

Watermann, Rainer; Stanat, Petra ; Kunter, Mareike ; Klieme, Eckhard ; Baumert, Jürgen
**Schulrückmeldungen im Rahmen von Schulleistungsuntersuchungen: Das
Disseminationskonzept von PISA-2000**

Zeitschrift für Pädagogik 49 (2003) 1, S. 92-111

urn:nbn:de:0111-opus-38708

in Kooperation mit / in cooperation with:

BELTZ

<http://www.beltz.de>

Nutzungsbedingungen / conditions of use

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

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

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

Kontakt / Contact:

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

Inhaltsverzeichnis

Thementeil:

Bildungssystem im Wandel – Zwischen Eigendynamik, Politik und Pädagogik

Jörg-W. Link/Axel Nath/Heinz-Elmar Tenorth

Bildungssystem im Wandel – Zwischen Eigendynamik, Politik und Pädagogik.
Zur Einführung in den Themenschwerpunkt 1

Axel Nath

Bildungswachstum und äußere Schulreform im 19. und 20. Jahrhundert.
Individualisierung der Bildungsentscheidung und Integration der
Schulstruktur 8

Ulrich G. Herrmann

Kommentar zum Beitrag von Axel Nath 26

Peter Lundgreen

„Bildungspolitik“ und „Eigendynamik“ in den Wachstumsschüben
des deutschen Bildungssystems seit dem 19. Jahrhundert 34

Hanno Schmitt

Kommentar zum Beitrag von Peter Lundgreen 42

Gerhard Kluchert

Die Entwicklung der Lehrer-Schüler-Interaktion und die
Bildungswachstumsschübe. Zur inneren Schulreform im 20. Jahrhundert 47

Heidemarie Kemnitz

Kommentar zum Beitrag von Gerhard Kluchert 61

Heinz-Elmar Tenorth

Wachstumsschübe des Bildungssystems und Konjunkturen seiner
Thematisierung. Über Kontinuität und Variation pädagogischer Reflexion 69

Peter Drewek

Kommentar zum Beitrag von Heinz-Elmar Tenorth 86

Allgemeiner Teil

Rainer Watermann/Petra Stanat/Mareike Kunter/Eckhard Klieme/Jürgen Baumert
Schulrückmeldungen im Rahmen von Schulleistungsuntersuchungen:

Das Disseminationskonzept von PISA-2000 92

Lilian Fried

Pädagogisches Professionswissen als Form und Medium der Lehrerbildungs-
kommunikation – empirische Suchbewegungen

112

Diskussion

Kai S. Cortina

Der Schulartwechsel in der Sekundarstufe I: Pädagogische Maßnahmen oder
Indikator eines falschen Systems?

128

Ernst Rösner

Schulartwechsel in der Sekundarstufe I. Eine Erwiderung auf Kai S. Cortina

143

Besprechungen

Hans-Werner Fuchs

Rudolf Tippelt (Hrsg.): Handbuch Bildungsforschung

145

Andreas Helmke

Christoph von Burkard/Gerhard Eikenbusch: Praxishandbuch Evaluation
in der Schule

148

Dietrich Benner

Niklas Luhmann: Das Erziehungssystem der Gesellschaft

151

Alois Suter

Helmut Heiland: Die Spielpädagogik Friedrich Fröbels

155

Dokumentation

Pädagogische Neuerscheinungen

157

Content

Topic: Processes of Change in the Educational System – between internal dynamic, politics and pedagogy

Jörg-W. Link/Axel Nath/Heinz-Elmar Tenorth

Processes of Change in the Educational System – between internal dynamic, politics and pedagogy. An introduction 1

Axel Nath

Educational Expansion and Organizational Reforms of Schooling during the 19th and 20th Centuries – Individualization of educational career decisions and integration of the school structure 8

Ulrich G. Herrmann

A Commentary on the contribution by Axel Nath 26

Peter Lundgreen

“Educational Policy” and “Autonomous Momentum” in the Growth of the Educational System since the 19th Century 34

Hanno Schmitt

A commentary on the contribution by Peter Lundgreen 42

Gerhard Kluchert

The Development of Teacher-Student-Interaction and Periods of Educational Growth – On internal school reform during the 20th century 47

Heidemarie Kemnitz

A commentary on the contribution by Gerhard Kluchert 61

Heinz-Elmar Tenorth

Periods in the Growth of the Educational System and Peaks In its Topicalization – On the continuity and variation of pedagogical reflection 69

Peter Drewek

A commentary on the contribution by Heinz-Elmar Tenorth 86

Articles

Rainer Watermann/Petra Stanat/Mareike Kunter/Eckhard Klieme/Jürgen Baumert

Feedback to Schools in the Context of Large Scale Assessment – Disseminating the results of PISA-2000 92

Lilian Fried

Pedagogical Professional Knowledge as a Form and a Medium of Communication in Teacher Education – Empirical Soundings 112

Diskussion

Kai S. Cortina

Switching School Types During Lower Secondary Education: Pedagogical measure
or indicator of a wrong system? 128

Ernst Rösner

Switching School Types During Lower Secondary Education.
A reply to Kai S. Cortina 143

Book Reviews 145

New Books 157

Rainer Watermann/Petra Stanat/Mareike Kunter/Eckhard Klieme/Jürgen Baumert

Schulrückmeldungen im Rahmen von Schulleistungsuntersuchungen: Das Disseminationskonzept von PISA-2000

Zusammenfassung: Mit Rückmeldungen schulbezogener Befunde aus Schulleistungsuntersuchungen wird häufig die Erwartung verknüpft, dass sie Schulentwicklungsprozesse in Gang setzen könnten. Es wird unzureichend berücksichtigt, dass der potenzielle Nutzen solcher Schulrückmeldungen vom jeweiligen Untersuchungsdesign abhängt und die Verarbeitung der Ergebnisse nicht unbedingt problemlos verläuft. In diesem Artikel wird eine realistische Position im Umgang mit Schulrückmeldungen beschrieben. Es wird dargestellt, wie in PISA unter Berücksichtigung vorliegender Implementationserfahrungen und Forschungsbefunde Schulrückmeldungen als Teil eines breiteren Disseminationskonzepts untersuchungsspezifisch entwickelt wurden.

1. Einleitung

Schulrückmeldungen gehören mittlerweile zum Standardrepertoire empirischer Schulforschung. Es gibt wohl keine größere Schuluntersuchung, bei der nicht die beteiligten Schulen in der einen oder anderen Form über Ergebnisse der Untersuchung unterrichtet worden wären. Neu sind jedoch die Erwartungen, die an diese Vorgehensweise geknüpft werden. Die Skepsis, die dem Nutzen von standardisierten Leistungsvergleichen für die Einzelschule in der Vergangenheit entgegengebracht wurde, scheint tendenziell der Annahme gewichen zu sein, solche Rückmeldungen könnten Hebelwirkungen für die Einleitung von Schulentwicklungsprozessen haben, die nicht nur zu einer Verbesserung der internen Arbeitsabläufe, sondern auch zu einer Verbesserung des Unterrichts führen. Gelegentlich wird sogar aus der Schulrückmeldung die Legitimation empirischer Schulforschung, zumindest für *Large-Scale Assessments* abgeleitet. Bemerkenswert ist dabei die Undifferenziertheit der Erwartungen: Es wird weder der Erfahrungsfundus mit Schulrückmeldungen in anderen Staaten – der sehr schnell Realismus einkehren ließ – berücksichtigt, noch werden die Ergebnisse der *School-Effectiveness-* und *School-Improvement-*Forschung rezipiert (z.B. Fitz-Gibbon 1996; Scheerens/Bosker 1997; Strittmatter 1999; Teddlie/Reynolds 2000). Ferner werden Schulrückmeldungen unabhängig von der Funktion der jeweiligen Untersuchung und ihrem Design unter einer Kategorie subsumiert, als wäre es bedeutungslos, ob die Ergebnisse aus Querschnitt- oder Längsschnittstudien, aus Stichproben- oder Totalerhebungen stammen. Schließlich wird mit den an Schulrückmeldungen verbundenen Erwartungen unterstellt, dass die dargestellten Ergebnisse in professioneller und problemloser Weise rezipiert und verarbeitet werden.

Im vorliegenden Beitrag wird eine realistische Position im Umgang mit Schulrückmeldungen beschrieben. Um diese Position zu verdeutlichen, wird in Abschnitt 2 zunächst der spezifische Stellenwert von Systemmonitoring-Studien wie PISA als Beitrag

der empirischen Bildungsforschung zur Qualitätssicherung und -entwicklung erörtert (für eine ausführliche Beschreibung vgl. Klieme/Baumert/Schwippert 2000). Es wird darauf hingewiesen, dass es insbesondere im Falle von Systemmonitoring-Untersuchungen keine überzeugenden Argumente für die Annahme gibt, Schulrückmeldungen würden automatisch Innovations- und Optimierungsprozesse an Schulen nach sich ziehen (Rolff i. Vorb.; Terhart 2002). Ferner wird betont, dass die Belastbarkeit und Generalisierbarkeit von schulspezifischen Befunden je nach Untersuchungstyp außerordentlich unterschiedlich sind. Dies gilt für Beschreibungen des Ist-Zustands sowie, in verstärktem Maße, für Rückschlüsse auf die Qualität von schulinternen Arbeitsprozessen. Bleiben die Besonderheiten der jeweiligen Studie unberücksichtigt, kommt es leicht zu einer Unterschätzung der Schwierigkeiten und Probleme bei der Rezeption und Verarbeitung von Schulrückmeldungen. Mit diesen skeptischen Argumenten soll der potenzielle Wert von Schulrückmeldungen nicht grundsätzlich in Zweifel gezogen werden. Sie sollen jedoch Erwartungen auf ein vertretbares Maß zurückführen und insbesondere den Blick für inhaltliche, prozessuale und organisatorische Voraussetzungen schärfen, die bei einer produktiven Nutzung von Schulrückmeldungen erfüllt sein müssen.

