

PÄDAGOGIK



WERNER BLUM • MARIANNE DEMMER
ANNEMARIE VON DER GROEBEN • KLAUS-JÜRGEN TILLMANN

LEISTUNGSVERGLEICHE - UND KEIN ENDE?

PÄDAGOGIK KONTROVERS: BRAIN GYM IN DER SCHULE?



THEMA

LEISTUNGSVERGLEICHE - UND KEIN ENDE?

Internationale und nationale Leistungsvergleiche werden seit TIMSS sowohl grundlegend in Frage gestellt als auch vehement verteidigt. PÄDAGOGIK fragt: Wem nützen solche Studien? Fördern Vergleichsstudien die Qualität von Schule? Was sind Risiken und Nebenwirkungen? Welche Erfahrungen machen Schüler, Lehrer, Schulen? Was erwartet uns in Zukunft? Information und Diskussion ist notwendig: Die Sachlage ist differenzierter geworden, Fragen und Einwände aber sind geblieben.

BEITRAG

GYMNASIALE OBERSTUFE
Zwei Experten fragen nach Entwicklungsaufgaben der gymnasialen Oberstufe. Ludwig Huber konzentriert sich auf die Zielperspektive „Selbständiges Lernen“. Heinz-Elmar Tenorth diskutiert curriculare und didaktische Entwicklungsmöglichkeiten. Ergänzende und kontroverse Positionen zu aktuellen Fragen.

PÄDAGOGIK:KONTROVERS

BRAIN GYM

Über „Gehirngymnastik“ - auch bekannt unter den Begriffen Edukinesetik und Kinesologie - in der Schule gibt es immer wieder Streit. Die Kontroverse diskutiert philosophische Hintergründe, praktische Bedeutung und kritische Einwände.

PÄDAGOGIK

52. JAHRGANG, HEFT 12/2000

THEMA

LEISTUNGSVERGLEICHE - UND KEIN ENDE?

MODERATION: KLAUS-JÜRGEN TILLMANN

ANNEMARIE VON DER GROEBEN/KLAUS-JÜRGEN TILLMANN
PRO UND CONTRA LEISTUNGSVERGLEICHSTUDIEN
Eine Kontroverse 6

ISABELL VAN ACKEREN/KLAUS KIEMM
TIMSS, PISA, LAU, MARKUS UND SO WEITER
Ein aktueller Überblick über Typen und Varianten von Schulleistungstudien 10

CHRISTIAN KLUG/SABINE REH
WAS FANGEN DIE SCHULEN MIT DEN ERGEBNISSEN AN?
Die Hamburger Leistungsvergleichsstudie aus der Sicht „beforschter“ Schulen 16

WERNER BLUM
QUALITÄTSENTWICKLUNG IM MATHEMATIKUNTERRICHT ...
... eine Folge von TIMSS? 22

JOHAN C. VAN BRUGGEN
QUALITÄTSENTWICKLUNG DURCH ZENTRALE TESTS?
Das Beispiel Niederlande 28

MARIANNE DEMMER
RISIKEN UND NEBENWIRKUNGEN VON SCHULLEISTUNGSVERGLEICHEN
Argumente gegen Testeuphorie 32

JULIA ANNI KROHNE
„ICH MUSS JETZT DEUTSCHLAND VERTRETEN“
Schülererfahrungen mit Pisa 36

BEITRAG

LUDWIG HUBER
SELBSTÄNDIGES LERNEN AUF DER OBERSTUFE
Lernumgebung und Orientierung 39

ELMAR TENORTH
PROPÄDAGOGIK UND REFLEXIVITÄT
Funktion und Lernorganisation der Oberstufe 46

