine tragfähige Konzeption für die *strategische Nutzung der Ergebnisse solcher Untersuchungen* zu entwickeln. Zwar hat die Rückmeldung der Untersuchungsergebnisse der Lernausgangslagenuntersuchung in Hamburg an die Grundschulen und an die weiterführenden Schulen einen intensiven Diskussionsprozeß in Gang gesetzt. Er bezieht sich auf die Klassenbildung, die pädagogische Arbeit innerhalb der Klassen und die Entwicklung von spezifischen Fördermaßnahmen, die Zusammenarbeit der Lehrkräfte innerhalb einer Schule ebenso wie auf die Zusammenarbeit von Grundschule und weiterführenden Schulen, Aktivitäten der Schulberatung, der Lehrerfortbildung und der Standardsicherung durch Vergleichsarbeiten (vgl. Behörde für Schule, Jugend und Berufsbildung Hamburg 1998). Insgesamt wäre es jedoch falsch zu glauben, daß auf diese Weise allein bereits alle notwendigen *Impulse zur Veränderung der Situation* ausgelöst werden können. Jedes der oben mitgeteilten Ergebnisse dieser Untersuchung muß Folgerungen und weitere Klärungen auf unterschiedlichen Ebenen nach sich ziehen, die vielfach nicht von einer einzelnen Schule allein bewirkt werden können.

Die Aussagen in These (2) legen den Schluß auf *Ursachenfaktoren im Lehrerhandeln und im Konzept der einzelnen Schule* nahe. Um herauszufinden, was man tun sollte, wäre es vor allem interessant zu klären, was die erwartungswidrig guten Schulen anders machen als andere und ob bzw. unter welchen Bedingungen ihre Vorgehensweisen auf andere Schulen übertragbar sind.

Die Aussagen zu These (3) legen Vermutungen nahe, die sich in gleicher Weise auf die Situation aller Hamburger Schulen beziehen. Wenn man auf dieser Ebene handeln will, muß man klären, was andere Länder in gleicher Situation anders machen als Hamburg (curriculare Vorgaben? Struktur, Einstellungen und Qualifikation der Lehrerschaft? Steuerungsverfahren? oder ...?).

Der Umgang mit den Folgerungen aus These (1) ist fraglos am schwierigsten. Sie bestätigt, was längst bekannt ist, aber zeitweilig im Glauben an die Wirksamkeit vielfältiger Förderprogramme und den mit ihnen verbundenen finanziellen Input zu sehr verdrängt wurde: Unterschiedliche Lernleistungen haben ihre Ursache auch in den sozialen Bedingungen, unter denen Schüler aufwachsen. Soweit sich dies daraus erklärt, daß Lernen eben nicht nur in der Schule stattfindet, sondern in starkem Maße auch von au-

50 Beschuß der Amtschefkonferenz der KMK vom 4./5. Februar 1999.

Berschulischen Stützsystemen abhängt, drängt sich die Frage auf, ob man dies resignierend auf sich beruhen lassen darf oder ob es nicht doch außerschulisch ansetzende wirksame Strategien gibt, soziale Unterschiede besser als bisher zu kompensieren. Entzaubert ist freilich der Glaube, daß man bereits durch eine Addition vielfältiger Förderprogramme ohne weiteres Probleme löst.

Bei alledem lassen sich viele der damit aufgeworfenen Fragen zuverlässig anhand von Querschnittsuntersuchungen noch nicht beantworten. Deshalb plant Hamburg *Folgeuntersuchungen für die Klassenstufen 7 und 9*. Sie werden gleichfalls flächendeckend angelegt sein. Die Tests für die Klassenstufe 7 sind abgeschlossen, aber noch nicht ausgewertet. Die Untersuchungen werden von der Frage geleitet, ob sich die Unterschiede in den Leistungen am Ende der Grundschule im weiteren Verlauf ausgleichen, ob sie unverändert bleiben oder ob sie sich vielleicht sogar verstärken. Erst auf dieser Basis ließen sich z.B. Entscheidungen über die weitere Entwicklung der Grundschulkonzeption treffen. Denn immerhin ist denkbar, daß auch bei schwächeren Leistungen zu einem bestimmten Zeitpunkt die Basis für positive Entwicklungen in der Folgezeit bereits gelegt ist.

Es zeigt sich mithin: Leistungsvergleiche erlauben *eine allgemeine Standortbestimmung* („bench-marking“) und geben erste Hinweise auf Handlungsnotwendigkeiten. Sie liefern jedoch nur begrenzt Aufschluß über mögliche Ursachenzusammenhänge und erst recht nicht bereits didaktische Konzepte zur Beseitigung erkannter Defizite. Dabei sind Aussagen stets nur in den Grenzen des gewählten Untersuchungsansatzes möglich. Dies darf indessen keine Rechtfertigung dafür sein, unbequeme Untersuchungsergebnisse mit dem Argument der begrenzten Reichweite von Untersuchungen vorschnell zu relativieren oder Untersuchungen nur dort anzusetzen, wo man sich für unangreifbar hält.