Ausgehend von diesen theoretischen und methodischen Überlegungen informiert Abschnitt 3 über den Aufbau und die Gestaltung der Schulrückmeldungen im Rahmen von PISA-2000. Dabei wird insbesondere erörtert, wie versucht wurde, eine angemessene Interpretation der schulbezogenen Ergebnisse vor dem Hintergrund der Ziele, der Anlage und der allgemeinen Befunde der Studie zu gewährleisten.

Abschnitt 4 thematisiert die Rolle der Schulrückmeldungen im Rahmen des allgemeinen Disseminations- und Rezeptionsprozesses von PISA-2000 und stellt die Schulung der PISA-Moderatoren vor. Mit dem Training wurden von den Ländern der Bundesrepublik ausgewählte Personen im Umgang mit den PISA-Befunden und den Schulrückmeldungen ausgebildet, um den Prozess der Rezeption und Verarbeitung der Schulrückmeldungen auf Nachfrage der Schulen unterstützen zu können. Abschließend werden Möglichkeiten der Verknüpfung von Systemmonitoring und der Evaluation von Einzelschulen diskutiert.

2. Funktionen, Prinzipien und Probleme von Schulrückmeldungen im Rahmen von PISA-2000

Schulleistungsuntersuchungen, die dem Begriff des Systemmonitoring zugeordnet werden, leisten einen spezifischen Beitrag zur Qualitätssicherung und -entwicklung in der pädagogischen Praxis. Charakteristisch für diese Studien ist, dass sie in vergleichender Absicht Klassen, Schulen, Schulformen und Schulsysteme auf der Grundlage zentral administrierter, standardisierter Tests untersuchen (Peek 2001a). Sie liefern auf der Bildungssystemebene Informationen über Kompetenzen und Einstellungen über Rahmenbedingungen von Schule und Unterricht sowie deren Zusammenhänge. Die Erträge dieser Studien liegen primär in Basisinformationen, die für die Bewertung und Steuerung auf Systemebene nützlich sind und weniger in der Evaluation von Einzelschulen

(Klieme/ Baumert/Schwippert 2000). In diesen Fällen liegt es deshalb auch nahe, die Effizienz der Stichprobe zu erhöhen, indem keine intakten Klassen oder Jahrgänge getestet, sondern pro Schule zufällig Schülerinnen und Schüler für den Test ausgewählt werden. Solche stichprobenbasierten Untersuchungen genügen dem Anspruch eines Systemmonitorings völlig (Baumert u.a. 2001).

Die PISA-2000-Studie ist an den Zielen und Anforderungen eines Systemmonitorings ausgerichtet konzipiert und durchgeführt worden. Dementsprechend sind die Adressaten von PISA nicht einzelne Schulen, sondern Akteure auf der Ebene des Bildungssystems. Das Rückmeldungskonzept von PISA-2000 hat diesen Aspekten Rechnung zu tragen, vor allem was die Funktionen und die Belastbarkeit der rückgemeldeten Informationen an die Schulen betrifft.

2.1 *Funktionen der Schulrückmeldungen*

Den an PISA beteiligten Schulen ist zugesagt worden, sie über die Ergebnisse des Projekts im Allgemeinen und ihrer Schule im Besonderen zu informieren, sofern die Schulen selbst dies wünschen. Die Rückmeldung, die von der weit überwiegenden Mehrheit der teilnehmenden Schulen angefordert wurden, soll folgende Funktionen erfüllen:

- a) Sie lässt die Personen und Institutionen, die mit ihrer Beteiligung die Untersuchung erst ermöglicht und dabei Belastungen in Kauf genommen haben, am Prozess der Auswertung und Nutzung der Untersuchungsergebnisse teilhaben. Dies ist nicht nur forschungsethisch wünschenswert, sondern dient auch der Akzeptanz von Qualitätssicherungsmaßnahmen im Schulbereich.
- b) Sie ist ein wesentliches Element der Disseminationsstrategie für PISA, denn mit mehr als 1400 Schulen bundesweit wird ein außergewöhnlich breiter Adressatenkreis erreicht, bei dem von hohem Interesse an Resultaten der Studie ausgegangen werden kann.
- c) Die mitgeteilten Daten bieten eine Momentaufnahme von Eingangsvoraussetzungen (z.B. sozialer Hintergrund der Schülerschaft), Prozessvariablen (z.B. Motivationsprofil der Schüler) und Leistungsstand eines Jahrgangs an der betreffenden Schule. Diese Befunde können – auch wenn sie nur begrenzt verallgemeinerbar sind – Anlass für pädagogische Diskussion in der Einzelschule sein.
- d) Die Schulen sollen sich im Umfeld der national repräsentativen PISA-Stichprobe einordnen und mit anderen Schulen, deren Eingangsbedingungen ähnlich sind, vergleichen können („Benchmarking“).

Dabei ist allerdings erneut zu betonen, dass der Ertrag von PISA vor allem in allgemeinen Erkenntnissen über die Bedingungen und Ergebnisse schulischer Lehr-Lern-Prozesse sowie der Identifikation möglicher Ansatzpunkte für Verbesserungen liegt (vgl. Pekrun 2002). Vor diesem Hintergrund legt das nationale PISA-Konsortium großen Wert darauf, dass in der bildungspolitischen Diskussion, in der Öffentlichkeit und von den Adressaten wahrgenommen wird, welche Ziele mit einer punktuellen Rückmeldung

schulbezogener Ergebnisse aus einer Stichprobenuntersuchung *nicht* angestrebt werden können und sollen:

- e) Die Rückmeldung stellt keine Evaluation der Einzelschule dar. Eine explizite Bewertung der Arbeitsergebnisse einer Schule würde ein Untersuchungsdesign voraussetzen, das erstens den pädagogischen Zielen der Schule entspricht und von der Schule mitgetragen wird und zweitens längsschnittlich angelegt ist, sodass Ergebnisse der schulischen Arbeit zugerechnet werden können. Dies ist im Rahmen der PISA-Studie, die sich auf die Erfassung zentral vorgegebener Komponenten beschränkt, jeweils nur eine kleine Schülerstichprobe untersucht und querschnittlich angelegt ist, nicht möglich, aber auch nicht nötig.
- f) Die Schulrückmeldung darf auch nicht mit der Erwartung belastet werden, sie könne unmittelbar Prozesse der Schulentwicklung in Gang setzen. Erfahrungen aus anderen Ländern (z.B. *California Assessment Program [CAP]*) und Ergebnisse der Evaluationsforschung und der Schulentwicklungsforschung (z.B. Fitz-Gibbon 1996; Scheerens/Bosker 1997; Strittmatter 1999; Teddlie/Reynolds 2000) zeigen, dass aus der Diagnose von Arbeitsbedingungen und -resultaten keineswegs direkt Impulse zur Weiterentwicklung entstehen. Hierzu sind günstige institutionelle Voraussetzungen und externe Unterstützungen notwendig. Ebenso kann die sachliche und zeitliche Passung der Rückmeldung zum Programm und Tätigkeitsprofil einer Schule eine wichtige Rolle spielen. Unter ungünstigen Bedingungen können Schulrückmeldungen sogar negative Auswirkungen haben. Dies ist auch einer der Gründe, weshalb die Entscheidung über die Anforderung einer Schulrückmeldung bei der Schule liegt. Die Schule soll diese Entscheidung unter Berücksichtigung der internen Situation, ohne äußeren Druck treffen können.

Der Zusammenhang zwischen Systemmonitoring einerseits, schulbezogener Evaluation und Entwicklung andererseits ist gegenwärtig Gegenstand zahlreicher Diskussionen im deutschsprachigen Bereich (z.B. Arnold 2002; Klieme/Baumert/Schwippert 2000; Pekrun 2002; Rolff 2001; i. Vorb.; Terhart 2002). Aus LAU, Länderstudien wie MARKUS, QuaSUM und anderen Projekten liegen Modelle für Schulrückmeldungen (Lehmann/Peek 1997; Helmke/Jäger 2002; Lehmann u.a 2000) und erste Erfahrungen mit deren Rezeption vor (Peek 2002). Angesichts des Diskussionsstandes in der Schulentwicklungsforschung, der zu Skepsis Anlass gibt, sowie des Mangels an systematischen Rezeptionsuntersuchungen ist nachdrücklich darauf hinzuweisen, dass die Schulrückmeldungen nicht mit Erwartungen an die Initiierung von Schulentwicklungsprozessen belastet werden dürfen.