PÄDAGOGIK:KONTROVERS

BRAIN GYM IN DER SCHULE?
PRO: MARIA ROBERG
CONTRA: MANFRED SYLVANUS 50

REZENSIONEN

ANNEGRET EICKHORST
THEORIE DER SCHULE
Bestandsaufnahmen und neue Perspektiven 53

MAGAZIN

U.A.: 17. KONGRESS DER DGfE IN GÖTTINGEN 57

PISA

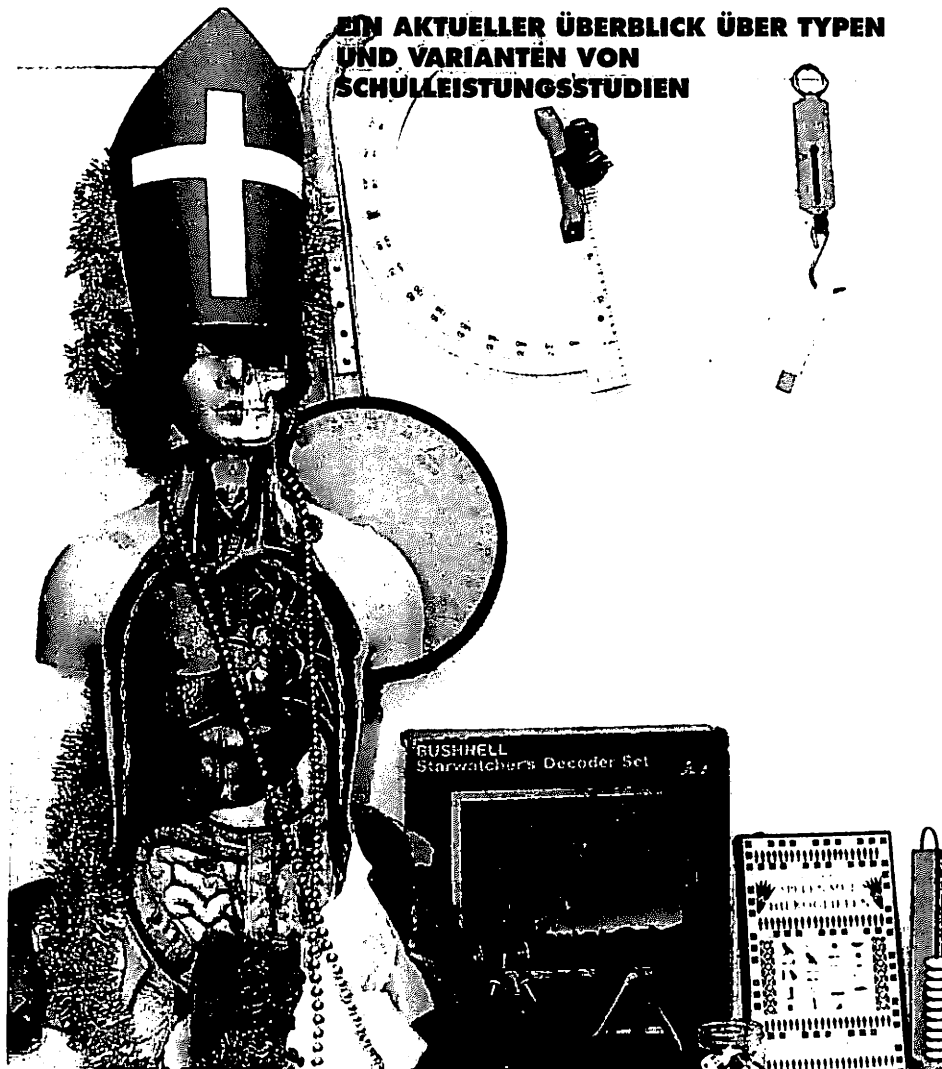
REINHARD KAHL'S KOLUMNE
Gemischte Verhältnisse 64

Selt TIMSS folgt eine Leistungsvergleichsstudie der nächsten: Die deutschen Beteiligungen an internationalen Studien gehören ebenso dazu wie die Deutschland-internen Bundesländervergleiche, außerdem führen einzelne Bundesländer zusätzlich noch eigene Studien durch. Dieser Beitrag verschafft einen Überblick und ordnet die verschiedenen Studien kritisch ein.

ISABELL VAN ACKEREN, KLAUS KLEMM

TIMSS, PISA, LAU, MARKUS UND SO WEITER

Ein aktueller Überblick über Typen und Varianten von Schulleistungstudien



In diesem Beitrag wird eine Übersicht über Typen und Varianten von Schulleistungstudien gegeben. Zunächst sollen aber diejenigen Tests und Prüfungen benannt werden, die nicht in diesen Überblick einbezogen werden. Dazu gehören zunächst die deutlich mehr als acht Millionen Klassenarbeiten, die jedes Jahr in den etwa 410.000 deutschen Schulklassen (bzw. Kursen) geschrieben werden. Es wird weiterhin nicht auf die Vergleichsarbeiten eingegangen, die im wachsenden Maße in den Parallelklassen einer Schule oder auch der Schulen einer Region zur Sicherung vergleichbarer Standards geschrieben werden. Schließlich werden auch nicht die zentralen Abschlussprüfungen behandelt, die in sieben Bundesländern vor allem im Rahmen der Abiturprüfungen, z. T. aber auch für das Erreichen des mittleren Bildungsabschlusses sowie des Hauptschulabschlusses vorgesehen sind (Klemm 1998).

Die Übersicht befasst sich vielmehr mit dem Typ von Leistungsvergleichen, der in der deutschen Schulentwicklung der neunziger Jahre Prominenz erlangt hat: Damit sind all die empirischen Leistungsvergleichs-Untersuchungen gemeint, die mit wissenschaftlichem Anspruch in bestimmten Jahrgängen (bzw. Altersgruppen) des Schulwesens eingesetzt werden, um dort den Stand der fachlichen – z. T. auch der überfachlichen – Kenntnisse zu ermitteln. Solche Studien beziehen sich immer auf eine bestimmte Region (z. B. ein Bundesland, die Bundesrepublik), dabei werden entweder alle Schüler der entsprechende Gruppe in dieser Region (z. B. alle Fünftklässler in Hamburg) einbezogen, oder aber es wird eine repräsentative Stichprobe dieser Gruppe ausgewählt. Solche Tests werden (zumindest bisher) nicht regelmäßig durchgeführt, sie sind vielmehr als gesonderte Forschungs- und Evaluationsprojekte angelegt. Die dabei erbrachten Leistungen fließen bei den einzelnen Schülern nicht in deren Benotungen ein, sie bleiben vielmehr in aller Regel anonym.