Vergleichsuntersuchungen sind nach alle dem *nur der Einstieg in ein Feld, dem weitere Schritte in Form gezielter Untersuchungen folgen müssen*. Je nachdem, auf welcher Ebene man ansetzen will und welche Fragestellungen dabei leitend sein sollen, wird auch die Frage unterschiedlich zu beantworten sein, wie breit (flächendeckend oder stichprobenartig, bestimmte Gruppen), in welcher Zeitabfolge (einmalig, regelmäßig, anlaßbezogen) und auf welchen Veranlassungsebenen (Einzelschule zum Zwecke der Selbstevaluation, Steuerungswissen der Schuladministration) weitere Untersuchungen anzulegen sind. Dabei ist es eine offene Frage, ob gelegentliche Untersuchungen zu unterschiedlichen Fragestellungen ausreichen oder ob es einer regelmäßigen Berichterstattung über die Leistungsstände von Schülerinnen und Schülern bedarf. Auch wenn hinter dieser Frage ein erhebliches Maß an Sprengstoff stecken könnte, wird sie in absehbarer Zeit zu beantworten sein.

Die Folgerungen aus bereits gewonnenen Erkenntnissen und aus weiteren Untersuchungen werden das Handeln vieler Akteure erfordern. Sie beziehen sich auf Änderungen in der Lehrerausbildung ebenso wie auf verstärkte

Anstrengungen der Lehrerfortbildung, auf curriculare Entwicklungen und die Entwicklung didaktischer Konzepte<sup>51</sup> wie auf Bemühungen im Rahmen der Schulentwicklung und des Schulmanagements.

### Risiken und Nebenwirkungen ausgeschlossen?

Untersuchungen, wie sie derzeit durchgeführt werden, sollen Steuerungswissen erzeugen. Sie dienen nicht einer Überprüfung der individuellen Leistung einzelner Schüler oder einzelner Lehrer. Zentrale Tests *sollen weder die geltenden Formen der Leistungsbewertung von Schülerinnen und Schülern (Noten, Zeugnisse) noch der Beurteilung des einzelnen Lehrers ersetzen*. Sie sind weder die Grundlage der individuellen Entscheidung über Berechtigungen zum Besuch weiterführender Bildungseinrichtungen noch ein Element der Personalbeurteilung und Karriereplanung. Insofern ist die Situation in Deutschland der Situation in England oder in den Vereinigten Staaten nicht vergleichbar, wo Tests bereits in großer Breite zur regelmäßigen Leistungsbewertung im Vergleich einzelner Schulen untereinander („Ranking“) wie auf der individuellen Ebene der Schülerinnen und Schüler Verwendung finden. Eine solche Entwicklung zuzulassen, kann auch kein vernünftiges schulpolitisches Ziel sein. Trotzdem muß man sich darüber klar sein, daß auch die in Deutschland eingeleiteten Entwicklungen nicht frei von Risiken und Nebenwirkungen sind:

(1) Es können sich *Rückwirkungen für Curricula und Unterrichtsmaterialien* ergeben, die Ziel und Mittel umdrehen. Statt zu testen, ob man erreicht, was man erreichen will, versucht man zu erreichen, was sich auf eine einfache Weise testen läßt.

(2) Eine zunehmende Bedeutung von Testergebnissen für die Wertschätzung der einzelnen Schule im Urteil der Eltern kann Schulen *zu unsinnigen Reaktionen veranlassen*, z.B. Tests in extensiver Weise zu trainieren oder Schüler, von denen ein negativer Einfluß auf das Testergebnis erwartet wird, vom Test fernzuhalten.

(3) Die Funktion der Rechenschaftslegung („accountability“) kann die der Verbesserung („improvement“) auf Dauer überlagern und statt zu einer Qualitätssteigerung *zum Festhalten am (scheinbar) Bewährten verleiten*. Diese Gefahr wäre insbesondere dann gegeben, wenn sich die Meßverfahren in der Fixierung auf einmal gewählte Parameter gegenüber der kritischen Reflexion komplexer Zusammenhänge unterschiedlicher Dimensionen von Leistung und Qualität verselbständigen.

(4) Standardisierung und Vereinheitlichung *können die Reduzierung von Vielfalt bewirken*, die gerade das Kennzeichen eines föderalen Systems

---

51 Vgl. in diesem Zusammenhang als Reaktion auf die Ergebnisse von TIMSS etwa die Erklärung der Fachverbände DMV, GDM und MNU vom 21. Mai 1998, die die Notwendigkeit einer konzeptionellen Veränderung des Mathematikunterrichts in Deutschland unterstreicht, sowie das Gutachten von Baumert et. al. für die Bundesländer-Kommission für Bildungsplanung und Forschungsförderung (BLK 1997).

sein soll. Auch wenn es unsinnig ist, unterschiedliche Bildungsziele gegen einander auszuspielen, so kommt man doch nicht daran vorbei, daß sich nicht alle Ziele (zumal angesichts der Restriktionen im Zeitbudget von Schülerinnen und Schülern) gleichzeitig optimal realisieren lassen. Man muß sich also entscheiden. Diese Entscheidung kann unterschiedlich ausfallen. Man muß sie in ihrer Unterschiedlichkeit gelten lassen, solange sie sich innerhalb einer Bandbreite hält, die Durchlässigkeit und Vergleichbarkeit nicht ernstlich gefährdet.