2.2 Belastbarkeit und Generalisierbarkeit der Informationen aus Schulrückmeldungen

Die Einzelschule ist Fokus und Handlungseinheit für die Qualitätsentwicklung im Bildungswesen. Im gewissen Sinne ist diese Aussage der „Fundamentalsatz der Schulentwicklungsforschung“ (Klieme/Baumert/Schwippert 2000, S. 387). Diese Vorstellung

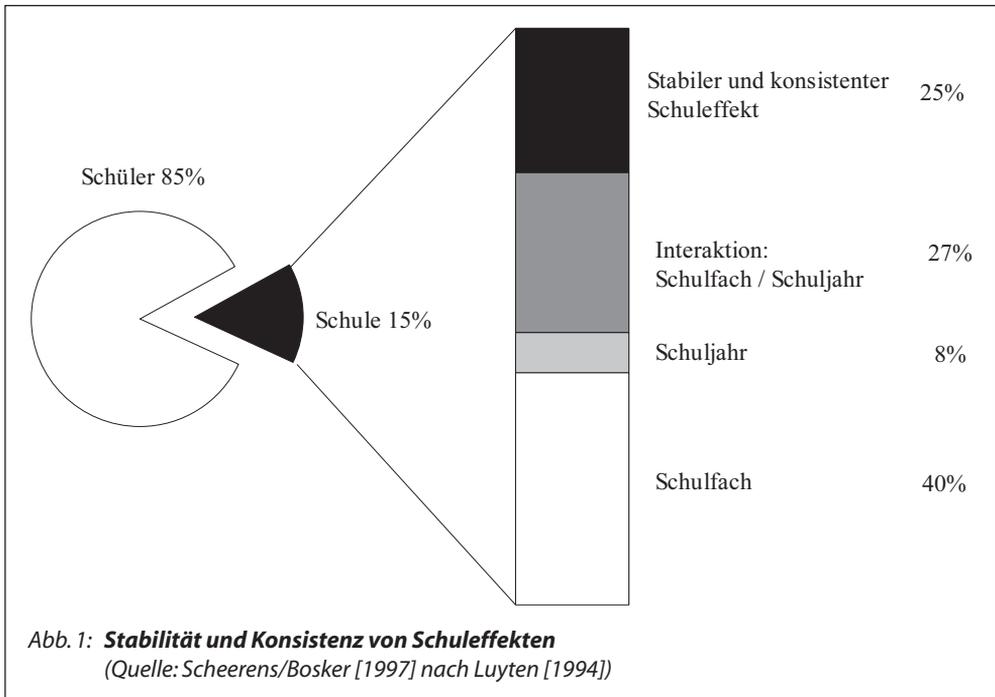
setzt das Konzept konsistenten Handelns auf Schulebene voraus, das förderliche organisatorische Rahmenbedingungen sichert und für die Entwicklung eines gemeinschaftlich geteilten Schulethos verantwortlich ist, der wiederum die Qualität in allen Bereichen einer Schule beeinflusst. Eine effektive Schule ist, wie Rutter u.a. (1980) formuliert haben, stabil und konsistent effektiv. Diese Überzeugung ist die Grundlage der meisten Vorschläge, die eine direkte Verbindung von *Large-Scale Assessment* und Schulentwicklung empfehlen oder hierin sogar eine zentrale Legitimation von Analysen auf Systemebene sehen. Wie Befunde der *School-Effectiveness-* und *School-Improvement-*Forschung zeigen (Sammons 1996; Scheerens/Bosker 1997; Sammons/Mortimore/Thomas 1996; Teddlie/Reynolds 2000), ist sie jedoch empirisch nicht haltbar. Deshalb wird im Rahmen von PISA eine vorsichtige Position vertreten, die sowohl die auf Grund des großen Stichprobenfehlers begrenzte Belastbarkeit von Schulrückmeldungen betont, als auch die relativierenden Befunde der Schuleffektivitätsforschung der letzten zwanzig Jahre berücksichtigt.

Sollen schulbezogene Befunde über die an der Untersuchung unmittelbar beteiligten Personen hinaus generalisiert werden, sei es – wie es in PISA nahe liegend ist – auf die Jahrgangsstufe und Altersgruppe oder gar die Schule insgesamt – wozu Ansätze der Schulentwicklung neigen –, sind die stichprobenbedingten Schätzfehler zu berücksichtigen, die in der Regel zu großen Vertrauensintervallen der jeweiligen Schätzungen führen. Mit diesem Problem kann man im Rahmen von Schulrückmeldungen adäquat umgehen, auch wenn wir über den Gebrauch entsprechender Informationen in den Schulen selbst wenig wissen. In Abschnitt 3 wird dargestellt, wie dieser Aspekt im Rahmen von PISA-2000 behandelt wurde.

Schwerwiegender sind sachliche, durch Resultate der Schuleffektivitätsforschung gestützte Bedenken gegen die Generalisierung von Ergebnissen über die untersuchte Domäne und den untersuchten Jahrgang hinaus (Ma 2001; Sammons/Mortimore/Thomas 1996; Scheerens/Bosker 1997). Sollen die Befunde als Indikatoren der Schulqualität genutzt werden, setzt dies Konsistenz der Ergebnisse über Fächer und Personengruppen sowie Stabilität über Geburtskohorten, den ‚Lebenszyklus‘ von Schulklassen und einen moderaten Personalwechsel hinweg voraus. Scheerens/Bosker (1997) haben die einschlägigen internationalen Befunde zusammengestellt, die eine solche Annahme für Sekundarschulen relativieren. Schuleffekte sind zu einem erheblichen Teil Effekte des Fachs bzw. Fachbereichs und der Geburtskohorte bzw. der Jahrgangsstufe. Die von Scheerens/Bosker (1997, S. 92) übernommene und adaptierte Abbildung, die auf Befunde von Luyten (1994) zurückgreift, gibt einen Eindruck vom Verhältnis der jeweiligen Einflusskomponenten (vgl. Abb. 1). ‚Echte‘ Schuleffekte sind nachweisbar, aber im Vergleich zum Effekt des Fachunterrichts von nachgeordneter Bedeutung.

Inwieweit die Befunde zur Konsistenz und Stabilität von Ergebnissen auch über unterschiedliche Schulsysteme bzw. Ländergrenzen hinweg verallgemeinerbar sind, ist weitgehend ungeklärt (vgl. Scheerens/Bosker 1997). Für Deutschland ist die Datenlage extrem unbefriedigend. Empirische Untersuchungen von Ditton/Krecker (1995) haben gezeigt, dass zwischen Klassen innerhalb einer Schule größere Unterschiede bestehen können als zwischen verschiedenen Schulen derselben Schulart eines Landes. Die Studie

ist jedoch nur begrenzt aussagefähig, da sie auf einer selektiven Konvenienzstichprobe von Schulen und auf Leistungsmaßen beruht, die nicht unterrichtsbezogen sind. Baumert/Köller (2000) haben im Rahmen einer Untersuchung zur multiplen Zielerreichung an gymnasialen Oberstufen gezeigt, dass Verallgemeinerungen über Domänen (Fächer, Lernbereiche) hinweg nur begrenzt oder gar nicht zulässig sind. Eigene Berechnungen zur Sekundarstufe I, die wir auf der Grundlage der Mehrkohorten-Längsschnittuntersuchung BIJU („Bildungsprozesse und psychosoziale Entwicklung im Jugendalter“), die mehrere Fächer berücksichtigt und für vier Bundesländer repräsentativ ist, durchgeführt haben, belegen, dass die zeitliche Stabilität sowohl des Leistungsstatus als auch der Leistungszuwächse *innerhalb* eines Fachs relativ hoch ist, und dass man auch von einer gewissen Leistungshomogenität innerhalb von Fachbereichen (z.B. dem mathematisch-naturwissenschaftlichen Fachbereich) sprechen kann. Dagegen ist Konsistenz der Leistungszuwächse über Fachbereiche hinweg (z.B. zwischen Mathematik und den Fremdsprachen) praktisch nicht zu erwarten. Die Korrelationen auf Schulebene liegen nahe bei Null.



Angesichts dieser Befundlage wird in den Schulrückmeldungen ausdrücklich darauf hingewiesen, dass die Befunde von PISA keine Generalisierung über die untersuchte Jahrgangsstufe sowie vor allem über das erfasste Fachgebiet hinaus erlauben und selbst Interpretationen auf der Ebene der Jahrgangsstufe sorgfältig die Stichproben zu berücksichtigen haben. Ferner wird betont, dass die schulbezogenen Ergebnisse keine direkten Rückschlüsse auf die Effizienz und Qualität der schulischen Arbeit zulassen – und zwar

auch dann nicht, wenn Vergleiche über Schulen mit ähnlicher Schülerschaft vorgenommen werden (*Value-added-Ansatz*). Denn die querschnittliche Anlage von PISA erlaubt keine hinreichende Kontrolle der unterschiedlichen Eingangsselektivität von Schulen. Dies gilt auch für alle anderen *Large-Scale Assessment*-Studien in Deutschland, mit Ausnahme der längsschnittlich angelegten Untersuchungen LAU und TIMSS/II-Deutschland. Schulrückmeldungen in PISA können also bestenfalls *Anregung für ein professionelles Gespräch* über Schule und Unterricht sein. Aber selbst diese Funktion der Schulrückmeldungen ist nicht voraussetzungslos. Erfahrungen mit Rückmeldungen im Rahmen von QuaSUM in Brandenburg zeigen, dass weder einfache Darstellungen noch explizite Vorsichtshinweise eine Garantie für eine adäquate Rezeption und Nutzung der Befunde bieten (Peek 2002). Dies spricht für die Notwendigkeit von Unterstützungsmaßnahmen bei der Dissemination von Ergebnissen an die PISA-Schulen.

2.3 *Prinzipien der Gestaltung von Schulrückmeldungen und ihre Umsetzung im Rahmen von PISA*

Auf der Grundlage des Diskussionsstands zur Funktion von Schulrückmeldungen können einige Gestaltungskriterien zusammengestellt werden, die nützlich oder wünschenswert sind. Während im Kontext von Schulentwicklung und -evaluation bislang vor allem Fragen der Funktion und Gestaltung von Feedback diskutiert wurden, sind im Systemmonitoring methodische Kriterien ausgearbeitet worden, die bedacht werden müssen, wenn die Rückmeldung zuverlässig und valide sein soll. Die folgende Zusammenstellung integriert beide Sichtweisen, soweit sie auf PISA anwendbar sind (zur weiteren Diskussion vgl. Arnold 1999; Klieme/Baumert/Schwippert 2000).