Bei der deskriptiv angelegten Darstellung dieser Vergleichsstudien, die auch als „large scale assessments“ bezeichnet werden, unterscheiden wir im folgenden drei Typen (vgl. Abb. 1):

- Studien, die international angelegt sind (mit deutscher Beteiligung).
- Studien, die bundesländerübergreifend national angelegt sind.
- Studien, die ausschließlich innerhalb eines einzelnen Bundeslandes durchgeführt werden.

An dieser Systematik orientiert sich die folgende Darstellung.

INTERNATIONALE SCHULLEISTUNGSSTUDIEN MIT DEUTSCHER BETEILIGUNG

Der Überblick über die international angelegten Studien beginnt – nicht der Chronologie folgend – mit den zwischen 1996 und 1998 vorgelegten TIMS-Studien, welche die Reihe der seit 1959 von der „International Association for the Evaluation of Educational Achievement“ (IEA) durchgeführten Leistungsstudien fortsetzen. Diese TIMS-Studien, die bei Schülern und Schülerinnen der 7. und 8. Klassen sowie der Abschlussklassen der Sekundarstufe II durchgeführt wurden (an der entsprechenden Studie für den Primarbereich hatte sich Deutschland nicht beteiligt), testeten in den Bereichen Mathematik und Naturwissenschaften. Neben der reinen Lernstandserfassung bei über einer halben Million Schülerinnen und Schülern mehrerer Altersgruppen enthielt sie zusätzliche Untersuchungskomponenten zu schulischen wie außerschulischen Kontextbedingungen; dadurch wurde TIMSS zur bis dahin umfangreichsten Schulleistungstudie. Für Deutschland waren ihre Ergebnisse insgesamt – für alle Bildungsgänge, für die unterschiedlichen Kompetenzbereiche und für alle Bundesländer – im internationalen Vergleich derart mittelmäßig, dass sie eine heftige und noch anhaltende Debatte zur Leistungsfähigkeit des deutschen Schulsystems einleiteten (Baumert/Lehmann u. a. 1997; Baumert/Bos/Watermann 1998). Eine Folge dieser Diskussion war es, dass die in Deutschland bis dahin verbreitete Distanz gegenüber derartigen großflächigen Vergleichsstudien innerhalb weniger Jahre aufgegeben, ja nahezu in ihr Gegenteil verkehrt wurde. Auch wenn es vorher schon in den neunziger Jahren mit der „International Study of Reading Literacy“ eine international angelegte Vergleichsuntersuchung mit deutscher Beteiligung gegeben hat (Elley 1992), rechtfertigt die Debatte um TIMSS und ihre innerdeutsche Wirkung die Feststellung, dass mit TIMSS eine Zäsur im deutschen Umgang mit „large scale assessments“ erreicht wurde.

Diese internationale Lesestudie von 1991, deren deutscher Teil 1995 veröffentlicht wurde (Lehmann u. a. 1995), ermittelte die Leistungen am Ende des Erstleseunterrichts und am Ende der allgemeinen Pflichtschulzeit. Sie erbrachte für Deutschland im internationalen Vergleich ebenfalls nur ein mittelmäßiges Abschneiden. Gleichwohl wurde sie aber selbst in der Fachöffentlichkeit, zunächst kaum wahrgenommen. Erst das in Folge der TIMS-Studien erwachte Interesse am Leistungsstand deutscher Schülerinnen und Schüler sicherte dieser Studie

nachträglich eine höhere Aufmerksamkeit. Die Verzögerung der Wahrnehmung hängt sicher auch damit zusammen, dass die öffentliche Interesse an den Vergleichsstudien und an ihren Ergebnissen von Anfang an in die deutsche Standortdebatte eingebettet war. Im Kontext dieser ökonomisch ausgerichteten Diskussion erscheinen der Öffentlichkeit mathematisch-naturwissenschaftliche Kompetenzen offensichtlich wichtiger als z. B. Bereiche wie Leseverständnis und Lesegewohnheiten.

Zehn Jahre nach dieser ersten IEA-Lesestudie startet im nächsten Jahr PIRLS, die „Progress in International Reading Literacy Study“ zum Leseverständnis beim Wechsel von der Primar- in die Sekundarstufe, die durch Schüler- und Lehrerfragebögen zur schulischen und häuslichen Situation ergänzt wird, und deren Ergebnisse für das Jahr 2003 erwartet werden. Unter Leseverständnis wird dabei die Fähigkeit verstanden, geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren. Für die bereits 1991 beteiligten Länder, darunter Deutschland, besteht nunmehr mit Hilfe einer Trendstudie die Möglichkeit, Entwicklungstendenzen in diesen zehn Jahren aufzudecken, was in Zukunft regelmäßig durch Wiederholungen der Untersuchung alle vier Jahre möglich sein soll.

