

1. Jahrgang

Heft 2
2001

Newsletter



Ö F E B

- Beitrag: Jürgen Baumert, Evaluationsmaßnahmen im
Bildungsbereich. Eine kritische Sicht auf mögliche Zugänge ... S. 3
- Interna ... S. 15
- Bericht ... S. 16
- Ankündigung ... S. 18
- Tagung der ÖFEB im September in Salzburg ... S. 20

Ö F E B

Österreichische Gesellschaft
für Forschung und Entwicklung
im Bildungswesen

Impressum:

Der Newsletter erscheint halbjährlich.

Herausgeberin und Medieninhaberin:

Österreichische Gesellschaft für Forschung und Entwicklung im Bildungswesen

Redaktion:

Geraldine Kraus, Georg Neuweg, Franz Rauch, Andrea Seel, Josef Thonhauser

Für den Inhalt verantwortlich:

Josef Thonhauser, Universität Salzburg, 5020 Salzburg, Akademiestr. 26.

Layout:

Kirstin Eckstein, Universität Salzburg, 5020 Salzburg, Akademiestr. 26

Karin-Gratiana Wurm, Pädagogische Akademie der Erzdiözese Wien, 1210 Wien, Mayerweckstr. 1 (Titel)

Jürgen Baumert, MPI für Bildungsforschung, Berlin:

Evaluationsmaßnahmen im Bildungsbereich.

Eine kritische Sicht auf mögliche Zugänge¹

Qualitätsentwicklung ist seit vielen Jahren ein wichtiges Thema im Bildungsbereich, das Thema Qualitätssicherung durch vergleichende Leistungsmessung dagegen relativ neu. Noch vor wenigen Jahren konnten in Deutschland internationale Vergleichsuntersuchungen nahezu ohne jede öffentliche Aufmerksamkeit durchgeführt werden, obwohl deren Ergebnismuster sich von den Befunden von TIMSS kaum unterschieden (Lehmann, Peek, Pieper & v. Stritzky 1995). Bei TIMSS und PISA scheint alles anders zu sein.

1. Typisierung von Evaluationsmaßnahmen

Eine Reihe von Streitpunkten, die in der Diskussion über Qualitätssicherung an Schulen immer wieder auftauchen, sind darauf zurückzuführen, dass weder ausreichend zwischen unterschiedlichen Evaluationsmaßnahmen und deren spezifischen Funktionen noch zwischen unterschiedlichen Instrumenten und deren spezifischer Eignung unterschieden wird. Wenn alles, was über den kommunikativen Austausch im Kollegium hinausgeht unter dem Oberbegriff standardisierter Testuntersuchungen zusammengefasst wird, ist es schwierig, sich über Evaluationsmaßnahmen und mögliche ungeplante Nebenfolgen zu verständigen. Als erster Schritt zu einer Systematisierung soll im folgenden eine grobe Unterscheidung hinsichtlich der Lokalisation der Verantwortung für eine Evaluationsmaßnahme und der Verfügbarkeit von Evaluationsergebnissen getroffen werden. In die eine Kategorie fallen Maßnahmen, die von der Einzelschule initiiert werden und deren Ergebnisse innerhalb der Schule verbleiben; zur zweiten Kategorie gehören zentral durchgeführte Untersuchungen, an denen eine größere Anzahl von Schulen beteiligt ist und die in der Regel in der einen oder anderen Form die Schulaufsicht involvieren. Innerhalb beider Kategorien sind wiederum unterschiedliche funktionale Ebenen zu unterscheiden. Dezentrale Evaluationsmaßnahmen dienen entweder der individuellen Bewertung von Schülerinnen und Schülern, der schulinternen Selbstvergewisserung zur Optimierung von Schulentwicklung oder der externen Bewertung von Programmen, wobei die Überprüfung entweder von der Schule initiiert und mit einem außenstehenden Partner ausgehandelt werden kann oder im Rahmen der schulaufsichtlichen Tätigkeit als Prozess- und Ergebnisevaluation erfolgt. In beiden Varianten ist die Verbesserung der Praxis eigentliches Anliegen.

Es gibt keine öffentliche Einrichtung, in der regelmäßiger und häufiger evaluiert wird, als in der Schule. Die Bewertung von Schülerleistungen ist so selbstverständlich, dass sie bei der Diskussion um Evaluation in der Regel vergessen wird. Die Bewertung der Schülerleistung hat in der Regel alle Merkmale, die bei Systemevaluationen gerade im höchsten Maße umstritten sind. Lehrkräfte bilden Rangreihen von Schülern innerhalb des Referenzrahmens einer Lernergruppe (Ranking anhand von Noten oder Punkten), die Note gibt keine inhaltlichen

¹ Seit einigen Jahren sind internationale Vergleichsuntersuchungen, allen voran TIMSS und PISA, wie man so sagt, in aller Munde (Einen umfassenden Überblick bietet F. E. WEINERT (Hg.): Leistungsmessungen in Schulen. Weinheim und Basel: Beltz, 2001).

Jene Institution, die sich – jedenfalls im deutschsprachigen Raum – sowohl analytisch als auch praktisch am intensivsten damit auseinandergesetzt hat, ist das Max-Planck-Institut für Bildungsforschung in Berlin. Professor Jürgen BAUMERT, einer seiner Direktoren und ein profunder Kenner der Materie, hat uns dankenswerter Weise einen hochaktuellen Beitrag überlassen, der hier in gekürzter Form wiedergegeben wird. Die vollständige Fassung wird noch in diesem Jahr im 43. Beiheft der Zeitschrift für Pädagogik erscheinen. Wir danken seinem Herausgeber, Professor Jürgen OELKERS, Universität Zürich, für die freundliche Genehmigung des teilweisen Vorabdrucks.

J. T.

Auskünfte über die verfügbaren Kompetenzen, und das mit einer Note verbundene Fähigkeitsniveau kann von Lerngruppe zu Lerngruppe erheblich schwanken. Individuelle Lernfortschritte spielen - wenn überhaupt - bei der Notenvergabe nur eine ergänzende Rolle. Lernentwicklungsberichte sind insofern ein korrigierender Ansatz, als sie die soziale Bezugsnorm in den Hintergrund drängen möchten und die Rückmeldung individueller Profile betonen. Die Nutzung individueller Beurteilungsnormen gerät allerdings auch wiederum schnell an ihre Grenzen, wenn Konflikte mit meritokratischen Gesichtspunkten distributiver Gerechtigkeit sichtbar werden.

Zum Kernbereich der dezentralen Erfassung von Ergebnis- und Prozessmerkmalen gehören alle Maßnahmen der schulinternen Evaluation, die der Optimierung der Schulentwicklung dienen. Häufig wird die schulinterne Evaluation in Gegenüberstellung zur externen Evaluation von Einzelschulen, aber auch in Gegenüberstellung zum Systemmonitoring, das später vorgestellt werden soll, als wünschenswerte und -romantisierend - auch als problemarme Form der Ergebnissicherung verstanden. In dem Augenblick aber, in dem Unterricht in die schulinterne Evaluation einbezogen wird, ist sie ein konfliktreiches Unterfangen, das einen Sprengsatz für die traditionelle Organisationsstruktur der Schule darstellt, insofern das Autonomie-Paritätsparadigma der Lehrerschaft ausgehebelt wird. Hier wird ein Prüfstein für die professionelle Entwicklung des Lehrerberufs liegen. Gleichzeitig wird für die Lehrerfortbildung eine dauerhafte Herausforderung durch die Aufgabe, Unterrichtsentwicklung kooperativ zu unterstützen, entstehen. Denn schulinterne Evaluation setzt Schulentwicklung voraus (Gray u.a. 1996).

Dies gilt gleichermaßen für die externe Evaluation von Schulen und Schulprogrammen, insbesondere dann, wenn ausgehandelte Vergleichskriterien Referenzpunkte darstellen sollen. Einen kritischen Punkt stellt die externe Evaluation von Einzelschulen nach *vorgegebenen* Vergleichskriterien dar. Dabei ist insbesondere die Rolle der Schulaufsicht und die Reichweite ihrer Eingriffsrechte strittig. Allerdings deutet sich in der einschlägigen Literatur ein Verständnis darüber ab, dass die Schulaufsicht sich zunächst auf eine Überprüfung der Einhaltung der Gütestandards der Schulprogrammarbeit und der internen Evaluation beschränken sollte (Prozessevaluation bzw. Metaevaluation). Erst beim Vorliegen von Mängelrügen sollte die Schulinspektion folgen (Posch & Altrichter 1997; Strittmatter 1999; Rolff 1995; Grogger & Specht 1999). Erfolgt die externe Evaluation nach vorgegebenen Kriterien durch Dritte - etwa Wissenschaftler - wird es immer vorherige Absprachen zwischen den beteiligten Schulen und den Partner geben müssen, bei denen sich jede Schule in spezifischer Weise zu den Vergleichskriterien positioniert. Zwei Beispiele aus jüngerer Zeit liegen in publizierter Form vor (Klieme, Baumert & Schwippert 2000; Köller & Trautwein, in Vorbereitung). In diesen Fällen wollten Schulen ihre Arbeit zu TIMSS-Normen in Beziehung setzen, um eine generelle Niveauorientierung zu erhalten - allerdings sehr wohl in dem Bewusstsein, dass die in TIMSS erfassten Leistungsaspekte keineswegs den Kern ihrer pädagogischen Bemühungen treffen.