2.3.1 *Freiwilligkeit und Vertraulichkeit*

PISA ist eine Stichprobenstudie, die auf Freiwilligkeit der Teilnahme von Personen und deren Anonymität sowie auf Vertrauensschutz gegenüber den beteiligten Schulen setzt. Öffentlich wird über derartige Untersuchungen nur in anonymisierter bzw. aggregierter Form berichtet, sodass weder Personen noch einzelne Schulen identifiziert werden können. Nur die Schule selbst erhält Informationen über die eigenen aggregierten Daten, bei deren Interpretation sie gegebenenfalls nach eigenem Ermessen Beratungsangebote nutzen kann. Andere Regelungen können für flächendeckende Erhebungen in einem Land sinnvoll sein, sie sind jedoch aufgrund der begrenzten Aussagekraft in Bezug auf die Ergebnisse einzelner Schulen im Rahmen von PISA nicht angemessen.

2.3.2 *Verankerung der schulbezogenen Befunde*

Eine Rückmeldung an die Einzelschule, die lediglich absolute Werte mitteilt (beispielsweise mittlere Resultate in Leistungstests oder mittlere Kennwerte für Motivation und

Selbstkonzept der Schüler), ist für eine Schule wenig aussagekräftig. Solche Befunde sind nur anhand von Vergleichsmaßstäben oder Bezugsnormen sinnvoll interpretierbar. Dabei lassen sich drei Bezugsnormen unterscheiden:

- Vergleich mit früheren Leistungen derselben Schule (*ipsative oder individuelle Bezugsnorm*)
- Vergleich mit einer Normgruppe von Schulen (*soziale Bezugsnorm*)
- Vergleich mit sachlichen Standards (*kriteriale Bezugsnorm*)

Alle drei dieser Bezugsnormen können potenziell nützliche Informationen liefern und lassen sich prinzipiell auf Rückmeldungen schulbezogener Ergebnisse im Rahmen von Schulleistungstudien anwenden. Da die ipsative Bezugsnorm ein längsschnittliches Design voraussetzt, kommt diese Vergleichsperspektive in PISA nicht in Betracht. Stattdessen wurde eine Kombination sozialer und kriterialer Bezugsnormen gewählt.

2.3.3 Soziale Bezugsnorm: Berücksichtigung von Kontextbedingungen schulischer Arbeit

Die Auswahl sozialer Vergleichsgruppen ist eng mit dem Sachverhalt verknüpft, dass sich in Querschnittuntersuchungen Effekte der Eingangsselektivität nicht von Effekten der Schuleffektivität trennen lassen. Je weniger vergleichbar die Kontextbedingungen sind, unter denen Schulen arbeiten, desto weniger sicher kann man sein, dass gefundene Unterschiede auf differenzielle Schuleffektivität zurückzuführen sind. Es gehört daher zum *state of the art* vergleichender Schulforschung, in Befunddarstellungen Kontextmerkmale zu berücksichtigen, die mit den Ergebnissen zusammenhängen, von der Schule jedoch nicht oder kaum beeinflussbar sind. Methodisch wird so vorgegangen, dass innerhalb der Gesamtstichprobe die jeweilige abhängige Variable (z.B. fachliches Interesse oder Lesekompetenz) regressionsanalytisch auf Hintergrundmerkmale von Schülerinnen und Schülern zurückgeführt wird. Die Abweichung des beobachteten Scores vom regressionsanalytisch geschätzten Score eines Schülers (der sogenannte Residualwert) dient dann approximativ als Messwert für den durch die Schule erzielten *value-added*. Da diese Residualwerte jedoch inhaltlich schwer zu interpretieren sind, werden bei Ergebnismrückmeldungen häufig beobachtete und erwartete Werte nebeneinander mitgeteilt (vgl. Peek 2001a, b). Bei diesem Verfahren handelt es sich um eine Hilfskonstruktion, die in Querschnittuntersuchungen eingesetzt wird, weil keine Informationen über das Vorwissen der Schülerinnen und Schüler vorhanden sind. Durch Berücksichtigung von Hintergrundmerkmalen können diese lediglich approximiert werden.

Auch im Rahmen von PISA-2000 wird bei der Rückmeldung schulbezogener Ergebnisse nach dem *value-added* Ansatz verfahren. In den Darstellungen wird das Ergebnis der jeweiligen Schule ihrem regressionsanalytisch ermittelten Erwartungswert gegenüber gestellt. Um zu vermeiden, dass triviale Unterschiede überinterpretiert werden, haben wir den Vergleichswert in der Regel nur dann dargestellt, wenn das Ergebnis der

Schule bedeutsam davon abweicht (zur Definition von Bedeutsamkeit s.u.). Die Schätzung der Erwartungswerte basiert auf den Ergebnissen von Schulen derselben Schulform innerhalb des jeweiligen Bundeslandes. Lediglich im Fall von Schulen, für die innerhalb des jeweiligen Bundeslandes keine angemessene Vergleichsgruppe vorhanden ist (z.B. aufgrund geringer Beteiligungsquoten für die Schulform in diesem Land), werden Schulen derselben Schulform über alle Bundesländer hinweg einbezogen. Als Hintergrundmerkmale der Schülerinnen und Schüler gehen in die Schätzungen der Erwartungswerte die Variablen Geschlecht, Muttersprache (deutsch vs. nicht deutsch), Schul- und Berufsbildung der Eltern, sozioökonomische Stellung der Eltern, Ausstattung der Familie mit Wohlstandsgütern sowie Fähigkeit des Jugendlichen zum schlussfolgernden Denken, gemessen mit dem kognitiven Fähigkeitstest (Heller/Gaedike/Weinläder 1985), ein. Die Einbeziehung schlussfolgernden Denkens zur Schätzung von Erwartungswerten stellt insbesondere für die Beurteilung der Testleistungen ein konservatives Vorgehen dar. So haben Längsschnittstudien gezeigt, dass Schulen auch die kognitiven Grundfähigkeiten beeinflussen, die wiederum für die Entwicklung fachlicher Kompetenzen von Bedeutung sind (vgl. Köller/Baumert 2002). Bei einem Teil der Varianz, die hier kontrolliert wird, handelt es sich also um Unterschiede, die eigentlich dem Einfluss der Schule zuzurechnen sind. Die Abweichungen zwischen den Leistungswerten von Schulen und den Erwartungswerten dürften den Einfluss der Schule demnach tendenziell unterschätzen. Die Einbeziehung der kognitiven Grundfähigkeiten ist jedoch zur Approximation des Vorwissens, für das in PISA kein Maß verfügbar ist, notwendig.

2.3.4 *Kriteriale Bezugsnorm: Kompetenzstufen*

Für pädagogische Diskussionen ist eine Ergebnisdarstellung, die einen Kennwert inhaltlich interpretierbar macht, letztlich nutzbringender als die Einordnung im Vergleich mit einer Referenzgruppe (Watermann/Klieme 2002). Voraussetzung ist jedoch, dass die verwendete Skala (z.B. Motivationsskala, Kompetenzskala) inhaltlich bestimmbare Abstufungen besitzt. Für viele Skalen, wie zum Beispiel Selbstberichtskaalen zu motivationalen Merkmalen oder zur Wahrnehmung von Aspekten des Schulklimas, ist eine solche kriteriale Interpretation nur sehr bedingt möglich. Hier lässt sich anhand von Beispieltens und durch Kennzeichnung des Neutralitätspunktes lediglich ein Gefühl dafür vermitteln, welche Angaben die Schülerinnen und Schüler, die einen bestimmten Durchschnittswert erreicht haben, über ihr Erleben und Verhalten tendenziell gemacht haben. Für den Leistungsbereich sind jedoch Verfahren entwickelt worden, die eine inhaltliche Interpretation von Testwerten ermöglichen. Es gehört inzwischen zum Standard groß angelegter Schulleistungsstudien, mithilfe des *proficiency scaling* Verfahrens Kompetenzstufen zu definieren (Gonzales/Beaton 1994; Watermann/Klieme 2002; Klieme u.a. 2000). Damit lässt sich beschreiben, welche Anforderungen ein Schüler, der einen bestimmten Testwert erreicht hat, mit hinreichender Sicherheit bewältigen kann.

In PISA wurden in jeder der drei untersuchten Domänen fünf Kompetenzstufen definiert (vgl. Baumert u. a. 2001; OECD 2002). In den Rückmeldungen der Ergebnisse an

die PISA-Schulen werden die Kompetenzstufen zur Darstellung der Befunde im Bereich Lesen herangezogen¹. Die Verteilung der Jugendlichen auf die Stufen wird berichtet und der erwarteten Verteilung für Schulen mit entsprechender Zusammensetzung der Schülerschaft gegenüber gestellt (s.u.), wobei die Vergleichswerte nur dann berichtet werden, wenn der Mittelwert der Schule vom Erwartungswert bedeutsam abweicht (vgl. Abschnitt 4). Mit dieser Darstellungsform werden normorientierte und kriteriumsorientierte Vergleichsmaßstäbe kombiniert, und es wird den Schulen ermöglicht, die erreichten Ergebnisse inhaltlich zu interpretieren. Welche Verteilung auf die Stufen der Skala als zufriedenstellend gelten kann, bleibt dabei der Diskussion an der Einzelschule überlassen. Weder in den Lehrplänen der Sekundarstufe I für das Fach Deutsch noch in den von den Kultusministern der Länder formulierten Standards für den mittleren Abschluss werden explizit Kompetenzanforderungen für den Bereich Lesen definiert. Aus den darin beschriebenen Zielsetzungen des Deutschunterrichts und Anforderungen, die Jugendliche in der neunten Klassenstufe bewältigen sollten, kann lediglich abgeleitet werden, dass das Erreichen der Kompetenzstufe II als allgemeiner Mindeststandard zu bewerten ist (Artelt u.a. 2001). Inwieweit eine Schule über das Erreichen dieser Mindeststandards hinausgehende Ziele anstrebt, bedarf der internen Klärung.