Größere Aufmerksamkeit erzielt derzeit allerdings eine international vergleichende Erhebung des Indikatorenprogramms INES (Indicators of Educational Systems) der OECD: PISA, das für einen Gesamterhebungszeitraum von zehn Jahren angesetzt, seit 1998 geplante „Programme for International Student Assessment“ unter Beteiligung von über 30 OECD-Mitgliedstaaten steht im Mittelpunkt der Diskussionen. Getestet werden 15-Jährige zu drei Erhebungszeitpunkten mit unterschiedlicher Schwerpunktsetzung in den Bereichen Leseverständnis, Mathematik und Naturwissenschaften sowie Cross-Curricular Competencies (Lernstrategien, motivationale Aspekte, Kooperationsfähigkeit ...), ergänzt durch sozio-ökonomische Hintergrundmerkmale und Variablen zu schulischen Lernbedingungen. Die Haupterhebung des ersten Projektzyklus mit dem Schwerpunkt Lesen, deren Ergebnisse für Ende 2001 erwartet werden, fand bereits Mitte 2000 statt (vgl. den Beitrag von Krolne in diesem Heft). Im Jahr 2003 wird der Schwerpunkt dann auf Mathematik und im Jahr 2006 auf Naturwissenschaften verlagert. Im Unterschied zu TIMSS ist bei PISA geplant, die Ergebnisse in ihrer Kombination von Daten zu schulischen Kontext-, Prozess- und Wirkungsvariablen so an Schulen zurückzumelden, dass sie dort für die Unterrichts-

entwicklung genutzt werden können. Den Überblick über die internationalen Studien abschließend, sei noch auf die zweite *Civic Education Study* der IEA (nach 1971) verwiesen, die 1999 durchgeführt wurde und die einen weiteren Kompetenzbereich erfasst: *CivEd* versucht, Auskünfte über Art, Umfang und mögliche Ursachen des politischen Wissens von 14-jährigen Jugendlichen – an der Zusatzserhebung unter 16- bis 18-Jährigen nimmt Deutschland nicht teil – mittels qualitativer und quantitativer Messmethoden zu geben. Dafür wurden in einer ersten Phase zwischen 1995 und 1997 Experteninterviews und curriculare Analysen durchgeführt, in einer zweiten Phase wurden politisches Wissen, Einstellungen, Engagement sowie außerschulische, z. B. familiäre und mediale Einflüsse erfasst.

BUNDESLÄNDER-ÜBERGREIFENDE LEISTUNGSVERGLEICHE

Die bereits erwähnte internationale PISA-Studie wurde durch zusätzliche Erhebungsinhalte und quantitative Erweiterungen der Stichprobe im Rahmen der zeitgleich durchgeführten nationalen Ergänzungsstudie PISA-E ausgebaut. So sind in Deutschland z. B. Schüler(innen) mit einer vergleichsweise deutlich unterdurchschnittlichen Lesekompetenz ebenso wie Schüler(innen) fremder Muttersprache mit erwartungswidrig guten Leistungen überrepräsentiert, um Schwächen besser nachvollziehen und Förderperspektiven ableiten zu können. Weiterhin werden sogenannte schulunabhängige Stützsysteme wie die häusliche Lesekultur, die Mediennutzung im Elternhaus und bei den Peers differenzierter als auf der internationalen Ebene untersucht. Die Er-

gänzungen sollen zudem sicherstellen, dass die getesteten Inhalte den deutschen Schulen behandelten Stoffgebieten entsprechen. In quantitativer Hinsicht wurde die altersbasierte durch eine jahrgangsbasierte Stichprobe ergänzt, um die Leistungsfähigkeit im Abschlussjahrgang der Sekundarstufe I abschätzen zu können. Schließlich wird PISA auf KMK-Beschluss auch für den innerdeutschen Vergleich genutzt, wofür die internationale Stichprobe von etwa 8.000 Jugendlichen um knapp 50.000 erweitert wurde. Daneben hat die KMK im September beschlossen, „Vergleichsuntersuchungen zum Leistungsstand von Schülerinnen und Schülern im Englischen und in der aktiven Beherrschung der deutschen Sprache“ (Akronym DESI für Deutsch-Englisch-Schülerleistungen-International) an ein Konsortium unter Federführung des Deutschen Instituts für Internationale