(5) Das Leistungsniveau der Schülerinnen und Schüler einer Schule hängt stark von deren sozialen *Kontextbedingungen* ab. Ein „ranking“ von Schulen, das diesen Aspekt vernachlässtigt, ist nicht zulässig. Dennoch kann niemand garantieren, daß es angesichts der zu erwartenden plakativen Darstellung von Untersuchungsergebnissen zu hinreichend differenzierten Urteilen kommt. Eine Schule in objektiv ungünstiger Situation wird vermutlich auf Dauer eine ungünstige Wettbewerbssituation haben. Demotivation von Lehrkräften und Schülern könnte die Folge sein. Es ist offen, ob sich dieser Gefahr bereits dadurch ausreichend begegnen ließe, daß Gegenstand von Untersuchungen nicht der Leistungsstand zu einem bestimmten Zeitpunkt, sondern der Leistungszuwachs innerhalb eines bestimmten Zeitraums (value added approach) wird.

Die Gefahr von Fehlentwicklungen in dem damit angedeuteten Sinne ist durchaus real. Sie darf dennoch kein Grund sein, den Kopf in den Sand zu stecken und darauf zu verzichten, Urteilsmaßstäbe auch aus den Vergleichen mit anderen abzuleiten. Aber man muß sich der Möglichkeit einer mißbräuchlichen oder unkundigen Verwendung von Untersuchungsergebnissen sehr bewußt sein. Auch wenn niemand daran denkt, derartigen Entwicklungen Vorschub zu leisten, kann man nicht sicher sein, daß alle den Versuchungen auf Dauer widerstehen können. Die Faszination einer scheinbaren Objektivität, die sich ergibt, wenn man in einem allzu simplen Urteil die vielfältigen Aspekte vernachlässtigt, die bei der Interpretation von Untersuchungsergebnissen bedacht sein wollen, ist nicht gering. Welche Dynamik den eingeleiteten Entwicklungen innenwohnt, läßt sich deshalb derzeit kaum abschätzen. *Wachsamkeit ist unerlässlich.*

## Literatur

- Baumert, J./Bos, W./Watermann, R.: TIMSS/III – Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe 11 im internationalen Vergleich. Zusammenfassung deskriptiver Ergebnisse, Berlin, 1998.
- Baumert, J./Lehmann, R./Lehrke, M./Schmitz, B./Clausen, M./Hosenfeld, I./Köller, O./Neubrand, J.: TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im Vergleich. Deskriptive Befunde, Opladen, 1997.
- Behörde für Schule, Jugend und Berufsbildung Hamburg: Standardisierte Leistungsmessung als Instrument der Qualitätssicherung im Bildungswesen. Materialien für die Referendarinnen und Referendare, Hamburg, 1998.

BLK: Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, Materialien zur Bildungsforschung, Heft 60, Bonn, 1997.

Lehmann, R./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.

### THOMAS BETHGE

### Zum Umgang mit den Ergebnissen von TIMSS

Durch die TIMS-Studie (Baumert u.a. 1997) wird der mathematische und naturwissenschaftliche Unterricht zum Gegenstand der öffentlichen Debatte. Seine Leistungsfähigkeit wird angezweifelt, das Selbstbild der Fachkolleginnen und -kollegen wird infrage gestellt. Es werden von der Öffentlichkeit Rezepte zur schnellen Veränderungen der Situation gefordert. Das deutsche Bildungssystem muß seinen Mittelpunkt verlassen und einen Spitzensplatz im Nationenranking einnehmen. Die öffentliche Debatte ist nicht bei der Kritik der mathematischen und naturwissenschaftlichen Fachleistungen stehengeblieben, sie bezieht sich auf das gesamte Bildungssystem. Die große Bedeutung, die die TIMS-Studie durch die Verallgemeinerung ihrer Ergebnisse über die Naturwissenschaften hinaus erlangt hat, macht es nicht einfacher, die notwendige Weiterentwicklung des mathematischen und naturwissenschaftlichen Unterrichts zu diskutieren.

Die *Kritik an Mathematik und Naturwissenschaften* in der Schule ist nicht neu. Fachdidaktiker, Lehrer und Fachverbände haben die Kritik vorgetragen, sie erreichte nicht die Öffentlichkeit und erforderte damit keine bildungspolitischen Reaktionen.

Folgende Kritikpunkte wurden vorgetragen:

- Unterricht erlangt nur eine begrenzte Wirksamkeit für den Aufbau naturwissenschaftlichen Wissens und Könnens bei den Schülerinnen und Schülern. Ein kontinuierlicher Kompetenzzuwachs über die Schuljahre ist für sie nicht erfahrbar.
- Das Interesse am naturwissenschaftlichen Unterricht wird im Verlauf der Schulzeit geringer (s. Hoffmann/Lehrke 1986).
- Im Mathematikunterricht bestehen Defizite im Umgang mit der mathematischen Symbolsprache und bei der mathematischen Modellbildung (s. Heymann 1996).