Insgesamt scheint die Verpflichtung auf interne und externe Evaluation bei einer verstärkten Selbständigkeit von Schulen unstrittig zu sein. Verfahrensstandards sind in Deutschland dagegen noch weitgehend unklar, auch wenn es in der Literatur zur Evaluation gute Vorlagen gibt (Joint Committee 1994). Hoch umstritten sind dagegen zentrale Ergebnis- und Prozessevaluationen. Unter anderem wohl auch deshalb, weil nicht zwischen funktionalen Ebenen zentraler Evaluationsmaßnahmen unterschieden wird. Umso dringender ist die Klärung (siehe Abbildung 1, in der die funktionalen Ebenen zentraler Evaluationsmaßnahmen unterschieden werden).

Bei der individuellen Zertifizierung durch zentrale Tests oder Prüfungen, mit denen Zugangsberechtigungen erteilt werden, ist zwischen Abschlussprüfungen, die von der abgebenden Institution veranstaltet werden, und Zugangsprüfungen, für die Abnehmer verantwortlich sind, zu unterscheiden. In fast allen europäischen Ländern sind Abschlussprüfungen am Ende der Schulzeit, die den Übergang auf weiterführende Bildungseinrichtungen regeln,

gängiger Standard (Frankreich, England, Norwegen, Schweden, Dänemark, Niederlande). Die durchgängig dezentrale Organisation zum Beispiel der Matura-Prüfungen in der Schweiz oder in Österreich ist eher eine Ausnahme. Diese Abschlussprüfungen haben *intentionale* Rückwirkungen auf die abgegebene Institution: Auf individueller Ebene sollen sich Schüler ausreichend auf die Prüfungen vorbereiten und auf institutioneller Ebene haben die Schulen auf entsprechende Standards zu achten. Gleichzeitig erhalten die Schulverwaltungen Rückmeldungen über die Qualität der Arbeit von Einzelschulen. *Teaching to the test* oder *Testcoaching* finden selbstverständlich statt, auch wenn sie nicht mit diesen Namen belegt werden, sondern als Lehrplantreue oder wünschenswerte Prüfungsvorbereitung gelten.

Abbildung 1: Funktionale Ebenen zentraler Evaluationsmaßnahmen

- | | |
|-----|--|
| 1. | Individuelle Zertifizierung durch zentrale Tests oder Prüfungen mit berechtigender Wirkung |
| 1.1 | Abschlussprüfungen |
| 1.2 | Zugangsprüfungen |
| 2. | Flächendeckende Evaluation von Einzelschulen bzw. ausgewählten Jahrgängen an Einzelschulen |
| 1.1 | Wettbewerbsmodell |
| | a) ohne Berücksichtigung von Ausgangsbedingungen |
| | b) mit Berücksichtigung von Ausgangsbedingungen (value added approach) |
| 1.2 | Modell professioneller Qualitätsentwicklung und Qualitätssicherung |
| | a) Separierung von Beratung und Kontrolle/Steuerung |
| | b) Verbindung von Beratung und Kontrolle/Steuerung |
| 3. | Systemmonitoring auf Stichprobenbasis |
| 4. | Internationale Vergleichsstudien auf Stichprobenbasis |

In Deutschland sind die zentralen Abiturprüfungen der Länder Bayern, Baden-Württemberg, Sachsen, Mecklenburg-Vorpommern und des Saarlands Beispiele diesen Evaluationstyps. Die Wirkung der zentralen Prüfungen geht in einigen Fächern deutlich über die Länder hinaus, in denen sie abgenommen werden. Die veröffentlichten Prüfungsaufgaben wirken standardisierend auch auf die Abiturprüfungen in Ländern, die eine dezentrale Prüfungsorganisation haben.

Unterschiedlich ist die Sachlage bei Zugangsprüfungen, die von Abnehmern veranstaltet werden. Je nach Schulnähe der Tests gehen unterschiedliche Rückwirkungen auf die abgebende Institution aus. Musterbeispiele für diesen Testtyp sind der *Scholastic Aptitude Test* (SAT) oder die Berufseingangsprüfungen der Industrie- und Handelskammern in Deutschland. Der SAT ist ein zentraler Hochschulzugangstest der USA, der aber weitgehend curriculumunspezifisch ist (das gilt auch für den Konkurrenten ACT oder den Englischtest für Ausländer, den TOEFL). Bei diesen standardisierten zentralen Tests, die regelmäßig eingesetzt werden, ist ein Testcoaching nicht zu verhindern. Die einzige Möglichkeit, damit rational und auch einigermaßen fair umzugehen, besteht darin, ein *Testcoaching* für möglichst alle Testteilnehmer anzubieten. Dies ist für die Validität des Tests relativ unproblematisch, da bekannt ist, dass ein coaching relativ schnell an die Obergrenze seiner Wirksamkeit kommt (Powers & Rock 1999). In dieser Weise ist auch bei dem Test zur Zulassung für die medizinischen Studiengänge in Deutschland verfahren worden (Trost u.a. 1998). Rückwirkungen auf die Schulpraxis durch Tests dieser Art sind relativ gering, wenn sie überhaupt auftreten. Umgekehrt

sind diese Tests aber auch nicht geeignet, Schulprogramme zu evaluieren. Die Kritik Popham's (1999) an dem Verfahren, Schulprogramme mit standardisierten Tests zu evaluieren, bezieht sich exakt auf Tests diesen Typs. Ganz anders liegen die Dinge möglicher Rückwirkungen der Berufseingangstests der Kammern. Diese Tests sind curriculumnah, beschränken sich aber auf basale, überwiegend sogar technische Fertigkeiten, deren Validität für die Berufsausübung ebenso strittig ist wie für den Unterricht der abgebenden Schulen. Wahrscheinlich wird nicht zu Unrecht eine Einschränkung des curricularen Spektrums im Abschlussjahrgang der betroffenen Schulform durch *teaching to the test* befürchtet (Blum, im Druck).

Von der individuellen Zertifizierung ist die flächendeckende Evaluation von Einzelschulen bzw. ausgewählter Jahrgänge an Einzelschulen abzusetzen. Zwei Varianten sind hier zu unterscheiden: Die erste Variante ist das Wettbewerbsmodell englischer oder schottischer Prägung. In einem Fall werden die Rohwerte der Evaluationsergebnisse von Schulen in den sogenannten *League Tables* veröffentlicht (ein Überblick über die Kritik gibt Sammons 1999). In der schottischen Spielart werden beim Bericht der Ergebnisse unterschiedliche Eingangsbedingungen in Rechnung gestellt (*Value Added Approach*: MacPherson 1992; Woodhouse & Goldstein 1988). Das Rationale beider Verfahren ist die Annahme einer nachfragegesteuerten Qualitätsentwicklung. Dieses Wettbewerbsmodell hat in den deutschsprachigen europäischen Ländern meines Erachtens keine Anhänger.

Davon zu unterscheiden ist die flächendeckende Evaluation von Einzelschulen, die in ein Modell professioneller Qualitätsentwicklung und Qualitätssicherung eingebettet ist. Auch hier sind zwei Spielarten zu erkennen. In einem Fall werden Beratung und Kontrolle konsequent getrennt, insofern die Evaluationsergebnisse ausschließlich den betroffenen Schulen ohne Einschaltung der Schulaufsicht zur Verfügung stehen. Im zweiten Fall gibt es keine klare Unterscheidung von Beratung und Kontrolle: Die Evaluationsdaten gehen an die Einzelschule und an die zuständige Behörde. Für beide Verfahren gibt es in Deutschland erste Beispiele: Brandenburg setzt auf die Trennung von Beratung und Kontrolle, Hamburg und Rheinland-Pfalz folgen dem Mischmodell. Welche Auswirkungen diese Evaluationsmaßnahmen in den Einzelschulen haben, inwieweit sie ein Instrument professioneller Entwicklung darstellen können und welcher systematischen Unterstützung ein solcher Prozess bedarf, ist weitgehend unklar. Ebenso wenig ist ausgelotet, wieweit die Verwendungsmöglichkeiten dieser Evaluationsergebnisse für Zwecke der Systemsteuerung reichen. Sicher ist, dass die flächendeckende Evaluation von Einzelschulen, wenn sie über die Erfassung einer Baseline hinausgeht und regelmäßig wiederholt wird, nur Sinn macht, wenn Schulentwicklungsmaßnahmen intensiver Art vorgeschaltet sind. Die Vorstellung, Veränderungen über die Rückmeldung von Evaluationsergebnissen per se einleiten zu können, ist wenig begründet (Fitz-Gibbon 1996). Nutzt man die flächendeckende Evaluation von Einzelschulen als Steuerungsinstrument, sind an Tests hohe Ansprüche bezüglich ihrer Verträglichkeit mit politisch und fachlich gewünschten didaktischen Konzeptionen zu stellen. Denn auch hier wird sich *Testcoaching* und möglicherweise auch *Teaching to the test* einstellen.