Pädagogische Forschung (DIPF) zu vergeben, das gemäß Ausschreibungstext allerdings auf Möglichkeiten der Zusammenarbeit mit dem PISA-Konsortium hin angelegt sein soll. Eine Position zwischen landes- und bundesweit angelegten Studien nimmt das BIJU-Forschungsprojekt unter Federführung des Max-Planck-Instituts für Bildungsforschung in Berlin ein. Dieses als Längsschnittstudie angelegte Projekt erfasst *Bildungsverläufe* und psychosoziale Entwicklungen im *Jugendalter* in den Ländern Nordrhein-Westfalen, Mecklenburg-Vorpommern, Sachsen-Anhalt und Berlin. Im Mittelpunkt stehen – neben dem über lehrplanvalide Tests ermittelten Lernerfolg in fünf Fächern – vor allem psychosoziale Entwicklungsaspekte in Abhängigkeit von familiären und institutionellen Rahmenbedingungen. Dies alles wird im Kontext des schulinstitutionellen

Wandels in den neuen Bundesländern gesehen und mit den alten Bundesländern verglichen. Die 1991 in Klasse 7 gestartete Untersuchung wird nach mehreren bereits erfolgten Erhebungen eine weitere Erhebungswelle mit der Befragung der mittlerweile im Studium, der Berufsausbildung oder im Beruf befindlichen Teilnehmer im Winter 2000 haben, wobei weitere Erhebungen im Rahmen des BIJU-Projektes nach Auskunft des Instituts denkbar sind; ein Ende ist noch nicht absehbar. Mit der speziellen Sicht auf den Vergleich von Fachleistungen sei noch erwähnt, dass im Kontext von BIJU die 1999 veröffentlichte Teilstudie zum Vergleich der Mathematikleistungen von Oberstufenschülern an integrierten Gesamtschulen und Gymnasien in Nordrhein-Westfalen entstanden ist (*Köller/Baumert/Schnabel 1999*).

LANDESWEITE SCHULLEISTUNGSTESTS

In der Gruppe bundeslandinterner Schulleistungstests sind derzeit zahlreiche Aktivitäten zu verzeichnen. Die folgende Übersicht stellt daher eine Momentaufnahme dar, die beim Erscheinen dieses Beitrags möglicherweise schon aktualisierungsbedürftig sein wird. Die größte Aufmerksamkeit hat fraglos die seit 1996 in Hamburg laufende Untersuchung „Aspekte der *Lernausgangslage*“ (*LAU*) erzielt. In dieser Studie wurden in allen fünften Klassen der allgemeinbildenden staatlichen sowie der interessierten Sonder- und Privatschulen Hamburgs die Bereiche Sprach- und Leseverständnis, aktives und passives Rechtschreibwissen und Mathematik untersucht. Eine Fortsetzung fand diese Erhebung 1998 mit der in Klasse 7 angesiedelten Folgestudie „Aspekte der Lernausgangslage und der Lernerwicklung“, die zusätzlich die erste Fremdsprache einbezog und bei der stärker als bei der ersten Untersuchung schülerbezogene Merkmale sowie schul- und unterrichtsbezogene Einstellungen erhoben wurden. Die Untersuchung wird für die Jahrgangsstufen 9 und 11 fortgesetzt. Die bisherigen Auswertungen der Hamburger Studie geben Aufschluss über Unterschiede bezüglich der Testleistungen und Einstellungen bei besonderen Schülergruppen, wie etwa bei Jungen und Mädchen, Schulfremdwechslern, Kindern bildungsferner Elternhäuser sowie bei Kindern aus Ein-Eltern- oder Migrantenfamilien (*Lehmann u. a. 1997, 1999*). Eine weitere Besonderheit der Untersuchung ist darin zu sehen, dass Ergebnisse dieser breit angelegten Längsschnittstudie klassenbezogen rückgemeldet wurden, so dass die Studie auf diese Weise sehr konkret für die Unterrichtsentwicklung der einzelnen Klassen genutzt werden kann (vgl. den Beitrag von *Klug/Reh* in diesem Heft). Das Land Rheinland-Pfalz hat im Mai 2000 in allen 8. Klassen sämtlicher Schulformen – die Sonderschule ausgenommen – die Mathematikleistungen, die Einschätzung von Unterrichtsmerkmalen sowie den schulischen und außerschulischen Kontext der Leistungserbringung mit Hilfe von Tests und von Befragungen der Schüler,