2.3.5 Beachtung von Grenzen der Aussagefähigkeit der Daten

Befunde aus sozialwissenschaftlichen Stichprobenuntersuchungen sind grundsätzlich mit Schätzfehlern behaftet. Aufgrund der relativ kleinen Stichproben innerhalb von Schulen sind diese für die schulbezogenen Ergebnisse in PISA vergleichsweise groß. Weiterhin kommt hinzu, dass jeweils nur eine Kohorte einer bestimmten Jahrgangs- oder Altersstufe untersucht wird. Da die Probleme der Verallgemeinerbarkeit solcher Daten nicht allgemein vertraut sind, wurde besonderer Wert darauf gelegt, die Schulrückmeldung so zu gestalten, dass Über- und Fehlinterpretationen möglichst vermieden werden. Unter anderem wird in den Ergebnisdarstellungen gekennzeichnet, welche Abweichungen zwischen beobachteten und erwarteten Werten als bedeutsam gelten können.

Üblicherweise wird in sozialwissenschaftlichen Untersuchungen zur Bewertung von Differenzen zwischen Messwerten das Kriterium der statistischen Signifikanz herangezogen. Dabei wird in der Regel ein strenges Maß angelegt, um die Hypothese über das Vorliegen eines Unterschieds zu prüfen. Dieses Maß minimiert die Wahrscheinlichkeit des Fehlers erster Art, also einen Unterschied anzunehmen, der möglicherweise nicht vorliegt. Dieses Vorgehen ist für die Schulrückmeldungen in PISA nicht zielführend, da auf Grund der geringen Stichprobengrößen an Einzelschulen bei Annahme der üblichen Signifikanzniveaus von 1-10 Prozent selbst große Differenzen nicht gegen den

1 In den Bereichen Mathematik und Naturwissenschaften ist die Anzahl der Schülerinnen und Schüler, die internationale Aufgaben bearbeitet haben, innerhalb der Schulen zu klein, um Verteilungen auf Kompetenzstufen darstellen zu können.

Zufall abgesichert werden können. Die Wahrscheinlichkeit, einen Fehler zweiter Art zu begehen, und einer Schule mitzuteilen, die Testleistungen ihrer Schülerinnen und Schüler lägen im Rahmen des zu erwartenden Leistungsstandes, obwohl dies möglicherweise nicht der Fall ist, wäre also bei Verwendung des Kriteriums der statistischen Signifikanz relativ groß.

In Abwägung der zwei möglichen Fehler haben wir im Rahmen von PISA folgendes Vorgehen gewählt (vgl. hierzu auch Watermann/Stanat, in Vorb.):

- Zur Bewertung der Bedeutsamkeit von Unterschieden zwischen den Werten einer Schule und dem jeweiligen Erwartungswert werden Maße der praktischen Relevanz verwendet. Im Bereich der Testleistungen werden als Kriterium Kompetenzzuwächse herangezogen, die Schülerinnen und Schüler üblicherweise im Laufe eines Jahres erzielen. Anhand von Daten aus PISA, TIMSS und anderen Studien lässt sich schätzen, dass in einem Jahr je nach Fachgebiet Leistungsgewinne von etwa 25 bis 50 Punkten (PISA-Metrik) erreicht werden (Baumert u.a. 2002; Baumert u.a. 1997). Vor diesem Hintergrund werden Abweichungen des Leistungsmittelwerts einer Schule vom Erwartungswert als bedeutsam gekennzeichnet, wenn sie 20 Punkte oder mehr betragen, also einem Lernzuwachs von etwa einem halben Jahr entsprechen.
- Für andere Bereiche, wie beispielsweise motivationale Merkmale, steht kein ähnlich anschauliches Maß für die Bedeutung von Unterschieden zur Verfügung. Hier ist es sinnvoll, mit der Einheit der Standardabweichung zu arbeiten und Unterschiede zwischen dem Wert der Schule und dem Erwartungswert, die mindestens eine Drittel Standardabweichung betragen, als bedeutsam auszuweisen.
- Die Rückmeldung von Merkmalen des familiären Hintergrunds der Schülerinnen und Schüler schließlich erfolgt auf der Grundlage eines Vergleichs mit anderen Schulen derselben Schulform innerhalb eines Bundeslandes (bzw. in Ausnahmen über alle Bundesländer hinweg). Hier wird den Schulen in Form von Balkendiagrammen mitgeteilt, in welchem Bereich sich der Wert der Schule mit einer 90-prozentigen Wahrscheinlichkeit befindet und wie weit dieser vom entsprechenden Durchschnittswert der Schulform entfernt ist (vgl. auch Abschnitt 3).

3. Gestaltung der Schulrückmeldungen im Rahmen von PISA-2000

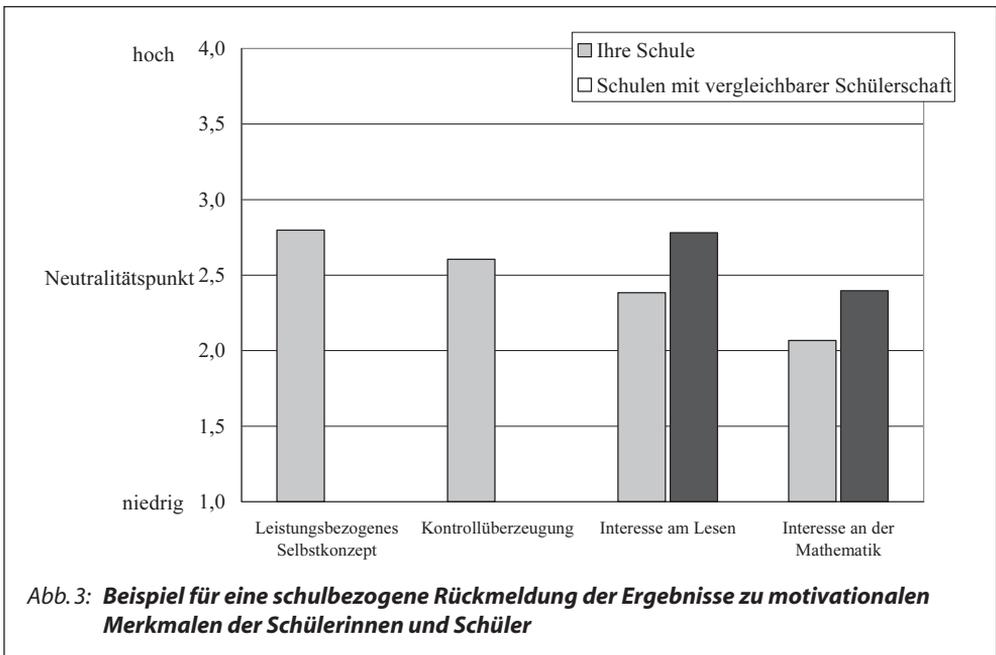
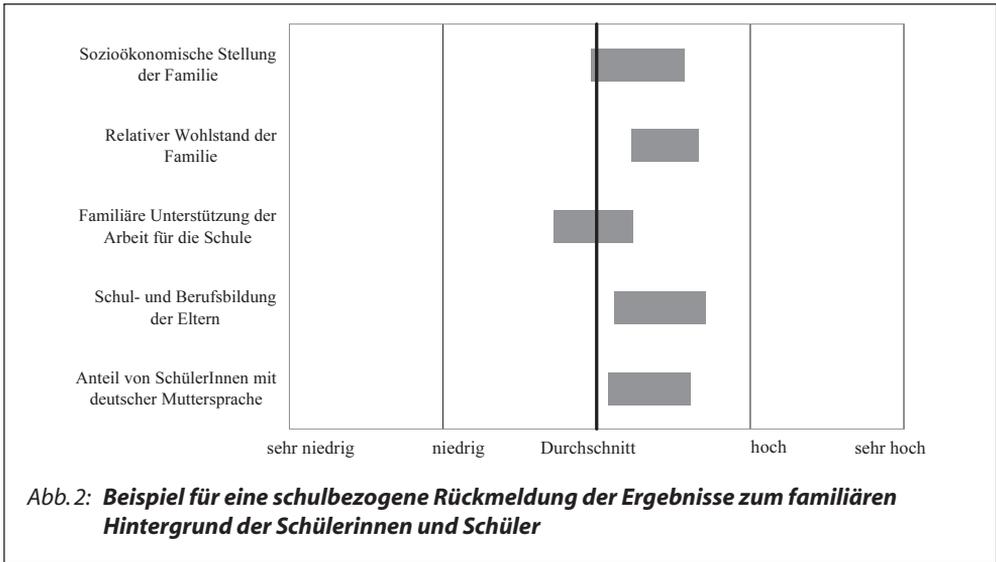
Mit den Schulrückmeldungen im Rahmen von PISA-2000 sind zwei allgemeine Funktionen verknüpft. Sie sollen zunächst einmal dazu genutzt werden, die beteiligten Schulen über die Studie zu informieren (Disseminationsfunktion). Der erste Teil der Rückmeldung enthält daher eine etwa 30 Seiten umfassende Beschreibung der Ziele, Methoden und Ergebnisse der Untersuchung sowie eine umfassende Liste von Quellen weiterführender und vertiefender Informationen (vgl. Stanat u.a. 2002). Es wird an verschiedenen Stellen betont, dass der Ertrag von PISA vor allem in allgemeinen Erkenntnissen über die Bedingungen und Ergebnisse schulischer Lehr-/Lernprozesse, über relative Stärken und Schwächen von Schülerinnen und Schülern in Deutschland und über mögliche An-

satzpunkte für Verbesserungen liegt. Als Ergänzung dieses Interpretationsrahmens für die schulbezogenen Befunde werden im zweiten Teil der Rückmeldung die in PISA definierten Kompetenzstufen für die Bereiche Lesen, Mathematik und Naturwissenschaften beschrieben.