In einer Reihe von Bundesstaaten der USA sind obligatorische Testprogramme eingerichtet worden, welche die Funktion der individuellen Zertifizierung und der Evaluation von Einzelschulen, Klassen und indirekt auch von Lehrern verbinden (z.B. STAR in Kalifornien oder ISAT in Illinois). Sie sind nicht nur als Abschlusstests konzipiert, sondern werden in der Regel auch jahrgangswise administriert. Vergleichbar sind in gewisser Weise die in Frankreich schulbezogen veröffentlichten Ergebnisse des Abiturs. Vielfach werden die Ergebnisse in sogenannten *School Report Cards* veröffentlicht. Die Tests sollen schulübergreifende Standards etablieren. Aufgrund der dezentralen Schulorganisation sind sie jedoch weitgehend curriculumunspezifisch. Testtraining findet in diesen Programmen selbstverständlich statt. Ob die Tests die Lehrkräfte in ihrer Unterrichtsgestaltung einschränken und sich das Spektrum der Unterrichtsgegenstände auf die abgetesteten Stoffe reduziert, ist schwer zu beurteilen (Shepard 1990). Vorliegende Berichte weisen eher darauf hin, dass die Lehrkräfte das Testtraining

von ihrem eigentlichen Unterricht abkoppeln und vor der Testadministration zwei- bis dreiwöchige Übungsphasen einlegen, mit der Folge, dass die verfügbare Unterrichtszeit faktisch verkürzt wird.

Von der zentralen Evaluation von Einzelschulen sind nationale und internationale Studien zum Systemmonitoring auf Stichprobenbasis abzusetzen. Diese beiden Evaluationsformen haben keine direkt steuernden Funktionen für die Einzelschule, sondern primär die Aufgabe, Steuerungswissen auf Systemebene zur Verfügung zu stellen. Inwieweit von solchen Untersuchungen normative Rückwirkungen auf Unterricht und Schule ausgehen, ist ungeklärt. Direkte Rückwirkungen auf Einzelschulen kann man jedoch praktisch ausschließen, ebenso wie ein *Testcoaching*. Wenn Rückwirkungen zu verzeichnen sind, müssen diese über Vermittlungsinstitutionen mit politischer Legitimation transportiert werden. Die institutionalisierte Lehrplanarbeit oder Lehrerfortbildung könnten solche Transmissionsriemen sein. In diesem Fall sind die Rückwirkungen nicht nur legitim, sondern in der Regel auch in ein didaktisches Konzept eingebettet. Abbildung 2 ordnet noch einmal Beispiele von Untersuchungen mit unterschiedlichem Evaluationsanspruch.

Abbildung 2: Unterstützung der Qualitätsentwicklung und Qualitätssicherung durch empirische Bildungsforschung

Arbeitsform		Ziele	Beispiele		
Kontrollierte Interventionsstudien	Wissenschaft	Erklärung von Bedingungen der Wirksamkeit pädagogischer Maßnahmen	DFG-Schwerpunkt "Bildungsqualität"		
Längsschnittstudien		Analyse von Entwicklungsverläufen im Zusammenwirken von institutionellen und psychosozialen Faktoren	Scholastik (Weinert) BIJU (Baumert)	TIMSS (Baumert/Lehmann) u.a.	
Systemmonitoring		Beschreibung von (a) Rahmenbedingungen (b) Schule/Unterricht (c) Kompetenzen/Einstellungen der Schüler sowie (d) Zusammenhängen	Markus (Helmke) LAU (Lehmann) Quasum (Lehmann)		PISA (Baumert u.a.)
Evaluation von Schulen und Programmen		Zielbezogene, handlungsorientierte Feststellung und Bewertung von Stärken und Schwächen	Netzwerk selbstwirksamer Schulen (Edelstein u.a.)		
Begleitung von Reforminitiativen		Praxis	Feststellung der Ausgangsbedingungen, Praxisberatung, Prozessevaluation	BLK-Modellversuchsprogramm SINUS (Prenzel u.a.)	

2. Probleme der Selektivität von Evaluationsmaßnahmen

Die traditionelle Lehrplanarbeit gewinnt ihre Stabilität durch die enge Rückkoppelung an vorgängige Veränderungen in der Praxis und durch eine sachlich und zeitlich gegliederte Abarbeitung von Problemen. Auch bei der Entwicklung von Schulprogrammen wird man davon

ausgehen können, dass sie vielfach der Explikation und Fortschreibung vorgängiger Praxis dienen werden. Insofern wird auch hier in gewisser Weise das Lizenzierungsprinzip der Lehrplanarbeit greifen. Zum Schwur wird es jedoch bei der schulinternen Evaluation kommen, wenn ernsthafte Zielklärung betrieben werden muss. Hier ist mit schwierigen Auswahl- und Prioritätsentscheidungen zu rechnen, die zu Kontroversen führen und beträchtlichen Begründungsaufwand erfordern können. Dies gilt verstärkt für zentrale Evaluationsmaßnahmen, wenn sich diese flächendeckend auf Einzelschulen oder auf das System insgesamt beziehen. Dann unterliegen Auswahlentscheidungen weitaus größeren Begründungspflichten, da mit jeder Selektionsentscheidung gleichzeitig ein bildungstheoretischer Gesamtzusammenhang thematisiert wird.

In diesem Zusammenhang steht auch die Mehrzahl der Vorbehalte gegen standardisierte Leistungsmessungen. Diese Vorbehalte lassen sich im wesentlichen auf zwei Basiseinwände zurückführen. Der erste Einwand behauptet einen Widerspruch zwischen Ganzheitlichkeit von Bildungsprozessen und den eingeschränkten Fragestellungen von Evaluationsmaßnahmen, insbesondere wenn sie mittels standardisierter Leistungserhebungen erfolgten. Der zweite Einwand besagt, dass standardisierte Untersuchungen oftmals den ihnen zugrunde liegenden Bildungsbegriff oder das vorausgesetzte Fachverständnis nicht explizierten und Bildungsqualität letztlich mit dem in eins setzten, was ein Test erfasse.

Der vermeintliche Widerspruch zwischen Ganzheitlichkeit von Bildungsprozessen und dem Reduktionismus empirischer Forschung wird mit folgenden Argumenten begründet:

- In standardisierten Schuluntersuchungen würden Aufgaben der Schule selektiv berücksichtigt und die schulischen Ziele implizit auf messbare Bereiche eingeeengt. Musterbeispiel für dieses Argument ist die Gegenüberstellung fachlicher Schulleistungen und fachübergreifender Qualifikationen.
- Innerhalb des Fächerspektrums würden wiederum Fächer primär kognitiver Rationalität bevorzugt, so dass selbst innerhalb des obligatorischen Fächerspektrums der Schule bestimmte Formen des Weltverstehens unbegründet privilegiert würden. Unberücksichtigt blieben fast immer der historisch-sozialwissenschaftliche und insbesondere der ästhetisch-expressive Bereich.
- Innerhalb der Fächer würde die Substanz der Fächer auf Faktenwissen, das allein durch Tests erfassbar sei, eingeschränkt. Theoretisches und begriffliches Verständnis, methodisches Können und die selbständige Auseinandersetzung mit einem Sachverhalt sowie fächerübergreifende Perspektiven würden nicht zu ihrem Recht kommen.

Diese Kritik geht davon aus, dass Leistungstests nicht in der Lage seien, anspruchsvolle Aspekte fachlichen Verständnisses zu erfassen, geschweige denn fachübergreifende Leistungen, und dass es sich bei dem Verhältnis von fachlichen Leistungen und fachübergreifenden Qualifikationen um ein didaktisches Optimierungsproblem handele, bei dem widersprüchliche Ziele auszubalancieren seien. In diesem Zielkonflikt seien ganz unterschiedliche Akzentsetzungen denkbar und auch mit guten Gründen zu rechtfertigen.