Bezeichnung der Studie ausführliche Bezeichnung	Abkürzung	Koordination/wiss. Begleitung		Planung	Zeitraum Erhebung	Ergebnisse	Testbereiche	Altersgruppe	Schulform	Anzahl Schüler		Erhebungsart
		internat.	national							internat.	national	
Civic Education Study	CivEd	IEA	Humboldt-Universität zu Berlin	1993	1995-1997 1999	Frühjahr 2001	politisches Wissen, Einstellungen, Engagement	14-Jährige	alle ohne Sonderschulen	120 000	5 000	•
International Study of Reading Literacy		IEA	Humboldt-Universität zu Berlin	1989	1991	1995	Leseverständnis und Lesegewohnheiten	3. und 8. Klassen	alle ohne Sonderschulen	210 000	10 000	•
Programme for International Student Assessment	PISA	OECD, ACER	MPIB, PISA-Kommission + PISA-Berater	1998	2000 2003 2006	Ende 2001 (erste Phase)	Leseverständnis, Mathematik, Naturwissenschaften, Lehrer-übergreifende Kompetenzen	15-Jährige	alle	180 000	8 000 + 50 000 (PISA-E)	•
Progress in International Reading Literacy Study	PIRLS	IEA	ISFA, Universität Hamburg	2000	2001	2003	Leseverständnis	9 bis 10-Jährige	Grundschule	210 000	6 000	•
Third International Mathematics and Science Study	TIMSS	IEA	IPN, MPIB, DIPF	1989	1994/1995	1996-1998	Mathematik, Naturwissenschaften	7. und 8. Klasse, Ende 5 II	alle ohne Sonderschulen	500 000	12 000	•
Bildungsverläufe und psychosoziale Entwicklung im Jugendalter	BIJU		MPIB	1989	1991, 1993, 1995, 1997, 2000	laufend	Deutsch, Englisch, Mathematik, Biologie, Physik, psychosoziale Aspekte	1991 Beginn mit 7. Klassen (Längsschnittstudie)	alle	letzte Erhebung 1997 ca. 7000	•	
Programme for International Student Assessment - Erweiterung	PISA-E		MPIB, PISA-Kommission + PISA-Berater	1998	2000, 2003, 2006	Ende 2001 (erste Phase)	vgl. PISA	Ende 5 I	alle	50 000	•	
Deutsch-Englisch-Schülerleistungen-International	DESI		DIPF	1999	2001	2005	Englisch, Deutsch	15-Jährige bzw. 9. Klassen	vgl. PISA	vgl. PISA	•	
Aspekte der Lernausgangslage (und der Lernerwicklung)	LAU		Amt für Schule/Humboldt-Universität zu Berlin	1995	1996, 1998, 2000	1997, 1999	Deutsch, Mathematik, 1. Fremdsprache (Englisch/Laten)	5., 7., 9. Klassen	alle ohne Sonderschulen	jeweils ca. 12 000	•	
Bayerischer Mathematiktest	BMT		Staatinstitut für Schulpädagogik und Bildungsforschung	1998	seit 1998 regelmäßig		Mathematik (ab 2000/01 auch Deutschtest)	Jahrgangsstufen 7 und 9	Realschulen, Gymnasien u. demnachst Hauptschulen	Gymn. 1999 32 000	•	
Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext	MARKUS		Bildungsministerium, Universität Koblenz-Landau	1998	2000	Ende 2000	Mathematik	8. Klassen	alle ohne Sonderschulen	44 000	•	
Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik	QuaSUM		Bildungsministerium, Humboldt-Universität	1997	1999	2000	Mathematik	5. und 9. Klassen	alle ohne Sonderschulen	13 000	•	
Vergleichsuntersuchung im Saarland			Schulbehörde	1998	1998/1999		Mathematik, Deutsch	Klassen 7 und 9	Gymnasien, Real- und Gesamtschulen	ca. 9 000	•	

- Anmerkungen:
 1. International Association for the Evaluation of Educational Achievement
 2. Australian Council of Educational Research
 3. Max-Planck-Institut für Bildungsforschung, Berlin
 4. Institut für Schulentwicklungs-forschung, Dortmund
 5. Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel
 6. Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt a. M.
 Die Zahlen sind gerundet

Abb. 1: Isabell von Ackeren, Klaus Klemm: Organisation, Anlage und Umfang von Schulleistungstests in Deutschland bzw. mit deutscher Beteiligung



Lehrer und Schulleiter untersucht. Für das Projekt 'Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext', kurz MARKUS, wurden unterschiedliche Testtypen entwickelt, die teilweise curriculumentorientiert sind, um eine gemeinsame Fähigkeitsskala für den landesinternen Vergleich zu haben, und die teilweise auch aus Aufgaben, wie sie etwa bei TIMSS zur Anwendung gekommen sind, bestehen. Dadurch soll die Vernetzbarkeit unterschiedlicher Studien gesichert werden. Im Vordergrund steht bei MARKUS eindeutig das Herausarbeiten von Ansatzpunkten eines breiten Qualitätsmanagements, das sich auf die Schulentwicklung, die Lehreraus-, Fort- und Weiterbildung richtet (Ministerium 2000). Angesichts dieser Zielsetzung wird bei MARKUS der Rückmeldung der Untersuchungsergebnisse an die Schulen und der Entwicklung von Hilfsangeboten zur Unterrichts-, Organisations- und Personalentwicklung eine besonders hohe Aufmerksamkeit gewidmet.