Vor dem Hintergrund der Grundlagen und allgemeinen Befunde der Studie sollen die Rückmeldungen den Schulen weiterhin Anhaltspunkte für ihre eigene Situation liefern (Rückmeldefunktion). Wir haben versucht, diesen Teil so zu gestalten, dass Fehlinterpretationen so weit wie möglich vorgebeugt wird. Dabei galt es zum einen, sowohl den potenziellen Nutzen als auch die Grenzen der schulbezogenen Informationen in ausgewogener Form zu erläutern. Zum anderen wurde angestrebt, eine angemessene Balance zwischen Verständlichkeit und Differenziertheit der Ergebnisdarstellungen zu erreichen. Zur Unterstützung bei der Interpretation der schulbezogenen Informationen werden in einer Einleitung die einzelnen Grafiken anhand eines fiktiven Beispiels erläutert. Die Gestaltung des schulbezogenen Teils der Rückmeldung orientiert sich an Modellen aus dem In- und Ausland (LAU; QuaSUM; MARKUS; Illinois State Board of Education 1993). Entwürfe wurden von einer Gruppe von Schulleiterinnen und Schulleitern sowie von den Moderatorinnen und Moderatoren der Länder (s.u.) im Hinblick auf Verständlichkeit geprüft und auf der Grundlage ihrer Hinweise anschließend optimiert.

Der schulbezogene Teil der PISA-Schulrückmeldungen enthält fünf Abschnitte (für eine ausführliche Beschreibung der einzelnen Skalen vgl. Kunter u.a. 2002):

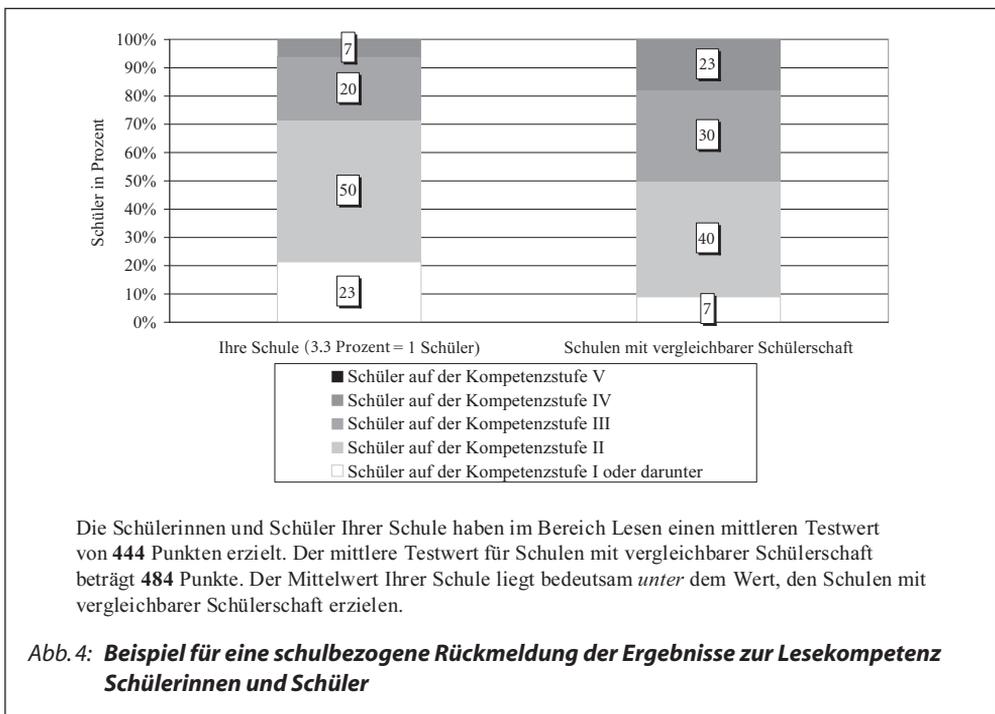
- 1) Abschnitt 1 stellt Rahmendaten zur Verfügung, die für die spezifische Aussagekraft der Untersuchung in der jeweiligen Schule von Bedeutung sind. Es wird dargestellt, wie viele Schüler in der Schule für die Teilnahme an der PISA-Erhebung in Betracht kamen, wie viele in die Stichprobe gezogen wurden, wie hoch die Beteiligungsquote war² und welcher Anteil der Testgruppe Mädchen waren. Weiterhin wird in Form eines Balkendiagramms im ersten Abschnitt dargestellt, in welchem Maße die Schülerinnen und Schüler bereit waren, sich bei der Bearbeitung des PISA-Tests anzustrengen. Der Erwartungswert für die Anstrengungsbereitschaft wird in das Diagramm nur dann eingetragen, wenn er bedeutsam über oder unter dem beobachteten Wert der Schule liegt.
 - 2) In Abschnitt 2 werden Merkmale des familiären Hintergrunds der Schülerinnen und Schüler der Einzelschule dargestellt und mit den entsprechenden Werten für Schulen derselben Schulform im jeweiligen Bundesland (bzw. über alle Bundesländer) verglichen (vgl. Abb. 2). Eine Abweichung wird als bedeutsam interpretiert, wenn das 90-Prozent-Konfidenzintervall um den Mittelwert der Schule den Durchschnittswert für die Schulform nicht einschließt (vgl. Balken in Abb. 2).
- 2 Schulen erhalten im Rahmen von PISA-2000 nur dann eine Rückmeldung schulbezogener Befunde, wenn mindestens 50 Prozent und mindestens 10 der in die Stichprobe gezogenen Neuntklässler an der Erhebung teilgenommen haben.



3) Um der motivationalen Zieldimension von Schule und Unterricht Rechnung zu tragen, werden in Abschnitt 3 Befunde für generalisierte schulleistungsbezogene Fähigkeitseinschätzungen (akademisches Selbstkonzept und Kontrollüberzeugung) sowie fachbezogene Interessen (Interesse am Lesen und Interesse an Mathematik) berich-

tet (vgl. Abb. 3). Weiterhin enthält Abschnitt 3 Informationen darüber, wie häufig die Schülerinnen und Schüler täglich zum Vergnügen lesen (ohne Abbildung). Auch in diesen Grafiken wird auf die Darstellung von Erwartungswerten verzichtet, wenn der beobachtete Wert nicht bedeutsam vom Erwartungswert abweicht.

- 4) Abschnitt 4 widmet sich der Darstellung der Testleistungen (vgl. Abb. 4). Für jeden der drei Kompetenzbereiche wird berichtet, welchen Mittelwert die Schülerinnen und Schüler in der jeweiligen Schule erreicht haben. Um zu vermeiden, dass Schulen zur Einordnung ihres Ergebnisses die unadjustierten Mittelwerte aus den allgemeinen Ergebnisberichten heranziehen, wenn in den Schulrückmeldungen kein angemessener Vergleichsmaßstab zur Verfügung gestellt wird, werden in diesem Abschnitt die Erwartungswerte grundsätzlich berichtet. Es wird jedoch deutlich gekennzeichnet, ob die Abweichung des beobachteten vom erwarteten Ergebnis als bedeutsam zu bewerten ist (Abweichung von 20 Punkten oder mehr). Im Bereich Lesen wird zusätzlich die Verteilung der Jugendlichen auf die Kompetenzstufen dargestellt. Ein Vergleich mit der erwarteten Verteilung wird jedoch nur dann durch die Darstellung des entsprechenden Balkens ermöglicht, wenn der mittlere Testwert der Schule bedeutsam vom Erwartungswert abweicht.



Im dargestellten Beispiel erreicht die Schule 444 Punkte, womit sie unter dem Wert für Schulen mit vergleichbarer Schülerschaft liegt. Dieser Befund korrespondiert mit dem Ergebnis, dass in dieser Schule auch das Interesse im Lesen geringer ausgeprägt

ist als in der Vergleichsgruppe (vgl. Abb. 3). Es wäre denkbar, dass die Schule dies zum Anlass nimmt, in eine Diskussion über die Lesemotivation der Schülerinnen und Schüler einzutreten. Die Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen weist darauf hin, dass der Anteil der Jugendlichen, die Kompetenzstufe II nicht erreichen und somit im Hinblick auf selbstständiges Weiterlernen und auf einen erfolgreichen Übergang ins Berufsleben als potenzielle Risikokandidaten einzustufen sind, in der betreffenden Schule vergleichsweise groß ist (23 Prozent vs. 7 Prozent in der Vergleichsgruppe). Im höheren Leistungsbereich (Kompetenzstufe IV: Detailliertes Verständnis komplexer Texte) befindet sich hingegen ein geringerer Anteil von Schülern als erwartet. Dieses Muster könnte als Hinweis darauf gewertet werden, dass in der betreffenden Schule die Förderung im unteren und oberen Leistungsbereich weniger gut gelingt als in vergleichbaren Schulen.