Richtig ist zunächst, dass eine einzelne empirische Evaluationsmaßnahme nicht das gesamte Spektrum schulischer Ziele abbilden kann. Es ist notwendigerweise jeweils eine Auswahl zu treffen. Damit wird auch immer Aufmerksamkeit in eine bestimmte Richtung gelenkt. Daraus ist aber zunächst nur der Schluss zu ziehen, dass einzelne Leistungstests jeweils spezifische Dimensionen von Schule thematisieren und kein Gesamturteil, weder über eine Schule noch über ein Schulsystem, erlauben.

Zum Beispiel haben die an PISA teilnehmenden Staaten sich mit gutem Grund dafür entschieden, die Prüfung der Lesekompetenz und die Erfassung der mathematischen und naturwissenschaftlichen Grundbildung in den Mittelpunkt zu stellen und dieses Programm schritt-

weise um die Erfassung fächerübergreifender Kompetenzen zu ergänzen (OECD 1999). Leseverständnis ist in modernen Gesellschaften eine zentrale Schlüsselqualifikation. Sie ist nicht nur Voraussetzung dafür, in allen schulischen Fächern den Anschluss zu halten, sondern Grundlage für die aktive Teilnahme am gesellschaftlichen Leben und notwendige Voraussetzung für jede Form selbständigen Weiterlernens. Sie ist Voraussetzung und Teil sprachlich-literarischer Bildung, aber selbstverständlich nicht mit dieser identisch. Ein hinreichendes mathematisch-naturwissenschaftliches Verständnis, das deutlich über einfache Rechenfertigkeiten und die Anwendung von Formeln hinausgeht, ist der Schlüssel zu Kernbereichen moderner Kulturen. Man kann sich mit guten Gründen eine Erweiterung der Erhebungsgegenstände wünschen - Fremdsprachenkenntnisse wären wahrscheinlich ein wichtiger Kandidat für eine Erweiterung. Die für PISA getroffene Auswahl ist jedoch nicht beliebig. Allerdings ist PISA auch keine Studie, die generelle Aussagen über das erreichte Allgemeinbildungsniveau von Schülerinnen und Schülern erlaubt. Gleichwohl ist kaum zu bestreiten, dass mit der Auswahl von Untersuchungsbereichen immer auch eine Entscheidung über die Bedeutung von Fächern oder Domänen verbunden ist. In der Regel sind diese Präferenzen - was häufig übersehen wird - bereits aus den Stundentafeln und Vorschriften über die Leistungsfeststellungen in der Schule zu ersehen, die für die Institutionalisierung eines Bedeutungsgefälles von Bildungsgegenständen sorgen.

Berechtigt ist auch, dass Testkritiker darauf insistieren, dass die zu erfassenden Konstrukte theoretisch expliziert werden. Dieses Argument verliert auch nicht an Gültigkeit durch den Hinweis, dass die Leistungsmessung in der Schule diesen Anspruch gerade nicht erfüllt. Bei internationalen Vergleichsstudien hat man sich mit unterschiedlicher Intensität und in der Regel mit begrenztem Erfolg um die Entwicklung theoretischer Rahmenkonzeptionen bemüht. Im Rahmen der Testkonstruktion von TIMSS ist auch deutlich geworden, wie konfliktrichtig ein solcher Versuch beim Aufeinandertreffen unterschiedlicher didaktischer Ansätze sein kann. Über die Qualität der erreichten Kompromisse wird man streiten können. Dies gilt besonders dann, wenn man hinzufügt, dass sich die *frameworks* bislang empirisch nur begrenzt bewährt haben.

Dieser Mangel lässt sich in gewissem Umfang post hoc durch Maßnahmen der Konstruktvalidierung kompensieren. Im Rahmen von TIMSS haben einige Länder, darunter auch die Schweiz und Deutschland, entsprechende Schritte unternommen (Ramseier, Keller & Moser 1999; Klieme 2000; Klieme, Baumert, Köller & Bos 2000; Watermann & Baumert 2000). Die wichtigste Maßnahme ist die Definition von Fähigkeitsniveaus und deren Operationalisierung durch ausgewählte Testitems, die bei einem gegebenen Niveau mit hinreichender Sicherheit bearbeitet werden können. Durch dieses sogenannte *Proficiency Scaling* kann man auch leicht landläufigen Einwänden entgegentreten, standardisierte Tests erfassen nur Faktenwissen. Im Rahmen von PISA haben die Expertengruppen neue Anläufe unternommen, die theoretischen Konstrukte Lesekompetenz, mathematische und naturwissenschaftliche Literalität a priori theoretisch zu konzipieren. Mittlerweile liegen relativ explizierte Testkonzeptionen vor (OECD 1999; Baumert, Artelt, Klieme & Stanat, 2001).

Die Abbildungen 3 und 4 zeigen zwei Mathematikaufgaben aus dem Mittelstufen- und Oberstufentest von TIMSS, die stellvertretend für jene Gruppe von Aufgaben stehen, die mathematisches Problemlösen erfassen. Beide Aufgaben sind Indikatoren für das jeweils oberste Fähigkeitsniveau.

Abbildung 3:

Mathematisches Problemlösen (8. Klasse)

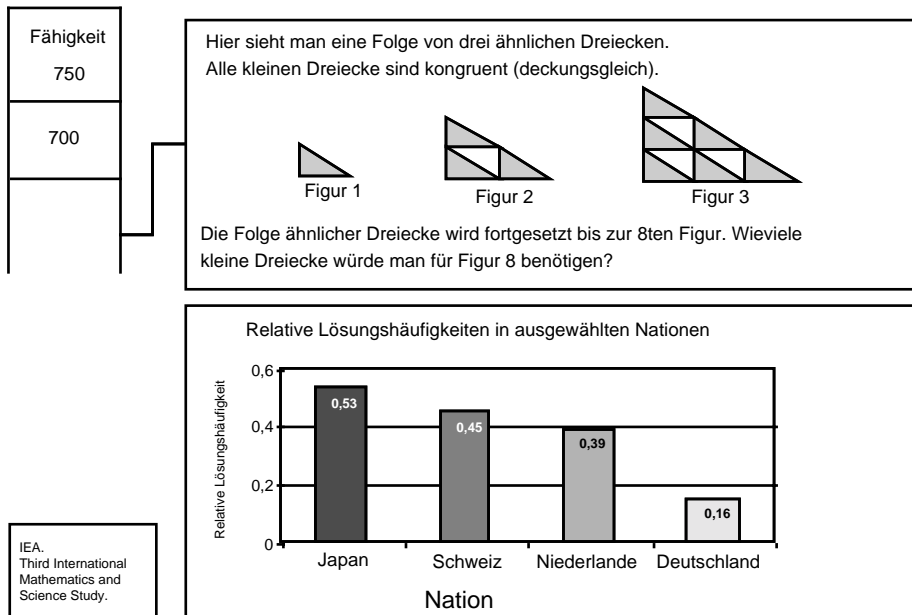
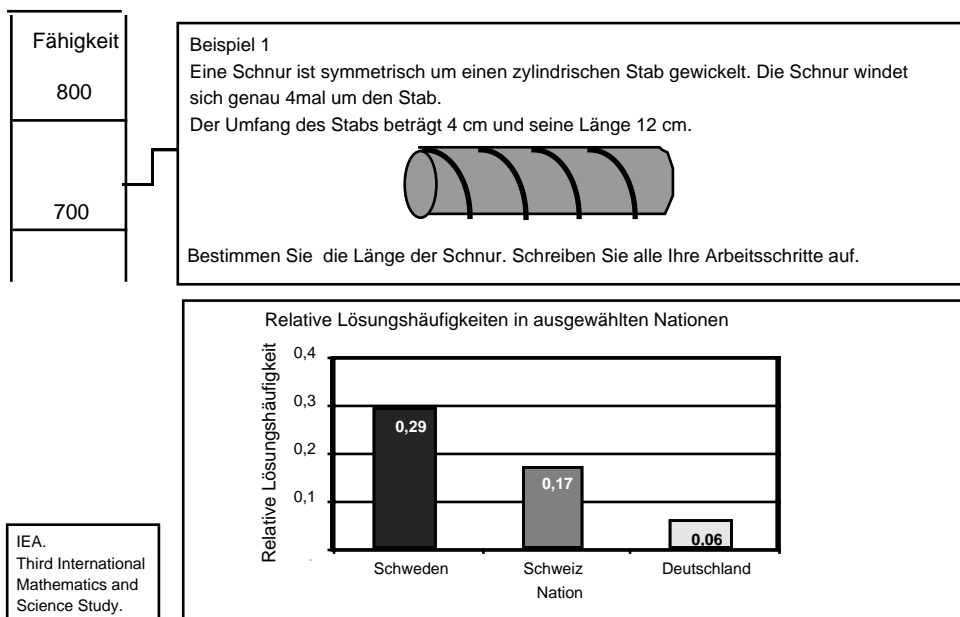


Abbildung 4:

Mathematisches Problemlösen (Gymnasiale Oberstufe)



Die Aufgaben sind entsprechend schwer. Ihre Schwierigkeit ist jedoch nicht darauf zurückzuführen, dass exotische Stoffe oder komplizierte Algorithmen abgeprüft werden, sondern darauf, dass Standardstoffe der Mittelstufe in einem Kontext präsentiert werden, in dem die schlichte Anwendung von Routinen versagt und die Situation mathematisch rekonstruiert werden muss. In einem Fall wird ein erstes Verständnis nichtlinearen Wachstums und im anderen die Fähigkeit der geometrischen Exploration einer alltäglichen Situation erfasst. Die extrem unterschiedlichen Lösungswahrscheinlichkeiten in den ausgewählten Ländern geben erste Hinweise auf eine differentielle Verständnisorientierung des Mathematikunterrichts (Baumert, Klieme & Watermann 1998; Baumert, Bos & Watermann 2000).

3. Fragen der Inhalts- und Kriteriumsvalidität

Ein weiteres Problem, das insbesondere bei internationalen Vergleichen, aber nicht nur dort, auftritt und häufig gegen standardisierte Leistungsmessung ins Feld geführt wird, ist die Festlegung der inhaltlichen Kriterien für die Aufgabenauswahl eines Tests. Bei Schulleistungstests lassen sich zwei grundsätzlich unterschiedliche Vorgehensweisen unterscheiden: In einem Fall wird curriculare Validität in der Regel auf Lehrplan- und Unterrichtsebene angestrebt. Dies ist das übliche Vorgehen, wenn es sich um Unterrichtsgegenstände handelt, für deren Vermittlung die Schule weitgehend verantwortlich ist (also Mathematik, Physik usw.). Im Vorfeld der Testkonstruktionen werden Unterrichtsstoffe identifiziert, die in möglichst vielen Ländern curriculare Validität beanspruchen können und dann den inhaltlichen Rahmen der Itementwicklung bestimmen. In der Regel wird bei diesem Verfahren ein Kompromiss zwischen kleinstem gemeinsamen Nenner und möglichst breiter Erfassung von einschlägigen Stoffgebieten gesucht. Leitendes Prinzip ist dabei, in ähnlicher Weise unfair gegenüber allen beteiligten Ländern zu sein. Wie gut dieses Verfahren gelingt, hängt vor allen Dingen von dem Grad der latenten internationalen Standardisierung eines Unterrichtsgebietes ab. Es setzt also ein - zumindest bereichsspezifisch - kulturübergreifend geteiltes Verständnis einer modernen Schule, nicht aber die Annahme transkultureller Universalien voraus, wie Eckensberger und Römhild (2000) anzunehmen scheinen. In der letzten Generation der IEA-Studien zur Mathematik und den Naturwissenschaften ist dieses Vorgehen für die Untersuchungen in der Sekundarstufe I und den vorakademischen Bildungsgängen gewählt worden. Im Rahmen von TIMSS ist die internationale curriculare Validierung in Deutschland und den Niederlanden durch eine Unterrichtsvalidierung ergänzt worden. In diesen beiden Ländern wurden Lehrkräfte anhand von Testaufgaben befragt, inwieweit die in den TIMSS-Tests repräsentierten Stoffe tatsächlich im Unterricht unterrichtet worden waren (Baumert, Lehmann u.a. 1997; Klieme 2000). Um die Fairness der Kompromissentscheidung gegenüber den teilnehmenden Ländern zu prüfen, kann man darüber hinaus das Curriculum als variierendes Systemmerkmal betrachten. Dieser Weg ist erstmalig in TIMSS beschritten worden. Nachdem in jedem beteiligten Land die ausgewählten Testaufgaben einer Lehrplanvalidierung unterzogen worden waren, wurden national angepasste Tests konstruiert, die nur die für das jeweilige Land curricular validen Aufgaben enthielten. Anschließend wurden die internationalen Vergleiche mit jeder der nationalen Testversionen wiederholt und die Stabilität der Rangreihen geprüft. Bleiben die Rangreihen stabil, ist dies ein starkes Argument für relativ große interkulturelle Fairness (Beaton, Mullis u.a. 1996; Beaton, Martin u.a. 1996; Arnold 1999).

Ein zweiter Ansatz geht von einem normativ-didaktischen Entwurf als Kriterium der Aufgabenauswahl aus. Dies setzt voraus, dass es eine internationale Verständigung über das normative Konzept gibt. Mit diesem Ansatz wird explizit ein internationales Benchmarking (Vergleichsnormierung) angestrebt. Dieses Vorgehen bietet sich immer dann an, wenn Kompetenzen erfasst werden sollen, die über die Schule hinausreichen, nicht allein in der Schule erworben werden und funktionale Bedeutung im Rahmen der Bewältigung allgemeiner Lebenssituationen haben. Lesekompetenz als Kulturwerkzeug ist ein solches Beispiel. Hier werden in der Regel Anwendungssituationen vorgegeben, an deren Bewältigung die verfügbare Kompetenz abgelesen wird. Im Rahmen von PISA wurde allerdings auch ein normativ-didaktischer Ansatz für den Mathematik- und Naturwissenschaftstest gewählt. Im Falle der Mathematik ist das Testkonzept weitgehend an holländische Vorstellungen von *realistic mathematics* angelehnt, die auf Hans Freudenthal (1977) zurückgehen (vgl. NCTM 2000). In ähnlicher Weise ist der Naturwissenschaftstest funktional angewandt orientiert. Er folgt den von AAAS entwickelten Standards für naturwissenschaftliche Grundbildung (AAAS 1993).

Die funktionale Orientierung des internationalen PISA-Mathematiktests wird vermutlich eine interessante Herausforderung für eingebürgerte epistemologische Überzeugungen über Mathematik und Mathematikunterricht in Deutschland ergeben. An einem Beispiel soll dies verdeutlicht werden. Die in der internationalen PISA-Rahmenkonzeption als Beispiel für eine

mittlere Kompetenzstufe wiedergegebene Aufgabe: "Maria lebt 2 km von der Schule entfernt, Martin 5 km. Wie weit leben Maria und Martin voneinander entfernt?" wurde von einer Referentin während einer PISA-Konferenz als unbrauchbares Mathematikitem bezeichnet, da die Aufgabe keine eindeutige Lösung habe (Demmer, 2000). Hierin kommt eine epistemologische Vorstellung von Mathematik und Mathematiktreiben zum Ausdruck, die Grigutsch (1996) und Törner und Grigutsch (1994) als Schemaorientierung bezeichnet haben. Sie ist für den deutschen Mathematikunterricht selbst in der gymnasialen Oberstufe charakteristisch (Köller, Baumert & Neubrand 2000). Im Rahmen dieser Konzeption hat die mathematische Exploration einer Situation keinen Platz. Gerade diese Fähigkeit erfasst aber das PISA-Item auf einem noch niedrigen mathematischen Anspruchsniveau.

4. Mögliche Erträge von *Large Scale-Schulleistungsstudien*

Häufig wird kritisch gefragt, welchen Nutzen die Schulen und insbesondere die beteiligten Schulen von Untersuchungsprogrammen wie TIMSS oder PISA hätten. In der Regel steht hinter dieser Frage die mehr oder minder explizite Vorstellung, pädagogische Forschung und allemal empirische Forschung habe nicht nur schulspezifische Diagnosen sondern gleichzeitig auch die Anleitung zur Therapie mitzuliefern. Und nur unter dieser Voraussetzung seien Leistungsstudien zu rechtfertigen. Die Frage scheint mir in doppelter Hinsicht falsch gestellt zu sein. Einmal handelt es sich bei Studien wie TIMSS oder PISA nicht um die Evaluation von Einzelschulen, sondern um Untersuchungen, die dem Systemmonitoring dienen und in erster Linie Wissen über Systemzusammenhänge erzeugen. Dieses Wissen wird in der Regel nur vermittelt über politisch administrative Entscheidungen, die Lehrplanarbeit, die Revision von Lehrbüchern oder die Lehrerfortbildung, gelegentlich vielleicht auch durch die Orientierung von Einzelschulen zur Schulentwicklung beitragen, in keinem Fall ersetzt es aber die konstruktive Phantasie, die allemal Grundlage konkreter Handlungsentwürfe ist. Diese aber fallen in den Kernbereich der professionellen Zuständigkeit des Lehrers und der Lehrerin. Aber selbst unter der Perspektive der (wenigen) Schulen, die an einer solchen Stichprobenuntersuchung teilnehmen, sollte die Frage umformuliert werden und lauten: Welchen Gebrauch machen Einzelschulen in professioneller Verantwortung von den Ergebnissen aus Schulvergleichen, die sie als Schulrückmeldungen erhalten? Mit dieser Frage betreten wir weitgehend unbekanntes Land. Erste Studien, die untersuchen, wie Schulen mit rückgemeldeten Ergebnissen dezentraler Evaluationsmaßnahmen umgehen, zeigen, dass es keine automatische Verbindung von Rückmeldung und Schulentwicklung gibt (Specht, Altrichter & Soukup-Altrichter 1998). PISA könnte Anlass sein, der Informationsnutzung in Schulen systematisch nachzugehen. Bei bisherigen internationalen Vergleichsstudien haben die beteiligten Schulen Rückmeldungen über Schulmerkmale und Ergebnisprofile im internationalen, nationalen und regionalen Vergleich sowie unter Berücksichtigung ihrer spezifischen Rahmenbedingungen erhalten.

In jüngster Zeit mehrten sich die Vorschläge, diese Schulrückmeldungen auszubauen und systematisch in Schulentwicklungsprozesse zu integrieren (Rolff 1999; Helmke, 2000). Umso wichtiger wird es, gleichzeitig auf die Grenzen der Aussagefähigkeit und Belastbarkeit solcher Schulrückmeldungen hinzuweisen. Die schulbezogenen Ergebnisse - auch wenn sie ein breites Spektrum von Merkmalen umfassen - basieren in der Regel auf Stichproben einzelner Jahrgänge, Klassen und Fächer, so dass sie keine generalisierten Aussagen über die gesamte Schule erlauben. Denn die Erträge der Arbeit einer Schule können sich von Klasse zu Klasse, von Jahrgangsstufe zu Jahrgangsstufe und von Fach zu Fach unterscheiden (Sammons 1999). Darüber hinaus handelt es sich - von wenigen Ausnahmen abgesehen - fast immer um querschnittlich angelegte Survey-Untersuchungen, bei denen eine ausreichende Kontrolle der individuellen Leistungsvoraussetzungen der Schülerinnen und Schüler sowie der institutionellen Kontextbedingungen praktisch kaum möglich ist, so dass man nicht mit hinreichender Sicher-

heit entscheiden kann, inwieweit die Befunde einer Schule ein Ergebnis der Übergangselektion oder ihrer pädagogischen Arbeit sind.

Der eigentliche Nutzen von zentralen Schulleistungsstudien auf Stichprobenbasis liegt nicht in den einzelschulbezogenen Informationen sondern in den mehr oder minder generalisierbaren deskriptiven und analytischen Befunden, die sich bei Studien wie TIMSS oder PISA in charakteristischer Weise auf fast alle Ebenen des Schulsystems beziehen. Charakteristisch für jeden dieser Befunde ist, dass keiner direkte Entscheidungshilfen liefert, sondern eher die Komplexität von Entscheidungssituationen vergrößert. Die Ergebnisse eröffnen Raum für konstruktive Entwürfe. Dies heißt aber auch, dass Evaluation auf System- ebenso wie auf Schulebene nicht per se praktisch nützlich ist. Nutzen wird erst im reflexiven Gebrauch der Ergebnisse erzeugt. Die eigentliche Arbeit beginnt in der Schule, in den sie unterstützenden Einrichtungen und in der Politik erst nach der Untersuchung.

Literatur

- American Association for the Advancement of Science (Ed.). (1993). *Benchmarks for science literacy. Project 2061*. New York: Oxford University Press.
- Arnold, K.-H.: *Fairneß bei Schulsystemvergleichen: diagnostische Konsequenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation*. Münster: Waxmann, 1999.
- Baumert, J. & Lehmann, R. et al. (1997). *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Klieme, E. & Stanat, P. (im Druck). *PISA (Programme for International Student Assessment) - Zielsetzung, theoretische Konzeption und Entwicklung von Meßverfahren*. Erscheint 2001 in: F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen - Eine Zwischenbilanz*. Weinheim: Beltz
- Baumert, J., Klieme E., Watermann, R. (1998). *Jenseits von Gesamttest- und Untertestwerten: Analyse differentieller Itemfunktionen am Beispiel des mathematischen Grundbildungstests der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie der IEA (TIMSS)*. *Schulpädagogik und Lehrerbildung. Festschrift zum 60. Geburtstag von Josef Thonhauser*. Herber, H.-J. & Hofmann, F. (Hrsg.), Innsbruck, Wien: StudienVerlag, 301-324.
- Baumert, J., Bos, W. & Watrmann (2000). *Mathematisch-naturwissenschaftliche Grundbildung im internaitonalen Vergleich*. In J. Baumert, W. Bos & R. Lehmann (Hrsg.). *Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Kapitel IV in Band I: TIMSS - Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. S. 135-197. Opladen: Leske + Budrich.
- Blum, W. (im Druck). *Was folgt aus TIMSS für Mathematikunterricht und Mathematiklehrausbildung?* Erscheint in E. Klieme & J. Baumert (Hrsg.). *Mathematik und Naturwissenschaften im Schulunterricht - Bestandsaufnahme und pädagogische Konsequenzen auf der Basis von TIMSS*. Bonn: Bundesministerium für Bildung und Forschung
- Demmer, M. (2000). *PISA-INFO 05/2000: Qualitätsdebatte*. Informationen der Gewerkschaft Erziehung und Wissenschaft, Vorstandsbereich Schule, 13.4.2000.
- Eckensberger, L. & Römhild, R. (2000). *Kulturelle Einflüsse*. 12. Kapitel in M. Amelang (Hrsg.) *Determinanten individueller Unterschiede*. Band 4 der Enzyklopädie der Psychologie. Göttingen u.a.: Hogrefe - Verlag für Psychologie.
- Fitz-Gibbon, C. (1996). *Monitoring school effectiveness: Simplicity and complexity*. In J. Gray, D. Reynolds, C. Fitz-Gibbon & D. Jesson, D. (Hrsg.). *Merging traditions: The future of research on school effectiveness and school improvement*. London: Cassell.
- Freudenthal, H. (1977). *Mathematik als pädagogische Aufgabe* (Bd. 1, Bd, 2) Stuttgart: Ernst Klett Verlag.
- Gray, J., Reynolds, D., Fitz-Gibbon, C. & Jesson, D. (Hrsg.) (1996). *Merging traditions: The future of research on school effectiveness and school improvement*. London: Cassell.

- Grigutsch, S. (1996). Mathematische Weltbilder von Schülern: Struktur, Entwicklung, Einflußfaktoren. Unveröffentlichte Dissertation vom Fachbereich 11/Mathematik der Gerhard-Mercator-Universität – Gesamthochschule Duisburg
- Grogger, G. & Specht, W. (1999). Evaluation und Qualität im Bildungswesen. Problemanalyse und Lösungsansätze am Schnittpunkt von Wissenschaft und Bildungspolitik. Dokumentation eines internationalen Workshops in Blumau/Steiermark, 18. bis 21. Februar. Graz: Bundesministerium für Unterricht und kulturelle Angelegenheiten.
- Helmke, A. (2000). TIMSS und die Folgen. Der weite Weg von der externen Leistungsevaluation zur Verbesserung des Lehrens und Lernens. In U. P. Trier (Hrsg.), Bildungswirksamkeit zwischen Forschung und Politik. Bern: Rüegger.
- Joint Committee of Standards for Educational Evaluation (1994). The program evaluation standards. How to assess evaluations of educational programs. Thousand Oaks: Sage.
- Klieme, E. (2000): Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In: Baumert, J., Bos, W. & Lehmann, R. (Hrsg.): Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Kapitel II in Band II: TIMSS - Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. S. 57-128. Opladen: Leske + Budrich.
- Klieme, E., Baumert J. & Schwippert, K. (2000). Schulbezogene Evaluation und Schulleistungsvergleiche. Eine Studie im Anschluss an TIMSS (erscheint in Rolff, H. G., Bos, W., Klemm, K., Pfeiffer, H. & Schulz-Zander, R. (Hrsg.). Jahrbuch der Schulentwicklung, Band 11. München Juventa
- Klieme, E., Baumert, J., Köller, O. & Bos, W. (2000). Mathematisch-naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.). Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Kapitel III in Band I: TIMSS - Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. S. 85-133. Opladen: Leske + Budrich.
- Köller, O. & Trautwein, U. (in Vorbereitung). Mehr als nur eine Momentaufnahme: Möglichkeiten, TIMSS als Basis für die Schuldiagnostik und Schulentwicklung zu nutzen.
- Köller, O., Baumert, J. & Neubrand, J. (2000). Epistemologische Überzeugungen und Fachverständnis im Mathematik- und Physikunterricht. In J. Baumert, W. Bos & R. Lehmann (Hrsg.). Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Kapitel VI in Band II: TIMSS - Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. S. 229-269. Opladen: Leske + Budrich.
- Lehmann, R. H., Peek, R., Pieper, I. & Stritzky, R. von (1995). Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen. Weinheim: Beltz.
- MacPherson, A. (1992). Measuring added values in schools. London: National Commission on Education.
- National Council of Teachers of Mathematics (NCTM) (2000). Principles and standards for school mathematics. Reston, VA: NCTM
- OECD: Measuring student knowledge and skills. A new framework for assessment. Paris, Organisation for Economic Co-Operation and Development, 1999.
- Popham, W. J. (1999). Why standardized test scores don't measure educational quality. Educational Leadership, 56 (6), 8-15.
- Posch, P. & Altrichter, H. (1997). Möglichkeiten und Grenzen der Qualitätsevaluation und Qualitätsentwicklung im Schulwesen. Innsbruck/Wien: Bundesministerium für Unterricht und kulturelle Angelegenheiten.
- Powers, D. E. & Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. Journal of Educational Measurement, 36 (2), 93-118.
- Ramseier, E., Keller, C. & Moser, U. (1999). Bilanz Bildung. Eine Evaluation am Ende der Sekundarstufe II auf der Grundlage der Third International Mathematics and Science Study. Zürich: Rüegger.

- Rolff, H.-G. (1995). Steuerung, Entwicklung und Qualitätssicherung von Schulen durch Evaluation. In H.-G. Rolff (Hrsg.), *Zukunftsfelder von Schulforschung*. Weinheim.
- Rolff, H.-G. (1999). PISA Initial Overall Evaluation. Gutachten. Erstellt im Auftrag der OECD für Treffen des Board of Participating Countries vom 13. bis 15. März 2000 in Melbourne.
- Sammons, P. (1999). *School effectiveness. Coming of age in the twenty-first century*. Lisse u.a.: Swets & Zeitlinger.
- Specht, W., Altrichter, H. & Soukup-Altrichter, K. (1998). *Qualitätsentwicklung mit Programm. Endbericht über die Begleitevaluation der Pilotphase. Report Nr. 41 des Zentrums für Schulentwicklung*. Graz: Bundesministerium für Unterricht und kulturelle Angelegenheiten.
- Strittmatter, A. (1999). *Qualitätsevaluation und Schulentwicklung*. In: J. Thonhauser & J.-L. Patry (Hrsg.), *Evaluation im Bildungsbereich*. Innsbruck-Wien: Studien Verlag.
- Törner, G. & Grigutsch, S. (1994). „Mathematische Weltbilder“ bei Studienanfängern – eine Erhebung. *Journal für Mathematik-Didaktik*, 15, 211–251.
- Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M. & Nauels, H.-U. (1998). *Evaluation des Tests für medizinische Studiengänge (TMS): Synopse der Ergebnisse*. Bonn: Institut für Test- und Begabungsforschung.(1998).
- Watermann, R. & Baumert, J. (2000). *Mathematisch-naturwissenschaftliche Grundbildung beim Übergang von der Schule in den Beruf*. In J. Baumert, W. Bos & R. Lehmann (2000). *Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Kapitel V in Band I: TIMSS - Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. S. 199-259. Opladen: Leske + Budrich.
- Woodhouse, G. & Goldstein, H. (1988). *Educational Performance Indicators and LEA League Tables*. *Oxford Review of Education*, 14 (3), 301-320.
-

Interna:

1. Die ÖFEB hat in ihrer 4. virtuellen Vorstandssitzung beschlossen, zu den bereits bestehenden Sektionen
 - Sektion Lehrerbildung und Lehrerbildungsforschung (B. Hackl und A. Kowarsch)
 - Sektion Unterrichtsforschung und Unterrichtsentwicklung (A. Seel)
 - Sektion empirische pädagogische Forschung (F. Eder)
 - Sektion Schulforschung und Schulentwicklung (H. Altrichter und M. Krainz-Dürr) und die
 - Sektion Sozialpädagogik (J. Scheipl)noch die
 - *Sektion Erwachsenenbildung* (E. Gruber und W. Lenz) und die
 - *Sektion Medienpädagogik* (B. Hipf und T. Hug) einzurichten.
2. Professor Reinhard Fatke, Universität Zürich, hat im März dieses Jahres in
- zwei Briefen angefragt, ob die ÖFEB bereit sei, den im Jahre 2004 voraussichtlich in Zürich stattfindenden Kongress der Deutschen Gesellschaft für Erziehungswissenschaft (DgE) gemeinsam mit der Schweizerischen Gesellschaft für Bildungsforschung (SGBF) zu veranstalten. Voraussetzung sei allerdings, dass die ÖFEB auch finanzielle Mittel einbringen könne. Der Vorstand hat sich dafür ausgesprochen, diesem Angebot näher zu treten.
3. Theo Hug hat an die ÖFEB die Anfrage gerichtet, ob diese bereit sei, gemeinsam mit dem Institut für Erziehungswissenschaft der Universität Innsbruck als Veranstalter der unter seiner Leitung stattfindenden Tagung „*Wie kommt Wissenschaft zu ihrem Wissen*“ (Ende 2001 / Anfang 2002) aufzutreten. Der Vorstand der ÖFEB hat sich für die Annahme dieser Einladung ausgesprochen.

Berichte:



IMST² (Innovations in Mathematics, Science and Technology Teaching).

Ein fachdidaktisches Projekt

Im Auftrag des Bildungsministeriums führt das Interuniversitäre *Institut für Interdisziplinäre Forschung (IFF)* unter der Leitung von Ao. Univ.-Prof. Dr. Konrad Krainer, Abteilung Schule und gesellschaftliches Lernen, das Projekt IMST² zur Weiterentwicklung der mathematisch-naturwissenschaftlichen Fächer in den Oberstufen durch. Mittelbar gaben die schlechten Resultate der österreichischen Schüler/innen der Sekundarstufe II bei *TIMSS (Third International Mathematics and Science Study)* den Anstoß zu diesem Projekt.

Um vor allem die selbstständige Problemlösungskompetenz von Schülerinnen und Schülern im mathematisch-naturwissenschaftlichen Bereich zu fördern, werden im Rahmen von IMST² Maßnahmen getroffen, die eine nachhaltige Qualitätsentwicklung im Unterricht versprechen. Neue Medien und Technologien, professioneller Erfahrungsaustausch unter den Lehrenden, Sammlung und Dokumentation von innovativem Unterricht („good practice“) und neue Schwerpunktsetzungen bei Fragen der Grundbildung sind wesentliche Inhalte der Initiative.

Je nach Grad der Beteiligung können Schulen verschiedene Formen der Mitarbeit bei IMST² in Anspruch nehmen. Der Bogen reicht von unverbindlicher Information (*Informationsschulen*) über gezielte Beteiligung (*Kontaktsschulen*, *Kooperationsschulen*) bis hin zu Schwerpunktsetzungen für die Projekte der Initiative (*Schwerpunktschulen*). Weit über hundert

allgemeinbildende und berufsbildende Schulen aus dem gesamten Bundesgebiet haben sich für eine Mitarbeit interessiert.

Im Rahmen des Pilotprojekts wurden vier *Schwerpunktprogramme* eingerichtet:

Das Schwerpunktprogramm ***Grundbildung (S1)*** beschäftigt sich u.a. mit folgenden Fragen:

Über welches mathematische und naturwissenschaftliche Wissen soll ein/e erfolgreiche/r Maturant/-in verfügen? Welche Fähigkeiten aus diesen Bereichen sind für die persönliche Entfaltung wertvoll und für die gesellschaftliche Entwicklung wichtig? Wie lernen Schüler/-innen, etwas zu verstehen, Begriffe zu verwenden, Zusammenhänge und Bedeutungen zu konstruieren? Wie können sie eine wertschätzende und zugleich kritisch-prüfende Einstellung zum Wissen entwickeln? Wie kann ihr Lernerfolg überprüft werden?

Lehrer/-innengruppen, die sich für eine Weiterentwicklung der Lern- und Prüfungskultur interessieren, sind zu einem Dialog mit dem Team des Schwerpunktprogramms über Fragen der Grundbildung eingeladen. Für interessierte Fachgruppen besteht die Möglichkeit zur Zusammenarbeit mit Kolleginnen und Kollegen der Fächer Biologie / Umweltkunde, Chemie, Mathematik und Physik.

Das zweite Schwerpunktprogramm ***Schulentwicklung (S2)*** beschäftigt sich u.a. mit folgenden Fragen:

Wie kann der Stellenwert von Mathematik und Naturwissenschaft an den Schulen gehoben werden? Wie können Unterrichts- und Schulentwicklung adäquat verbunden werden? Wie können mathematische und naturwissenschaftliche Schwerpunktsetzungen an Schulen initiiert und unterstützt werden? Wie können die in der Schulentwicklung gewonnenen Erfahrungen für andere Schulen genutzt werden? Wie kann die Qualität standortbezogener Schwerpunktsetzung (Schulprogramm) dokumentiert und evaluiert werden?

Dieses Schwerpunktprogramm will mathematische und naturwissenschaftliche Schwerpunktsetzungen unterstützen, indem Projektgruppen an den Schulen bei der Entwicklung von Schulprogrammen begleitet werden.

Beim Schwerpunktprogramm ***Geschlechtssensibler Unterricht (S3)*** geht es u.a. um folgende Fragen:

Wie kann die Schul- und Unterrichtskultur dazu beitragen, dass die Beziehung zur Mathematik, und zu den Naturwissenschaften nur mehr eine Frage der individuellen Neigungen ist und nicht eine des Geschlechts? Welche geschlechtssensiblen Angebote gibt es bereits? Welchen Einfluss haben solche Entwicklungen auf die allgemeinen fachlichen Zielsetzungen?

Dieses Schwerpunktprogramm befasst sich mit Gemeinsamkeiten und Unterschieden beim Mathematik- und Physiklernen der Mädchen und Burschen. Es sollen Lernumgebungen ausfindig gemacht werden, die den jeweiligen Lernbedürfnissen Rechnung tragen. Darauf aufbauend sollen dann konkrete Maßnahmen entwickelt, die Lehrer/-innen bei der Durchführung begleitet und die Erfahrungen fachdidaktisch reflektiert werden.

Im Rahmen dieses Schwerpunktprogramms werden u. a. Möglichkeiten zur Zusammen-

arbeit bestehen, hier werden u.a. Mitarbeiter/innen für die Erstellung einer CD-ROM zur fachdidaktischen Aus- und Fortbildung von Mathematik- und Physiklehrer/-innen gesucht.

Der letzte Schwerpunkt, ***Forschung und Entwicklung (S4)***, hat die Untersuchung der folgenden Fragen zur Aufgabe:

Wie können Oberstufenschüler/-innen zu selbstständigem und eigenverantwortlichem Arbeiten motiviert werden? Was sind Kontexte, in denen kreative und aktive Denkleistungen gefördert werden? Welche Unterrichtsformen unterstützen effektive Lernprozesse?

In seiner ersten Phase initiiert, fördert und finanziert dieses Schwerpunktprogramm Vorhaben von Lehrer/-innen, aber auch von Vertreterinnen und Vertretern der universitären Fachdidaktiken, die sich insbesondere die Entwicklung des Unterrichts in Richtung selbständiges und eigenverantwortliches Arbeiten und die Erforschung ihrer Bedingungen vorgenommen haben. Das kann von der Erprobung neuer Unterrichtsideen über die Selbstevaluation des eigenen Unterrichts bis hin zu größeren Forschungsprojekten reichen.

Informationen sind erhältlich bei:

Josef Hödl-Weißhofer
IFF/Abt. Schule und gesellschaftliches Lernen
Sterneckstraße 15, 9020 Klagenfurt

Projekt IMST²

Tel. 0463/2700-6138 oder 6139
E-mail: imst@uni-klu.ac.at
Homepage: <http://imst.uni-klu.ac.at>

Ankündigung:

Wie kommt Wissenschaft zu Wissen? **Tagungsankündigung und Einladung**

Nach der Veröffentlichung der Buchreihe und der CD-ROMs sind wir nun mit den Vorbereitungen einer Tagung befasst, zu der wir herzlich einladen.

Die Tagung „**Wie kommt Wissenschaft zu Wissen?**“ wird am **30.11./1.12.2001** am **Institut für Erziehungswissenschaften der Universität Innsbruck**

<<http://www.uibk.ac.at/c/c6/c603/>> in Kooperation mit der *Österreichischen Gesellschaft für Forschung und Entwicklung im Bildungswesen (ÖFEB)*

<<http://paedpsych.jk.uni-linz.ac.at/INTERNET/OEFEB/>> und der *Gesellschaft für Psychoanalyse (VSFI)* <<http://vsfi.uibk.ac.at/site/gfp/gfp.html>> durchgeführt.

Kontext

Im Übergang zum 21. Jahrhundert wird zunehmend deutlich, dass sich mit den technologischen, den sozio-kulturellen, den ökonomischen und den politischen Veränderungen auch die Lage des Wissens gewandelt hat. Die wissenschaftlichen Institutionen sind mit diesen Dynamiken eng verknüpft. Einerseits ermöglichen sie diese Veränderungsprozesse und tragen sowohl zur Problemverursachung als auch zur Problembewältigung bei. Andererseits verändern sie sich dabei auch selbst und mit ihnen die Art und Weise, wie wissenschaftliches Wissen hervorgebracht, diskutiert, organisiert, kritisiert, verbreitet und tradiert wird.

Vor allem die Sozial- und Kulturwissenschaften stehen heute in einem vielfältigen Spannungsfeld zwischen Tradition und Innovation, Stabilität und Reform, Fragmentierung und Systematisierung, Integration

und Differenzierung. Dies zeigen die zahlreichen Abhandlungen zur Identitäts- und Profilbildung, die disziplinübergreifenden Bemühungen und nicht zuletzt die Beiträge der gleichnamigen Buchreihe.

Themenspektrum

Die Leitfrage 'Wie kommt Wissenschaft zu Wissen?' geht nicht von einer vorgefertigten Wissensordnung aus, in der alle Teilbereiche von legitimierte SprecherInnen ihren Platz zugewiesen bekommen. Wissenschaft wird hier weniger als statisches Gebilde, sondern vielmehr als pulsierender Diskurszusammenhang aufgefasst. Dabei bietet die Integration der Informations- und Kommunikationsangebote mittels Buchreihe, CD-ROM und Internetdiensten vielfältige Möglichkeiten des Erwerbs von einschlägigen Kompetenzen im Hinblick auf wissenschaftliche Arbeitsweisen, Methoden, Paradigmen, Konzeptionen und Anwendungsformen.

Die Tagung geht von den thematischen Akzentuierungen der Buchreihe aus, die das gesamte Spektrum der Wie-Fragen ins Blickfeld rückt:

- Einführung in das wissenschaftliche Arbeiten
- Einführung in die Forschungsmethodik und Forschungspraxis
- Einführung in die Methodologie der Sozial- und Kulturwissenschaften
- Einführung in die Wissenschaftstheorie und Wissenschaftsforschung

ReferentInnen & WorkshopleiterInnen

Vorläufige Zusagen für Hauptvorträge bzw. für Round Tables oder Workshops liegen vor von Dr. Roland Benedikter (I), PD Dr. Gert Dressel (A), Dr. Eva Fleischer (A), Prof. Dr. Martina Kaller-Dietrich (A), Dr. Dieter Korczak (D), Prof. Dr. Otto Kruse (D), Mag. Thomas Lerche (D), PD Dr. Dieter Rainer Leschke (D), PD Dr. Reinhard Margreiter (A), Prof. Dr. Dieter Münch (D), Dr. Klaus Niedermair (A), Prof. Dr. Gertrud Simon (A), Prof. Dr. Dieter Spanhel (D), Dr. Stefan Weber (A) und Dr. Siegfried Winkler (A).

Organisation

Tagungsleitung:

A. Univ. Prof. Dr. Theo Hug

Infos, Anmeldung und Organisation:

Nicole Bickel <nicole.bickel@uibk.ac.at>

Die WWW-Infos zur Tagung werden nach Zusage der Fördermittel und Konkretisierung der Themen monatlich aktualisiert.



Ö F E B

Österreichische
Gesellschaft für
Forschung und
Entwicklung im
Bildungswesen

[Tagungs-
programm](#)

[Tagungsort](#)

[Vorträge](#)

[Symposien](#)

[Workshops](#)

[Poster](#)

[Kontakt](#)

[ÖFEB](#)

1. Tagung der ÖFEB
von 24. bis 26. September 2001
in Salzburg

Erziehung und Bildung zwischen
Selbstverantwortung und Rechenschaftspflicht:
Kompetenz - Kooperation - Professionalisierung

VeranstalterInnen:

[Institut für Erziehungswissenschaft, Universität Salzburg](#)
[Institut für Lehrerinnen- und Lehrer-Bildung, Universität Salzburg](#)
[Pädagogische Akademie des Bundes in Salzburg](#)
[Pädagogisches Institut des Bundes in Salzburg](#)

last update: 30-8-2001 [kirstin eckstein](#)

Diese Homepage mit aktuellen Informationen finden Sie unter
http://www.sbg.ac.at/erz/oefeb_tagung