Von den beiden flächendeckend angeleg-

ten Untersuchungen in Hamburg und in Rheinland-Pfalz unterscheiden sich Studien, die stichprobenartig angelegt sind, wie etwa die brandenburgische 'Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik' (QuaSUM), bei der im Juni 1999 die Lernstände der 5. und 9. Klassen im Fach Mathematik durch von den Lehrerinnen und Lehrern des Landes entwickelte Aufgaben repräsentativ erhoben wurden. Mit dem Ziel, Anhaltspunkte für die Weiterentwicklung der pädagogischen Arbeit in der Primar- und der Sekundarstufe I zu erhalten, wurde mittels Schüler-, Lehrer-, Schulleiter- und auch Elternfragebögen eine Aufklärung der für Varianzen verantwortlichen schulischen und außerschulischen Bedingungen versucht.

Weitere, flächendeckend angelegte Schulleistungstests in der Bundesrepublik werden im Unterschied zu den beiden genannten Studien regelmäßig durchgeführt, beziehen dabei aber Kontextvariablen, welche Beiträge zur Erklärung der Ergebnisse liefern können, nicht oder kaum mit ein. In Bayern verfolgt man beispielsweise

seit 1998 im Rahmen der 'Bildungs-offensive Mathematik' (Kelly 1999) die Strategie, Schülern, Eltern und Lehrern eine Orientierungshilfe in der Art an die Hand zu geben, wie man sie z.B. in Frankreich oder England auf unterschiedlichen Altersstufen, den sogenannten 'cycles' bzw. 'key stages', bereits seit langer Zeit kennt. Dazu werden bayernweite Mathematiktests eingesetzt, die für die Realschulen in Klasse 7 der sechsstufigen und in Klasse 9 der vierstufigen Realschule verpflichtend sind und die von den Gymnasien bisher freiwillig, aber mit enormem Zuspruch durchgeführt werden. In den nächsten Jahren sollen diese Tests bei einer gleichzeitigen Ausweitung auf andere Fächer auch für die Hauptschulen eingeführt werden; eine Teilnahme wird dann auch für die Gymnasien verpflichtend sein. Zudem wird überlegt, die Testergebnisse in Zukunft zu bewerten, um dadurch zu erreichen, dass derartige Tests von den Schülern ernster genommen werden.

Eine ähnliche Entwicklung zeichnet sich im Saarland ab, wo im Zuge der 'Qua-

litäts-offensive Saar' 1998 erstmals alle Siebtklässler der Gymnasien, 1999 die Jahrgangsstufe 7 der Real- und Gesamtschulen und ebenfalls 1999 die 9. Klassen der Gymnasien in Mathematik landesweit getestet wurden (Cornely 2000). Zum gleichen Zeitpunkt wurden die Leistungen der Jahrgangsstufe 7 der Gymnasien im Fach Deutsch extern erhoben. Derartige Vergleichsuntersuchungen mit Ergebnismeldungen an die einzelnen Schulen und Klassen sollen in Zukunft fortgesetzt und auf die vierten Klassen der Grundschulen ausgeweitet werden.

ENTWICKLUNGSTENDENZEN

Die Übersicht über Schulleistungstests dokumentiert die Vielfalt der Bemühungen auf Bundes- und Landesebene, im Bereich der externen Evaluation verstärkt tätig zu werden. Der Überblick zeigt auch, dass es im Verlauf der neunziger Jahre eine Entwicklung bei der Qualität der eingesetzten Instrumente und der dahinterstehenden Konzepte gibt. Diese Entwicklung ist dadurch gekennzeichnet,

- dass sich das Interesse der Leistungstudien auf erklärende schulische und außerschulische Variablen der Leistungserbringung ausweitet.
- dass Rückmeldestrategien entwickelt werden, die sich nicht nur an die bildungspolitisch interessierte Öffentlichkeit, sondern verstärkt an die Einzelschule und an die einzelne Lerngruppe sowie an ihre Lehrer und Lehrerinnen richten und
- dass die Leistungsstudien in umfassendere Konzepte der Qualitätsentwicklung eingebaut werden.

Zugleich verweist die hier vorgelegte Übersicht über Typen und Varianten von Schulleistungstests aber auch auf eine andere Tendenz: Nach Jahren der Abstinenz im Bereich großflächig angelegter Leistungstests beobachten wir seit Beginn der neunziger Jahre in Deutschland eine Beschleunigung beim Einstieg in regelmäßige Vergleichsuntersuchungen. Dafür gibt es fraglos einige gute Gründe: Die Aussagekraft von Schulleistungsuntersuchungen würden – die Hamburger LAU-Studie führt dies vor – durch Längsschnittstudien mit mindestens zwei Messungen qualitativ überproportional gesteigert. Auch ist es nachvollziehbar, dass auf die gut begründete Kritik an der Konzentration der eingesetzten Tests auf die Kulturtechniken mit einer Ausweitung auf weitere Fächer, z. B. auf geistes- und sozialwissenschaftliche Kompetenzbereiche, reagiert wird. Schließlich kann es kaum zurückgewiesen werden, dass die Schulpolitik in einem föderalen Staat aus Gründen der Sicherung der Vergleichbarkeit ein starkes Interesse an regelmäßigen innerdeutschen Ver-

gleichsuntersuchungen hat. Würde man jedoch diesen Weg weitergehen und Schulleistungsvergleiche zu einer regelmäßigen Veranstaltung ausbauen, so würde man sie mit all ihrem Entwicklungspotential von der Schulentwicklung abkoppeln: Schulen, Schulaufsicht und Forschung wären überfordert, regelmäßige Vergleichsstudien aufzunehmen und produktiv zu verarbeiten. Dass aus der Fülle von Tests nicht zwangsläufig Qualität erwächst, zeigt uns das Beispiel der USA: Dort können statt steigender Schulleistungen eher Abnutzungserscheinungen bei den Testanwendern beobachtet werden (Kempfert/Rolf 1999, S. 17). Wer 'large scale assessments' unter der Hand dazu nutzen will, eine Deutschland umfassende Form regelmäßiger zentraler Leistungsüberprüfung zu etablieren, der sollte diese Nebenwirkungen kalkulieren und zur Kenntnis nehmen, dass – so hat uns gerade die TIMS-Studie gezeigt – zentrale Tests vielleicht der Vergleichbarkeit von Zertifikaten dienen, gewiss aber keinen eigenständigen Beitrag zur Sicherung von international hoher Qualität darstellen.

LITERATUR

- Baumert, J./Lehmann, R. u. a.: TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Berlin 1997. (gekürzte Version im Internet unter: <http://www.mpib-berlin.mpg.de/TIMSS/II/Ergebnisse.htm>)
- Baumert, J./Bos, W./Watermann, R.: TIMSS/III. Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich. Berlin 1998
- Cornely, J.: Guter Unterricht schafft Qualität. Maßnahmen zur Qualitätsentwicklung im Saarland. In: Schulverwaltung Rheinland-Pfalz, Nr. 1/2000, S. 17–18
- Elley, W. B.: How in the world do students read? IEA study of reading literacy. IEA The Hague/Netherlands 1992
- Kelly, M.: Die Bildungsoffensive Mathematik. In: Schulverwaltung Bayern, Nr. 5/99, S. 177–182
- Kempfert, G./Rolf, H.-G.: Pädagogische Qualitätsentwicklung. Weinheim 1999
- Klemm, K.: Steuerung der Schulentwicklung durch zentrale Leistungskontrollen? In: Rolf, H.-G. u. a. (Hg.): Jahrbuch der Schulentwicklung 10. Weinheim 1998, S. 271–294
- Köller, O./Baumert, J./Schnabel, K.: Wege zur Hochschulreife: Offenheit und Vergleichbarkeit des Systems und Sicherung vergleichbarer Standards. Analysen am Beispiel der Mathematikleistungen von Oberstufenschülern an Integrierten Gesamtschulen und Gymnasien in Nordrhein-Westfalen. In: Zeitschrift für Erzie-

hungswissenschaft Nr. 3/1999, S. 385–422

Lehmann, R./Peek, R./Pieper, I./von Stritzky, R.: Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen. Weinheim u. a. 1995

Lehmann, R./Peek, R.: Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Hamburg 1997 (Broschüre der Hamburger Schulbehörde, siehe auch Internet)

Lehmann, R./Peek, R.: Aspekte der Lernausgangslage an der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen. Hamburg 1999 (Broschüre der Hamburger Schulbehörde, siehe auch Internet)

Ministerium für Bildung, Wissenschaft und Weiterbildung (Hg.): Informationsbroschüre MARKUS. Mainz 2000

Weitere Informationen im Internet (Stand der Recherche: August 2000)

BIJU: <http://www.mpib-berlin.mpg.de/EuB/program/areas/project11-1.htm> und <http://www.mpib-berlin.mpg.de/EuB/BIJU-Bericht2/Titel.htm>

BMT: http://www.isb.bayern.de/gym/math_inf/rahmen99.htm

CivEd: <http://www.iea-dpc.de/> und <http://www.indiana.edu/%7E7Essdc/ieadig.htm>

LAU, beide Endberichte: <http://lbs.hh.schule.de/laulau5/> und <http://lbs.hh.schule.de/laulau7/>

PIRLS: <http://www.iea-dpc.de/>

PISA: <http://www.mpib-berlin.mpg.de/PISA/>

QuaSUM: http://www2.hu-berlin.de/empir_bf/QuaSUM.htm

TIMSS: <http://timss.bc.edu/>

Dr. Klaus Klemm, Jg. 1942, ist Professor für Erziehungswissenschaft im Fachbereich 2 der Universität GH Essen.

Adresse:
Universität
GH Essen,
45117 Essen,
e-mail: k.klemm@uni-essen.de



Isabell van Akeren, Jg. 1974, ist wissenschaftliche Mitarbeiterin im Fachbereich 2 der Universität GH Essen.

Adresse:
Universität
GH Essen,
45117 Essen,
e-mail: isabell.vanacker@uni-essen.de