Diese Interpretation ist allerdings mit einer gewissen Vorsicht zu bewerten. Sie wurde dargestellt, um das Potenzial der in PISA gewählten Darstellungsform deutlich zu machen. Bei ausreichender Stichprobengröße innerhalb der Schule kann diese Art der Rückmeldung detaillierte Hinweise auf relative Stärken und Schwächen geben. Angesichts der vergleichsweise kleinen Anzahl von Schülerinnen und Schülern, die im Rahmen von PISA innerhalb der einzelnen Schulen getestet worden sind, sollten Abweichungen von der erwarteten Verteilung auf die Kompetenzstufen jedoch nicht überbewertet werden. Dies wird in den Schulrückmeldungen ausdrücklich betont.

- 5) Abschnitt 5 schließlich enthält Informationen zur Wahrnehmung der Schule durch die Schülerinnen und Schüler. Die Skalen entstammen der Schulklima-Forschung und beziehen sich auf die allgemeine Schulzufriedenheit der Jugendlichen, auf die Schüler-Lehrer-Beziehungen und auf die Schüler-Schüler-Beziehungen. Weiterhin wird die allgemeine Zufriedenheit der Eltern mit der Schule dargestellt. Die Darstellungsform und das Kriterium für die Bedeutsamkeit von Differenzen zwischen beobachtetem und erwartetem Wert entsprechen denen in Abschnitt 3.

4. Unterstützung des Disseminationsprozesses in PISA-2000

Der Disseminationsprozess im Rahmen von PISA umfasst eine Vielzahl von Publikationen und Aktivitäten, die dazu dienen, die Akteure im Bildungssystem über die Studie und ihre Ergebnisse zu informieren und in eine Diskussion über die Befunde, ihre Interpretation und die abzuleitenden Schlussfolgerungen einzutreten. Dieser Prozess hat lange vor Veröffentlichung der ersten Ergebnisberichte begonnen, indem unter anderem gemeinsame Veranstaltungen der Länder und des Konsortiums für Schulleiter der ausgewählten Schulen durchgeführt und kontinuierlich Informationen über die Konzeption und den Arbeitsstand des Projekts auf den Internetseiten des nationalen PISA-Konsortiums veröffentlicht wurden. Auch über die Ergebnisse der Studie wurde umfassend informiert. Den Schulrückmeldungen kommt in diesem Zusammenhang eine besondere Bedeutung zu, da sie eine relativ große Zahl von Schulen erreichen und diese in die Nutzung der Ergebnisse unmittelbar einbeziehen. Dies stellt auch eine Gelegen-

heit für Schulen dar, Erfahrungen im Umgang mit Ergebnissen empirischer Untersuchungen zu sammeln.

Um den Prozess der Rezeption und Nutzung der Schulrückmeldungen zu unterstützen, wurden von den Ländern der Bundesrepublik ausgewählte Moderatorinnen und Moderatoren darin geschult, allgemeine und schulbezogene Ergebnisse der Studie zu erläutern. Je nach Bundesland wurden für diese Aufgabe unterschiedliche Personengruppen herangezogen, wie beispielsweise Schulpsychologen oder Schulentwicklungsmoderatoren. Die Schulung umfasste drei Teile. Im ersten Teil wurden anhand eines Modellvortrags und Gruppenarbeit zu ausgewählten Themen die Grundlagen, Methoden und allgemeinen Ergebnisse der PISA-Studie dargestellt und vertiefend diskutiert. Als Hilfsmittel für eigene Präsentationen in Schulen erhielten die Moderatorinnen und Moderatoren CDs mit PowerPoint-Versionen aller in den nationalen PISA-Berichten enthaltenen Tabellen und Grafiken. Der zweite Teil der Schulung diente dazu, Grundlagen für die Interpretation der schulbezogenen Rückmeldungen zu vermitteln und die Grenzen ihrer Aussagekraft zu verdeutlichen. Dabei wurden insbesondere Fragen der Stichprobenziehung sowie der statistischen und praktischen Bedeutsamkeit erläutert und die hieraus folgenden Konsequenzen für die Generalisierbarkeit der Befunde diskutiert. Die Anwendung der so erarbeiteten Interpretationsgrundlagen auf Rückmeldungen schulbezogener Ergebnisse wurde anschließend in Gruppenarbeit anhand von Beispielen erprobt.

Im dritten Teil der Schulung, der von den Bundesländern konzipiert und durchgeführt wurde, ging es schließlich darum, die Schulrückmeldungen und die Rolle der Moderatoren in den Rahmen des allgemeinen Disseminationskonzepts des jeweiligen Landes einzubetten. Dabei wurde in der Regel auch erläutert, welche weiteren Ressourcen zur Verfügung stehen, um Schulen bei Anschlussaktivitäten, etwa bei der Erhebung weiterer Informationen über ihre Situation, zu unterstützen. Es ist geplant, mit einer Nachbefragung unter anderem zu erheben, wie häufig Schulen Hilfe von Moderatorinnen und Moderatoren eingeholt haben und auf welche Art von Unterstützung sich die Anfragen bezogen.

5. Möglichkeiten der Nutzung von Systemmonitoring für die pädagogische Praxis

Wir haben in diesem Beitrag mehrfach betont, die Schulrückmeldungen in PISA-2000 nicht mit zu hohen Erwartungen zu belasten. Die schulbezogenen Rückmeldungen stellen nur eine Momentaufnahme für einen Jahrgang innerhalb einer Schule dar und können – was die pädagogische Praxis betrifft – bestenfalls Anlass für Gespräche über Optimierungsprozesse sein. Sie können keinen direkten Weg der Verbesserung vorzeichnen; es gibt keinen ‚geheimen‘ Mechanismus, der aus der Diagnose Entwicklungsdynamik entstehen ließe. Zumindest liegen bis heute keine überzeugenden Argumente für die Annahme vor, Schulrückmeldungen würden automatisch zu Innovations- und Optimierungsprozessen an Schulen führen. Alles spricht vielmehr für den Vorrang von

Entwicklungsmaßnahmen. Mit Recht wird deshalb – nicht nur in Deutschland – die Arbeit an Schulprogrammen in den Vordergrund gestellt. Evaluationsmaßnahmen können an Entwicklungen anschließen – primär in schulspezifischer Form, gelegentlich aber auch unter Nutzung von externen *Benchmarks*. Vor diesem Hintergrund möchten wir abschließend thematisieren, wie man Systemmonitoring für die Evaluation von Schulprogrammen nutzen kann.

Eine mittlerweile gut erprobte Möglichkeit der Nutzung von Systemmonitoring für die pädagogische Praxis besteht in der Verknüpfung von entsprechenden Erhebungen und zeitlich versetzter schulbezogener Evaluation, wie sie in verschiedenen Projekten des Max-Planck-Instituts für Bildungsforschung an Reformschulen durchgeführt wurden (vgl. Klieme/Baumert/Schwippert 2000; Köller/Trautwein, i. Vorb.). Ein wesentliches Kennzeichen dieser Untersuchungen ist es, dass dabei Tests und Fragebögen eingesetzt werden, die einer jüngeren Erhebung zum Systemmonitoring entstammen. Mit diesem Vorgehen sind zwei Ziele verbunden: Zum einen wird gewährleistet, dass nach wissenschaftlichen Kriterien entwickelte, empirisch erprobte und bewährte Erhebungs- und Auswertungsverfahren verwendet werden. Zum Zweiten sind aus der repräsentativen Erhebung Vergleichswerte bekannt, die als soziale und kriteriale Bezugsnormen für die Evaluation dienen können. So ist es möglich, auf der Basis von PISA annäherungsweise repräsentative Normen zum Beispiel für Schulformen innerhalb von Bundesländern zu bestimmen, an denen sich die Einzelschule messen kann. Es steht also beinahe im Sinne der standardisierten pädagogischen Diagnostik ein Instrumentarium zur Verfügung, das an einer nationalen Stichprobe geeicht ist und den Vergleich einer Einzelschule mit den vorhandenen Normen ermöglicht. Zusätzlich können sich Schulen anhand der Verteilung ihrer Schülerinnen und Schüler auf die Kompetenzstufen kriterial verorten. So wäre es denkbar, dass sich eine Schule in erster Linie das Erreichen von Mindeststandards zum Ziel gesetzt hat und weniger an einem sozialen Vergleich orientiert ist.

Jenseits des Leistungsbereichs ist weiterhin von Bedeutung, dass in PISA, wie auch in anderen Schulleistungsuntersuchungen, ein breites Spektrum motivationaler und sozialer Merkmale erhoben worden sind. Dies ermöglicht zahlreiche Verknüpfungen mit den Fragestellungen der Schulen, etwa zum Aspekt der Kopplung von motivationalen und fachbezogenen Zielen. Dabei gilt es jedoch zu beachten, dass das Potenzial der Nutzung des Systemmonitorings für Evaluationsvorhaben an Einzelschulen von der Passung der Ziele und der zu erfassenden Konstrukte abhängt.

Die hier beschriebene Strategie der Verknüpfung von Systemmonitoring und Evaluation von Einzelschulen setzt voraus, dass innerhalb von Schulen Schulentwicklungsprozesse bereits in Gang gesetzt worden sind. Außerdem muss gewährleistet sein, dass mithilfe externer Unterstützung Standards der empirischen Forschungsmethodik eingehalten werden (vgl. Leutner 1999). Nur unter diesen Bedingungen bietet Systemmonitoring einen Referenzrahmen für die Schule, den interne Selbstevaluation allein nicht bereitstellen kann (Klieme/Baumert/Schwippert 2000).

Literatur

- Arnold, K.-H. (1999): Fairness bei Schulsystemvergleichen. Münster: Waxmann.
- Arnold, K.-H. (2002): Schulentwicklung durch Rückmeldung der Lernwirksamkeit an die Einzelschule: Möglichkeiten und Grenzen der Schuleffizienzforschung. In: Zeitschrift für Pädagogik 48, S. 741-764.
- Artelt, C./Stanat, P./Schneider, W./Schiefele, U. (2001): Lesekompetenz: Testkonzeption und Ergebnisse. In: Baumert, J./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, W./Stanat, P./Tillmann, K.-J./Weiß, M. (Hrsg.): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich, S. 69-137.
- Baumert, J./Köller, O. (2000): Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In: Baumert, J./Bos, W./Lehmann, R. (Hrsg.): TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie: Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Band 2. Opladen: Leske + Budrich, S. 271-315.
- Baumert, J./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, W./Stanat, P./Tillmann, K.-J./Weiß, M. (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- Baumert, J./Artelt, C./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, Tillmann, K.-J./Weiß, M. (Hrsg.) (2000): PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. Opladen: Leske + Budrich.
- Baumert, J./Lehmann, R./Lehrke, M./Schmitz, B./Clausen, M./Hosenfeld, I./Köller, O./Neubrand, J. (1997): TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Opladen: Leske + Budrich.
- Ditton, H./Krecker, L. (1995): Qualität von Schule und Unterricht. Zeitschrift für Pädagogik 41, S. 507-529.
- Fitz-Gibbon, C. (1996): Monitoring school effectiveness: Simplicity and complexity. In: Gray, J./Reynolds, D./Fitz-Gibbon, C./Jeseson, D. (Hrsg.): Merging Traditions. The Future of Research on School Effectiveness and School Improvement. London: Cassell, S. 74-90.
- Gonzales, E.J./Beaton, A.E. (1994): The determination of cut scores for standards. In: Tuijnman, A.C./Postlethwaithe, T.N. (Hrsg.): Monitoring the standards of education. Oxford: Pergamon, S. 171-190.
- Heller, K. A./Gaedike, A.-K./Weinläder, H. (1985). Kognitiver Fähigkeits-Test KFT 4-13. Weinheim: Beltz.
- Helmke, A./Jäger, R.S. (Hrsg.) (2002): Das Projekt MARKUS: Gesamterhebung Rheinland-Pfalz. Kompetenzen, Unterrichtsmerkmale, Schulkontext. Landau: Verlag Empirische Pädagogik.
- Illinois State Board of Education (1993): Understanding your Child's IGAP Scores. Illinois.
- Klieme, E./Baumert, E./Schwippert, K. (2000): Schulbezogene Evaluation und Schulleistungsvergleiche – Eine Studie im Anschluss an TIMSS. In: Rolff, H.-G./Bos, W./Klemm, K./Pfeiffer, H./Schulz-Zander, R. (Hrsg.): Jahrbuch der Schulentwicklung. Band 11. Daten, Beispiele und Perspektiven. Weinheim und München, S. 387-419.
- Köller, O./Baumert, J. (2002): Entwicklung schulischer Leistungen. In: Oerter, R./Montada, L. (Hrsg.): Entwicklungspsychologie. 5., vollständig überarbeitete Auflage. Weinheim/Basel/Berlin: Beltz, Psychologie Verlags Union.
- Köller, O./Trautwein, U. (2001): Evaluation mit TIMSS-Instrumenten: Untersuchungen in der 8. Jahrgangsstufe an fünf Gesamtschulen. In: Die Deutsche Schule 93, S. 167-185.
- Köller, O./Trautwein, U. (in Vorb.): Schulqualität und Schülerleistung. Evaluationsstudie über innovative Schulentwicklungsprozesse an fünf hessischen Gesamtschulen. Weinheim und München: Juventa.
- Klieme, E./Baumert, J./Köller, O./Bos, W. (2000): Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompeten-

- zen. In: Baumert, J./Bos, W./Lehmann, R. (Hrsg.): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie: Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn (Bd. 1): Opladen: Leske + Budrich, S. 85-133.
- Kunter, M./Schümer, G./Artelt, C./Baumert, J./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, W./Stanat, P./Tillmann, K.-J./Weiß, M. (2002): PISA 2000. Dokumentation der Erhebungsinstrumente. Berlin: Max-Planck-Institut für Bildungsforschung.
- Lehmann, R./Peek, R. (1997): Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Hamburg: Behörde für Schule, Jugend und Berufsbildung.
- Lehmann, R./Peek, R./Gänsfuß, R./Lutkat, S./Mücke, S./Barth, I. (2000): Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg. Potsdam: Eigendruck MBSJ.
- Leutner, D. (1999): Standards empirischer Forschung für Evaluation im Bildungsbereich. In: Thonhauser, J./Patry, J.-L. (Hrsg.): Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog. Innsbruck: StudienVerlag.
- Luyten, J.W. (1994): School effects: Stability and Malleability. Enschede: University of Twente.
- Ma, X. (2001): Stability of School Academic Performance Across Subject Areas. In: Journal of Educational Measurement 38, pp. 1-18.
- OECD (2001): Knowledge and skills for life. First results from the OECD Programme for International Student Assessment (PISA) 2000. Paris: OECD.
- Peek, R. (2001a): Rückmeldestrategien als Element der Qualitätsentwicklung durch großflächige Tests. In: Journal für Schulentwicklung 2, S. 55-64.
- Peek, R. (2001b): Die Bedeutung vergleichender Schulleistungsmessungen für die Qualitätskontrolle und Qualitätsentwicklung. In Weinert, F.E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim: Beltz, S. 323-335.
- Peek, R. (2002). Klassenbezogene Rückmeldungen aus Schulleistungsstudien und ihre Folgen in den beteiligten Schulen – das Beispiel QuaSUM in Brandenburg. Vortrag auf dem 18. Kongress der DGFE, München.
- Pekrun, R. (2002): Vergleichende Evaluationsstudien zu Schülerleistungen: Konsequenzen für die Bildungsforschung. In: Zeitschrift für Pädagogik 48, S. 111-128.
- Rolff, H.-G. (2001): Was bringt die vergleichende Leistungsmessung für die pädagogische Arbeit in Schulen? In Weinert, F.E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim: Beltz, S. 337-352.
- Rolff, H.-G. (in Vorb.): Rückmeldung und Nutzung der Ergebnisse von großflächigen Leistungsuntersuchungen. Grenzen und Chancen. In Rolff, H.-G. (Hrsg.): Jahrbuch der Schulentwicklung, Band 12, Weinheim: Juventa.
- Rutter, M./Maughan, B./Mortimore, P./Ouston, J. (1979): Fifteen Thousand Hours. Secondary Schools and their Effects on Children. Somerset: Open Books.
- Sammons, P. (1999): School Effectiveness. Lisse: Swets + Zeitlinger.
- Sammons, P. (1996): Complexities in the Judgement of School Effectiveness. In: Educational Research and Evaluation 2, S. 113-149.
- Sammons, P./Mortimore, P./Thomas, S. (1996): Do Schools Perform Consistently Across Outcomes and Areas. In: Gray, J./Reynolds, D./Fitz-Gibbon, C./Jesson, D. (Hrsg.): Merging Traditions. The Future of Research on School Effectiveness and School Improvement. London: Cassel, S 3-29.
- Scheerens, J./Bosker, R.J. (1997): The Foundations of Educational Effectiveness. Oxford: Pergamon.
- Stanat, P./Artelt, C./Baumert, J./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, W./Schümer, G./Tillmann, K.-J./Weiß, M. (2002): PISA 2000: Die Studie im Überblick. Berlin: Max-Planck-Institut für Bildungsforschung.

- Strittmatter, A. (1999): Qualitätsevaluation und Schulentwicklung. In: Thonhauser, J./Patry, J.-L. (Hrsg.): Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog. Innsbruck: StudienVerlag.
- Teddlie, C./Reynolds, D. (2000). The International Handbook of School Effectiveness Research. London, NY: Falmer.
- Terhart, E. (2002): Wie können die Ergebnisse von vergleichenden Leistungsstudien systematisch zur Qualitätsverbesserung in Schulen genutzt werden? In: Zeitschrift für Pädagogik 48, S. 91-110.
- Watermann, R./Klieme, E. (in press): Reporting Results of Large-Scale Assessment in Psychologically and Educationally Meaningful Terms - Construct Validation and Proficiency Scaling in TIMSS. In: International Journal of Psychological Assessment 18, (3).
- Watermann, R./Stanat, P. (in Vorb.): Methodische Aspekte der Schulrückmeldungen im Rahmen von PISA. In: Empirische Pädagogik.

Abstract: Providing schools with feedback on their results in assessment studies is often expected to trigger school-development processes. The fact that the usefulness of such feedback depends on the design of the study and that interpreting the results may be difficult for schools is frequently ignored. The current article describes a realistic position with regard to the use of school-level feedback. It presents a study-specific approach to providing schools with such information that was developed for PISA-2000 as part of a broader dissemination strategy. This approach takes into account research findings and implementation experiences pertaining to the potential and limitations of school-level feedback.

Anschrift der Autoren:

Dr. Rainer Watermann, Dr. Petra Stanat, Dipl. Psych. Mareike Kunter, Prof. Dr. Jürgen Baumert, Max-Planck-Institut für Bildungsforschung, Lentzeallee 94, 14195 Berlin.
Prof. Dr. Eckhard Klieme; Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt.