

An illustration of an open book with its pages fanned out, positioned on the left side of the page. The book is rendered in shades of purple and grey, with a white spine.

PISA

Aktuelle GEW-Informationen aus dem Vorstandsbereich Schule

INFO: 19₁₀₉

*Die Bildungsinter-
nationale – der Welt-
dachverband der
Bildungsgewerkschaften
– hat im Juli ein Papier
vorgelegt, in dem alter-
native Wege aufgezeigt
werden, die Ergebnisse
von Leistungsverglei-
chen wie PISA zu beur-
teilen und zu präsentie-
ren.*

*In den kritischen Blick
gerät dabei auch die
Haltung des Deutschen
PISA-Konsortiums.
Wir dokumentieren die
deutsche Übersetzung
von Silvia Schütze.*

9. September 2009

Peter Mortimore:

Alternative Modelle zur Analyse und Darstellung der PISA-Ergebnisse einzelner Länder

Juli 2009

Aus dem Englischen übersetzt von Sylvia Schütze

Peter Mortimore:

Alternative Modelle zur Analyse und Darstellung der PISA-Ergebnisse einzelner Länder

Juli 2009

Aus dem Englischen übersetzt von Sylvia Schütze

Zusammenfassung

Dieser unabhängige Bericht wurde von „Education International“ – der „Bildungsinternationale“ – in Auftrag gegeben. Er stellt Fragen in Bezug auf die derzeitige Form und Schwerpunktsetzung von PISA und will, wenn möglich, Verbesserungsvorschläge unterbreiten.

Es wird anerkannt, dass PISA allen, die mit Schulen und Bildungssystemen zu tun haben, sehr viele Informationen bietet. Es liefert eine Fundgrube von stichhaltigen und verlässlichen Daten.

Gleichzeitig leidet PISA aber auch unter einigen Begrenzungen: Es misst nur eine kleine Menge von dem, was in Schulen gelehrt wird; es kann nur Querschnittsergebnisse anbieten; es ignoriert die Rolle und den Beitrag von Lehrkräften; und die Art und Weise, wie die Ergebnisse präsentiert werden – zumindest in einigen der Tabellen –, lädt zu einer oberflächlichen Ranglisten-Wahrnehmung ein, wo doch ein interessanteres, aber vor allem komplexeres Bild entstehen könnte.

Es gibt kaum einfache Lösungen für diese Probleme. Aber mit etwas gutem Willen kann das PISA-Konsortium Einiges verbessern. Dieser Bericht rät zu etlichen Veränderungen. Insbesondere empfiehlt er:

- eine Verschiebung der Testziele von PISA von „Schnappschuss-Aufnahmen“ des jeweiligen nationalen Leistungsstandes hin zu einer detailreicheren Interpretation der Stärken und Schwächen der einzelnen Länder im Hinblick auf die Förderung lebenslangen Lernens ihrer Bevölkerung;
- eine Rückbesinnung auf die Frage, wie Schulen und Bildungssysteme ihre Leistungen steigern und für mehr Bildungsgerechtigkeit sorgen können;
- eine Ausweitung des Testbereichs, beispielsweise durch Einbeziehung der Kenntnisse in Geographie und Sozialkunde;
- die Beteiligung von Lehrkräften bei Design und Entwicklung der nächsten PISA-Runden;
- die Einbeziehung von Lehrerinformationen, um die Daten stärker zu kontextualisieren;
- eine Ausweitung der Methodologie durch Implementierung einer Längsschnittskomponente;
- die Umgestaltung der wenigen Tabellen, die derzeit noch nach Rangfolge geordnet sind, in eine alphabetische Reihenfolge.

Einleitung

PISA (das Programm zur Internationalen Schüler-Bewertung) ist ein internationales Großprojekt, das die OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung) gemeinsam mit mehr als fünfzig Ländern durchführt. Es wurde 1997 gestartet und führt in dreijährigem Abstand Tests zum Leseverständnis sowie zu mathematischen und naturwissenschaftlichen Kompetenzen durch (2003 außerdem zum Problemlösungsverhalten); diese Tests werden in den einzelnen Ländern mit fünfzehnjährigen Schüler/innen durchgeführt um herauszufinden, wie gut sie „darauf vorbereitet sind, sich den Herausforderungen der heutigen Wissensgesellschaft zu stellen“ (vgl. PISA 2006a, S. 20).¹ Laut PISA liefern die Tests folgende Befunde:

- ein Kenntnis- und Fähigkeitsprofil der Fünfzehnjährigen;
- Kontextindikatoren für den Zusammenhang von Schülerleistungen und Schulmerkmalen;
- einen aktuellen Kenntnisstand für Bildungspolitik und -forschung;
- Trendindikatoren, die anzeigen, wie sich Ergebnisse im Laufe der Zeit verändern (vgl. ebd., S. 22).

Der erste Überblick im Jahre 2000 fokussierte das Leseverständnis, drei Jahre später lag der Schwerpunkt auf mathematischen und 2006 auf naturwissenschaftlichen Kompetenzen, d.h. es wurde jeweils weniger detaillierter Wert auf die beiden anderen Kompetenzbereiche gelegt. Nach jeder Untersuchung wurde ein detaillierter Bericht vorgelegt. PISA verwendet auch Fragebögen, um Informationen über den familiären Hintergrund und die Schulen der Schüler/innen zu gewinnen. Der letzte Bericht beinhaltet außerdem Informationen aus einem Elternfragebogen, der in einigen der teilnehmenden Länder eingesetzt worden war. Nach Aussage von PISA werden die Informationen aus den Fragebögen gemeinsam mit den Leistungsergebnissen ausgewertet, um Folgendes darzustellen:

- Unterschiede zwischen einzelnen Ländern in Bezug auf das Verhältnis von schülerbezogenen Faktoren (z.B. Geschlecht, sozioökonomischer Hintergrund) und Leistungen;
- Unterschiede in Bezug auf das Verhältnis von schulbezogenen Faktoren und Leistungen in den einzelnen Ländern;
- Unterschiede im Hinblick auf die Breite der Leistungsabweichungen zwischen verschiedenen Schulen (eher als: innerhalb einzelner Schulen) und diesbezügliche Unterschiede in den einzelnen Ländern;
- Unterschiede zwischen den Ländern hinsichtlich des Ausmaßes, in dem Schulen die Auswirkungen individueller Schülerfaktoren auf Leistungen abschwächen oder verstärken;
- Unterschiede in den Bildungssystemen und in ihrem nationalen Stellenwert, die mit Unterschieden bei den Schülerleistungen in Zusammenhang stehen;
- durch Verknüpfung von PISA 2000 und PISA 2003 Veränderungen in diesen Beziehungen im Laufe der Zeit (vgl. ebd., S. 20f.).

PISA bietet den Regierungen der teilnehmenden Staaten zahlreiche Aufschlüsse, mit deren Hilfe sie die Leistungsfähigkeit ihres Bildungssystems überprüfen können. Es liefert Daten, die von Regierungen und ihren Wählern – sowie auch von Lehrkräften und anderen im Erziehungs- und Bildungsbereich Tätigen – genutzt werden können, um über die Stärken und Schwächen ihrer Schulen und ihres Bildungssystems im Vergleich mit anderen Ländern zu debattieren.

¹ Die englischen Originalzitate wurden ins Deutsche übersetzt. Den Literaturhinweisen ist deshalb jeweils ein „vgl.“ vorangestellt.

Durch das Angebot knapper Diskussionspunkte zu allen wichtigeren Ergebnissen in der Rubrik „Implikationen für die Politik“ versucht PISA auch, „... eine neue Basis für den politischen Dialog und für die Zusammenarbeit bei der Definition und Implementierung von Bildungszielen bereitzustellen“ (vgl. PISA 2006b, S. 3).

Zurzeit fehlen bei PISA aber viele Anknüpfungspunkte zum lebenslangen Lernen. Wenn man von der allgemein anerkannten Tatsache ausgeht, dass ein produktives – und damit glückliches – Leben im 21. Jahrhundert von fortwährendem Lernen abhängig ist, dann handelt es sich dabei um eine ernstzunehmende Auslassung. Ließe sich vielleicht auf Erhebungsinstrumenten wie dem „Index für zukunftsorientierte Motivation, Naturwissenschaften zu lernen“ – der bei der Untersuchung von 2006 verwendet wurde (vgl. ebd., S. 149) – aufbauen, um diesem Bedarf abzuwehren?

Die Bildungsinternationale und PISA

Einer Publikation der Bildungsinternationale (Education International; im Folgenden: EI) zufolge sind „die Wirkungen von PISA auf die Medien nicht zu übersehen“ (vgl. Education International 2007, § 12). Aus diesem Grunde wäre es dumm, wenn Bildungsgewerkschaften sich nicht damit beschäftigen. Aber es gibt auch andere Gründe, um sich genauer mit PISA zu befassen: Die reichhaltigen Informationen über Schülerleistungen und deren Zusammenhang mit dem sozioökonomischen Hintergrund der Schüler/innen sowie Merkmalen von Schulen und Bildungssystemen können uns darüber aufklären helfen, inwiefern Schulen und Bildungssysteme Leistungen fördern und Gerechtigkeit bei Bildungsabschlüssen steigern können.

Das Hauptanliegen von EI ist es, die Daten von PISA konstruktiver zu nutzen. Es ist beklagt worden, dass das Schwergewicht auf den Leistungen beim Leseverständnis, in der Mathematik und in den Naturwissenschaften den Rest der schulischen Arbeit abwertet, z.B. den Unterricht in anderen Fächern. Außerdem führt die Faszination der Medien über simple Ranglisten einzelner Länder zu einer oberflächlichen Debatte über Dinge, die an sich sehr komplex sind. Außerdem hat eine Untersuchung von EI zu den PISA-Daten von 2003 gezeigt, dass „Bildungsminister und ihre Berater/innen ... versuchen, die Daten so zu ‚drehen‘, dass sie zu ihren eigenen Absichten passen“ (vgl. ebd., § 13).

Der hier vorliegende unabhängige Bericht wurde von EI in Auftrag gegeben, um diese Aspekte und weitere Kritikpunkte am Vorgehen von PISA zu untersuchen und zu überlegen, ob alternative Methoden der Datenpräsentation besser geeignet sind, um eine ernsthafte Debatte über die Qualität und die Gerechtigkeit von Bildungsmethoden und -systemen zu führen. Die Perspektive des Berichts ist die eines „Daten-Nutzers“, der sowohl mit Wissenschaftler/inne/n als auch mit Schulpraktiker/inne/n zusammenarbeitet und der, vor dem Hintergrund vieler Jahre Schulforschung (vgl. Mortimore 1998), den Begriff der „Effektivität von Bildungssystemen“ verwendet, um einen Zustand zu beschreiben, bei dem alle Schüler/innen Zugang zu maximaler Schulqualität erhalten – und zwar in pädagogischer, nicht in simpler ökonomischer, Hinsicht.

Generelle Kritik an PISA

Es hat bereits viel Kritik an den konzeptionellen, methodischen und politischen Aspekten des Programms gegeben (vgl. Prais 2003 und 2004; Smithers 2004; Goldstein 2004; Burkhardt 2005; Wuttke 2006; Hopmann/Brinek/Retzl 2007 – darunter etliche kritische europäische Beiträge), zuletzt von Grek (2009).

Aus diesen kritischen Beiträgen haben sich zehn wichtige Themen herauskristallisiert:

1. **Kulturelle Unterschiede**

Schüler/innen aus mehr als fünfzig Nationalitäten zu testen, ist für jede Studie eine große Herausforderung. Schüler/innen mit sehr unterschiedlichem kulturellem Hintergrund können auf allgemeine Fragen sehr unterschiedlich reagieren, sogar auf eine ganz gewöhnliche Testsituation. Das Problem mag im Laufe der Zeit und mit fortschreitender Entwicklung des Programms geringer werden, da auch die Erwartungen der Lehrkräfte und der Schüler/innen sich anpassen. Auch kann es natürlich nicht die Absicht der Testentwickler/innen sein, solche kulturellen Unterschiede „glatt zu bügeln“ (die für die Identität einzelner Völker sehr wichtig sein können). Es ist aber von großer Bedeutung, dass die Existenz solcher Unterschiede bei der Interpretation von PISA-Ergebnissen im Hinterkopf behalten wird.

2. **Übersetzungsfragen**

Es bestehen auch Unterschiede in der Formulierung einzelner Items. Einige Sprachen sind komplizierter als andere. Solche mit mehr regelmäßigen grammatischen Konstruktionen verursachen z.B. vermutlich auch weniger Lese- und Schreibprobleme. Außerdem muss berücksichtigt werden, dass einige Schüler/innen nicht in ihrer Muttersprache getestet werden. Das trifft auf viele Immigrant/inn/en der ersten Generation zu, aber auch auf Schüler/innen in Ländern wie Luxemburg, wo sie entweder auf Französisch oder auf Deutsch getestet werden. PISA ist 2006 in seinem Technical Report (2006a, S. 89f.) auf das Problem der Übersetzungsfragen eingegangen und hat die Maßnahmen dargestellt, die diesbezüglich ergriffen wurden – z.B. Übersetzung und Rückübersetzung von Fragen und Test-Items –, um wenigstens einige dieser Probleme besser in den Griff zu bekommen.

3. **Zusammenstellung der Untersuchungssamples**

Bei jeder Untersuchung – die im Gegensatz zu einer Volksbefragung nicht alle Personen eines bestimmten Alters umfasst – kann man Fragen zur Repräsentativität des ausgewählten Samples stellen. Es besteht die Gefahr, dass einige Länder versuchen, den Anteil der fähigeren Schüler/innen zu erhöhen und den Anteil derjenigen, von denen man annehmen kann, dass sie am schwächsten abschneiden, zu verringern. Aber Schüler/innen mit Lernschwierigkeiten und solche, die die Landessprache nicht fließend beherrschen, gehören bei einem korrekten Ausschuss dazu. Mit diesem Thema haben sich etliche Kritiker/innen befasst, u.a. der Beitrag „Die verschwundenen Schüler/innen: PISA und Schüler/innen mit Behinderungen“ von Bernadette Hörmann (Universität Wien) (vgl. Hörmann 2007).

Die PISA-Leitlinien befassen sich in bemerkenswerter Detailgenauigkeit mit diesem Problem. Das Sample jedes Landes wird sorgfältig untersucht und mit bestimmten Kriterien abgeglichen. Es gab bereits Fälle, wo entschieden wurde, dass das Sample nicht gut genug war – z.B. die Ablehnung des britischen Samples im Jahre 2003. (Aus diesem Grunde wurden die britischen Ergebnisse nicht in der üblichen Weise publiziert, sondern in einen Anhang „verbannt“. Nachfolgende Analysen bestätigten, dass das britische Sample nicht repräsentativ gewesen wäre, da es tatsächlich leistungsfähiger war, als es ein zufällig ausgewähltes Sample gewesen wäre [vgl. Micklewright/Schnepf 2006].)

4. **Außerachtlassung länderspezifischer Curricula**

Es ist auch kritisiert worden, dass ein Schwergewicht auf solchen Fragen liegt, die sich mit dem gesunden Menschenverstand beantworten lassen und nicht auf Kenntnisse aus einem spezifischen Curriculum abzielen. Kommentatoren wie Thomas Jahnke von der Universität Potsdam argumentieren, dass Politiker/innen und die Öffentlichkeit dadurch dazu verleitet werden, den eigentlichen Lehrplan geringer zu achten (vgl. Jahnke 2007). Auch EI ist dieser Frage bereits nachgegangen und hat festgestellt, dass „PISA kein Gesamtbild der Bildung bietet“ (vgl. Education International 2007, § 14). Gerade in letzter Zeit, wo PISA zunehmend als ultimativer Maßstab für „Bildungsqualität“ angesehen wird, richtet sich die Aufmerksamkeit der Öffentlichkeit und von Politiker/innen fast ausschließlich auf die Hauptfächer Mathematik, Na-

turwissenschaften und Lesen. Das marginalisiert Fächer wie Geschichte, Geographie, Sozialkunde, Fremdsprachen und viele andere.

Das Ziel von PISA ist es zu messen, wie gut junge Leute in der Lage sind, ihre Kenntnisse und Fertigkeiten einzusetzen, „um mit den Herausforderungen des wirklichen Lebens zurecht zu kommen, anstatt festzustellen, wie gut sie mit einem spezifischen Curriculum fertig geworden sind“. Deshalb legt PISA sein Schwergewicht „innerhalb der Bereiche Leseverständnis, Mathematik und Naturwissenschaft“ auf „die Beherrschung von Prozessen, das Verständnis von Konzepten und die Fähigkeit, in verschiedenen Situationen zu funktionieren“ (vgl. PISA 2006a, S. 22).

Wenn die PISA-Daten nur als grobes Schätzergebnis für die Leistungsfähigkeit des Bildungswesens eines Landes und seine Bemühungen, Bildungsgerechtigkeit herzustellen, genommen werden, dann mögen die Grenzen des getesteten Curriculums nicht ins Gewicht fallen. Wenn die Daten jedoch als definitives Urteil über nationale Bildungssysteme angesehen werden, dann werden die Resultate verheerend sein. Sie werden die gewaltige Sogwirkung haben, dass dauerhaft zwischen einem sogenannten Kerncurriculum (Leseverständnis, Mathematik und Naturwissenschaften, die von PISA getestet werden) und einer weniger wichtigen Peripherie mit anderen Fächern unterschieden wird.

Eine Möglichkeit, um diese Gefahr zu vermeiden, liegt darin, auch Inhalte anderer Fächer, die Wissen von internationaler Bedeutung enthalten, zu integrieren – wie es PISA 2003 mit dem Einbezug von Problemlösungsfähigkeiten schon versucht hat. So könnten die Fächer Geographie oder Sozialkunde einige passende Themen bieten, um die derzeitigen PISA-Tests zu erweitern.

5. **Verwerfungen in der Bildungspolitik**

EI führt an, dass PISA durch die Konzentration auf eine begrenzte Zahl von Fähigkeiten Regierungen dazu verleitet, sich für einen vor allem „wirtschaftlichen“ Zugang zu Bildungsfragen zu entscheiden – zur Durchführung von „Reformen zum Testen von Schüler/innen und Evaluieren von Lehrkräften, die von Effektivitäts- und Wettbewerbskriterien inspiriert sind“ (Education International 2007, § 25). Einige Kritiker/innen haben die Vermutung geäußert, dass es den privaten Firmen, die für PISA arbeiten, vor allem darum geht, ihren Anteil am „beständig wachsenden Test-Markt“ zu vergrößern (vgl. Bracey 2005, S. 797) oder – noch krasser – dass „PISA ein Absatzmarkt des New Public Government der neo-liberalsten Sorte ist“ (vgl. Huisken 2005).

6. **Mangelhafte Einbindung von Lehrkräften**

PISA erhebt zurzeit keine Daten über die Lehrerschaft. Zwar wurden zusätzlich zu den Testverfahren Fragebögen für Schulleitungen und Schüler/innen ausgewertet und 2006 in sechzehn Ländern auch Informationen von einem Sample von Eltern erhoben: „Der Fragebogen erbrachte Informationen über die Investitionen von Eltern in die Bildung ihrer Kinder und über ihre Ansichten zu Themen und zu Karrierechancen aus dem Bereich der Naturwissenschaften“ (vgl. OECD 2007b). Es erscheint daher etwas seltsam, dass diejenigen, die in den Klassenzimmern ihre Arbeit tun – im Herzen des Bildungsprozesses – und die zur Verbesserung der Qualität und der Gerechtigkeit am meisten beitragen können, ignoriert werden.

Natürlich werden einige Lehrer/innen auch gar nicht beteiligt werden wollen; andere möchten ihre kostbare Zeit nicht auf etwas verwenden, was ihnen marginal vorkommt. Andere aber heißen die Gelegenheit zur Teilnahme möglicherweise sehr willkommen und können auf diese Weise dazu beitragen, die Lücke zu füllen, die zur Zeit zwischen der Schulstrategie – die die Schulleitung zum Ausdruck bringt – und den Ansichten und Leistungen der Schüler/innen klafft.

7. **Die Nutzung von Querschnittsdaten**

Das Forschungsdesign von Pisa ist auf Querschnitt angelegt. Alle drei Jahre werden nationale Samples gleichaltriger Schüler/innen getestet. Obwohl einiges dafür spricht, dass jedes Sample grundsätzlich seinem Vorgänger ähnelt, so können sich doch die Merkmale des Samples verändern (beispielsweise durch einen Anstieg oder eine Abnahme des Immigrantenteils), und das macht Vergleiche zwischen den verschiedenen Jahrganggruppen schwierig.

Ein Kritiker bemerkte zu diesem Problem: „Die festgestellten Unterschiede spiegeln zweifellos Unterschiede zwischen Bildungssystemen, aber sie spiegeln auch soziale und andere Unterschiede, die man nicht genau bestimmen kann“ (vgl. Goldstein 2004). Darüber hinaus braucht man, um valide Vergleiche der Auswirkungen von verschiedenen nationalen Bildungssystemen anstellen zu können, im Grunde genommen Längsschnittdaten. Goldstein bringt es auf den Punkt: „Der Bericht von 2003 behauptet, dass das Leseverständnis ‚einen unmittelbaren Effekt auf das Bruttoeinkommen, das Beschäftigungsverhältnis, die Gesundheit hat‘. Solche kausalen Effekte mögen zwar existieren, aber sie können nicht aus Querschnittsuntersuchungen gewonnen werden“ (vgl. ebd.).

Viele Kommentatoren merken an, dass PISA – insbesondere von Regierungen und den Medien – benutzt wird, um Kausalitäten zu konstruieren, obwohl die Studie sich darauf beschränkt, Korrelationen zwischen verschiedenen Messungen herzustellen. Zumindest für einen Teil der Untersuchung ein Längsschnittdesign einzuführen, würde den Einfluss von PISA stärken. Es würde natürlich auch die Kosten steigern.

8. **Die Datenmodellierung**

Goldstein (2004) hat kritisch auf die Methode hingewiesen, mit deren Hilfe PISA entscheidet, ob Items auf eindimensionale Skalen passen – Antwortmodellierung –; auch Blum u.a. haben dieses Verfahren kritisiert, „weil es zu subtilen Verzerrungen führen und wichtige Unterschiede zwischen Ländern ‚glätten‘ kann“ (vgl. Blum/Goldstein/Guerin-Pace 2001). Goldstein stellt außerdem fest, dass „Items, die in einigen Ländern unterschiedlich angenommen wurden, ... in den Verdacht geraten sind, eine kulturelle Verzerrung herbeizuführen. ... Aus diesem Grunde wurden einige Items gestrichen.“ Solche Items werden im PISA-Bericht als „uneindeutig“ bezeichnet (vgl. z.B. PISA 2006b, Report 3b, S. 152).

PISA hat einige Mehrebenen-Modellierungs-Techniken eingeführt, um Unterschieden zwischen Schulen dadurch gerecht zu werden, dass Merkmale der Schülerschaft berücksichtigt werden – z.B. das Geschlecht und der soziale Hintergrund. Das ist hilfreich, denn es kann dazu beitragen, weitere Varianten in die Modelle einzubeziehen.

9. **Auswirkungen der Aufstellung von Ranglisten auf die nationalen Schulsysteme**

Wenngleich die PISA-Daten sehr detailliert sind, akribisch auf die Berücksichtigung von Standardabweichungen geachtet wird und zahlreiche Vergleiche der Durchschnittsleistungen durchgeführt werden um festzustellen, welche nationalen Unterschiede statistisch signifikant sind, beziehen sich die Diskussionen in den Medien fast ausschließlich auf den „Ranglistenplatz“ der einzelnen Länder. Die Tabellen dominieren die Nachrichten so sehr, dass ein großer Teil der sorgfältigen Arbeit des PISA-Konsortiums, z.B. zum Verhältnis zwischen den einzelnen Wertungsbereichen und zu den kontextuellen Faktoren, den Einstellungen und den Schul- und Systemfaktoren ignoriert wird – außer von Wissenschaftler/inne/n.

Die Konzentration auf diesen begrenzten Datenbestand durch die Medien hat einzelne Länder veranlasst, sich an den Tabellen zu orientieren. EI schreibt dazu: „Das führt in der Regel zur Einführung von noch mehr Leistungsmessungen, von nationalen Tests, die sich an Modell und Methodologie von PISA orientieren, zu mehr Inspektionen und schließlich zur Verknüpfung von Leistung und Bezahlung der Lehrkräfte mit den Testergebnissen der Schüler/innen“ (vgl. Education International 2009).

In England (vgl. Mortimore 2009), wo das Testen ganzer Alterskohorten mit Standardleistungstest inzwischen gesetzlich vorgeschrieben ist, zeigt sich, dass Tabellenplätze die öffentliche Diskussion über Schulen dominieren. Die Ranglisten befeuern die Entwicklung einer marktwirtschaftlichen Sicht auf Schulstandorte, obwohl sie sehr fragwürdig sind, weil sie die Unterschiede der jeweiligen Schulklientel nicht berücksichtigen (vgl. Mortimore 2008).

Viele Eltern sind jedoch durch die staatliche Unterstützung von Ranglisten dahingehend fehlgeleitet worden, dass sie diese für eindeutige Wegweiser zu den besten Schulen halten. Infolgedessen werden diese Schulen von ehrgeizigen Eltern, die für ihre Kinder das Beste wollen, eifrig nachgesucht. Ein weiteres Problem besteht darin, dass dort, wo es mehr Bewerber/innen als Schulplätze gibt, die Entscheidung von den Eltern auf die Schulen übergeht. In England sind Fälle vorgekommen, wo Eltern falsche Adressen angegeben haben, um ihre Kinder in gewissen Schulen unterzubringen, und sich dafür vor Gericht verantworten mussten. Sogar der Preis von Häusern wird durch die Ranglisten beeinflusst; Grundstücke im Einflussbereich von hochrangigen Schulen erzielen Spitzenpreise. Diese Schulen werden in der Regel mit Anmeldungen überschüttet. Schlecht platzierte Schulen hingegen müssen oftmals darum kämpfen, genug Schüler/innen zu gewinnen, um nicht geschlossen zu werden.

Ranglisten schaden auch der Bildungsgerechtigkeit, weil sie Lehrkräfte dazu bringen, sich auf die Schüler/innen zu konzentrieren, die dazu beitragen können, den Listenplatz der Schule zu verbessern. Diejenigen, die im Hinblick auf die Tests als „Wackelkandidaten“ gelten, erhalten die meiste Unterstützung. Schüler/innen, die besondere Förderung benötigen, sind weniger willkommen, da sie mehr Ressourcen benötigen – und dennoch bei den Tests weniger gut abschneiden und dadurch den Listenplatz der Schule senken.

Außerdem haben Forschungen der OECD klar gezeigt, dass sozial, wirtschaftlich oder kulturell besser gestellte Eltern eher in der Lage sind, für ihre Kinder die richtigen Entscheidungen zu treffen (vgl. OECD 2007a). In einem System, das auf Ranglisten aufgebaut ist, ist es klar, dass die begehrtesten Schulen aller Wahrscheinlichkeit nach von diesen Eltern angesteuert werden, so dass die weniger attraktiven Schulen für die schlechter gestellten Familien übrig bleiben.

In den letzten zwanzig Jahren wurde der Alltag in englischen Schulen von Ausscheidungstests dominiert. Marshall (2004) hat ausgerechnet, dass jede/r Schüler/in durchschnittlich einhundert formale Tests oder Prüfungen zu absolvieren hat (vgl. Marshall 2004). Mansell (2007) hat die lähmenden Wirkungen eines so massenhaften Testens beschrieben (vgl. Mansell 2007). Das Schlüsselproblem von Leistungsmessungen liegt nicht so sehr im Testen an sich, sondern darin, wie die Informationen, die bei diesen Tests gewonnen werden, benutzt werden, um zu bestimmen, welche Leistungen als Auswahlkriterien zu gelten haben – mit ernsthaften Konsequenzen für Schüler/innen, Lehrer/innen und Schulen.

Das Durchführen von Tests spielt insofern für das Lernen von Schüler/inne/n eine positive Rolle, als ihnen und ihren Lehrkräften hilfreiche Rückmeldung gegeben wird. Das Problem entsteht daraus, dass der Fokus der Ausscheidungstests nicht auf dem Lernen liegt, sondern auf der Platzierung.²

Auch wenn Regierungen Ranglisten nicht unterstützen – die Medien tun es. PISA muss deshalb überlegen, wie es seine Daten so präsentieren kann, dass die Möglichkeiten für Ranglisten eingeschränkt werden und ein reflektierterer und positiverer Umgang mit den Ergebnissen gefördert wird.

10. **Herrschaftswissen und Heimlichtuerei**

Die OECD und andere internationale Einrichtungen haben zweifellos zunehmend an Einfluss gewonnen, da die Diskussion über Globalisierung bei den Regierungen zugenommen hat. Ob-

² Vgl. die ausführliche Diskussion dieses Aspektes im „House of Commons“, „Children, Schools and Families Select Committee Testing and Assessment“, 3. Bericht der Sitzungsperiode 2007/2008.

wohl die Wirtschaftspolitik bei der OECD weiterhin Vorrang genießt, hat ihr Interesse an Bildung in diesem Kontext „wachsende Bedeutung gewonnen, da diese im Rahmen des Konzepts von ‚Humankapital‘ für die nationale ökonomische Wettbewerbsfähigkeit als zentral angesehen und in Zusammenhang mit der expandierenden ‚Wissensökonomie‘ gebracht wird“ (vgl. Grek 2009). Nach Grek „hat die OECD als methodisch hochkompetente Agentur eine Nische für die Entwicklung von Bildungsindikatoren und vergleichenden Bildungsleistungsmessungen geschaffen“ (vgl. ebd.).

Andere Kommentatoren haben darauf hingewiesen, dass angesichts des schwächer gewordenen Einflusses der Regierungen auf ökonomische Entwicklungen – wie die letzte weltweite Finanzkrise gezeigt hat – „der Wettbewerbsvorteil zwischen Nationen häufig in Kategorien der Qualität der nationalen Bildung neu bestimmt wird und Bildungssysteme nach internationalen Standards beurteilt werden“ (vgl. Brown u.a. 1997). Dadurch ist PISA, das von Regierungen von über fünfzig Ländern unterstützt wird, mit seinem technischen Knowhow und seinem Anspruch, komplexe Informationen in leicht verständlicher Weise zu präsentieren, zu einem der Schlüssel-Weltschiedsrichter für die Qualität und Gleichheit von Schulsystemen geworden. Nach Meinung einiger deutscher Kritiker hat PISA jedoch durch den nachgewiesenen Mangel an Bereitschaft von Mitgliedern des deutschen Konsortiums, sich an einer offenen Diskussion über die Vorannahmen, Methoden und Folgen zu beteiligen, an Ansehen verloren. Auf einer offenen Internetplattform stellen Hopmann u.a. dar, dass „wiederholte Einladungen, diese Themen in offenen Symposien anzusprechen oder zu dem Band ‚PISA According to PISA‘ (‚PISA gemessen an PISA‘) beizutragen, unbeantwortet blieben oder zurückgewiesen wurden“ (Hopmann/Brinek/Retzl 2007). Die Kommentatoren listen eine Anzahl von Beschwerden über das deutsche Konsortium und seine Reaktionen auf die Einladung zu einem wissenschaftlichen Dialog über PISA auf. Wenn diese Behauptungen wahr sind, ist das sehr schade und eine verpasste Gelegenheit. Das Verhalten kontrastiert scharf mit der begeisterten Teilnahme des leitenden PISA-Konsortiums an einem Treffen in London im Jahre 2005 (vgl. Department for Children, Schools and Families 2005).

Es würde dem PISA-Konsortium sehr gut anstehen, eine Politik der Offenheit zu betreiben und seine Bereitschaft zu zeigen, sich in Frage stellen zu lassen. Es ist auch wichtig, dass Wissenschaftler/innen, die seit Jahren im Bereich von Leistungsmessung arbeiten und sich um eine bescheidene Forschungsförderung bemüht haben, gegenüber dem finanziell gut ausgestatteten, international breit unterstützten PISA-Projekt nicht verbittert werden. Es bieten sich ungeheure Gelegenheiten zur Zusammenarbeit und ausgezeichnete Möglichkeiten für Studien, die auf den vollständigen PISA-Daten, die das Konsortium zur Verfügung stellt, aufbauen. Die Computerlinks, über die man von jeder Tabelle eine Excel-Datei herunterladen kann, stellen eine herausragende Neuerung dar. Es ist zu hoffen, dass die Antwort des PISA-Konsortiums auf die hier vorliegende Veröffentlichung die Zusammenarbeit innerhalb dieses Konsortiums sowie mit Wissenschaftler/inne/n und Schulpraktiker/inne/n verbessert.

Notwendige Veränderungen

Vor allem die folgenden vier Punkte müssen nach Auffassung von EI verändert werden:

1. Einbeziehung von Lehrkräften

Es gibt zwei Wege, wie Lehrkräfte einbezogen werden können. Zum einen könnten sie viel stärker in die Entwicklung von Teststrategien und Test-Items eingebunden werden. Die Einbeziehung von Lehrer-Fokusgruppen würde es beispielsweise ermöglichen, dass das professionelle Wissen von Lehrer/inne/n an zentralen Stellen der Entwicklung weiterer PISA-Runden berücksichtigt wird.

Zum anderen könnte das PISA-Konsortium Fragebögen für Lehrkräfte an den untersuchten Schulen entwickeln. Die Daten, die auf diese Weise zu gewinnen sind, würden die Aussagen der Schüler/innen und Eltern ergänzen und damit einen besseren Kontext für die Interpretation der Testergebnisse bieten.

Es muss zwar akzeptiert werden, dass einige Lehrer/innen und natürlich auch einige nationale Gewerkschaften eine Teilnahme an den PISA-Tests ambivalent beurteilen, aber es würde dem PISA-Konsortium gut anstehen, diese Möglichkeit immerhin anzubieten.

Auf diese Weise könnten Informationen über die Erwartungen, Einstellungen, Lehrmethoden und Weiterbildungsabsichten von Lehrkräften gewonnen und ihre Ansichten über die Schülerschaft eruiert werden. Einige Fragen könnten vom OECD-Talis-Projekt (vgl. www.oecd.org/edu/talis; Zugriffsdatum: 24.08.2009) übernommen werden, andere müssten noch formuliert werden, um die Aussagen der Schüler/innen, Schulleitungen und Eltern sinnvoll zu ergänzen.

Es ist wichtig zu beachten, dass eine solche zusätzliche Mitarbeit sich auf die Arbeitszeit von Lehrkräften auswirken kann.

2. Verwendung von Längsschnittdaten

Die Grenzen des Test-Designs von PISA wurden in diesem Text bereits an früherer Stelle angesprochen. Es gibt drei wichtige Argumente für den Einsatz eines Längsschnitt-Designs:

- Progredienzmessungen sagen mehr als „Momentaufnahmen“;
- individuelle Längsschnittdaten sind aufschlussreicher als Querschnittsdaten;
- auf der Basis von Längsschnittinformationen können kausale Schlüsse gezogen werden.

Statistiker wie Goldstein (2004) treten für die Einbeziehung einer Längsschnittkomponente bei den nächsten PISA-Runden ein. Das PISA-Konsortium sollte die Hinzufügung einer solchen Komponente ernsthaft in Erwägung ziehen – sie könnte ja in kleinerem Rahmen stattfinden –, um die Querschnittsergebnisse zu ergänzen.

3. Verbesserungen bei der Anlage der Tabellen

Insgesamt ist es völlig unbestritten, dass die PISA-Daten sehr gut präsentiert werden. Band 1 enthält eine Fülle von vielfältigen Informationen in überwiegend gut geeigneten Formaten. Band 2 enthält mehr als 175 Tabellen, die die Länder in alphabetischer Reihenfolge auflisten.

Die Abschnitte des Berichtes, die sich mit dem beschäftigen, „Was PISA misst und wie“ sowie „Die Durchführung von PISA: Was gemessen wird“ sowie Abbildung 1.2, die einen „Über-

blick über die Test-Bereiche von PISA 2006“ bietet (vgl. PISA 2006a, S. 20), sind wahre Muster an Verständlichkeit. Kasten 2.2 auf Seite 50, „Die Interpretation der Sample-Statistiken“, stellt die Probleme der Unsicherheit, der Konfidenzintervalle und der statistischen Bedeutung sehr gut dar. Dadurch sollte ein/e Leser/in, der bzw. die den Umgang mit Statistiken nicht gewohnt ist, in der Lage sein, sowohl die Grenzen als auch die Stärken der Daten einzuschätzen. Kasten 3.2. – „Interpretation der PISA-Indizes“ – ist sehr hilfreich, ebenso Kasten 5.1, der sich mit den Grenzen der Analysen befasst, und die Kästen 5.2 bis 5.9, die die Mehrebenen-Modellierung betreffen.

Wie bei jedem komplexen Bericht werden die Daten auch bei PISA in verschiedenen Formaten präsentiert. Dazu zählen alphabetische Listen der teilnehmenden Länder, bei den OECD-Ländern von Australien bis zu den Vereinigten Staaten von Amerika und bei den Partnerländern von Argentinien bis Uruguay; Ranglisten, die nach Leistung gestaffelt sind; Streudiagramme, die den Zusammenhang von zwei oder mehr Variablen abbilden; und Mehrebenenendiagramme, die die Auswirkungen unterschiedlicher Modellierung der Daten zeigen.

- **Alphabetische Listen**

Diese werden verwendet, um die Ergebnisse verschiedener Tests aufzulisten (Abb. 2.13), die Messwerte einzelner Indizes des Berichts abzubilden (Abb. 3.4) und Informationen aus den Fragebögen wiederzugeben. Dieser Listentyp wird auch für die etwa 170 detaillierten Tabellen im zweiten Teil des Berichts verwendet, der die Einzelergebnisse enthält. Also verwenden etwa 200 der 300 Tabellen und anderen Aufstellungen des Berichts ein alphabetisches Format.

- **Ranglisten**

Diese werden verwendet, um die Ergebnisse der drei Haupttestbereiche zu präsentieren: Naturwissenschaften, Leseverständnis und Mathematik (vgl. Abb. 2.11b, 6.8a. und 6.20a), außerdem für die Abbildungen, die den Anteil von Schüler/inne/n auf den jeweiligen Leistungsniveaus der drei Tests zeigen (vgl. Abb. 2.11a, 6.1 und 6.19) sowie für einige besondere Unterkategorien (z.B. Abb. 2.24e).

Tabellen mit mehreren Variablen – darunter die meisten (allerdings nicht alle) erstellten Indizes – sind ebenfalls als Ranglisten aufgebaut: manchmal in aufsteigender Reihenfolge (vgl. Abb. 5.17), aber häufiger in absteigender Reihung (vgl. Abb. 3.14).

- **Streudiagramme**

Die Beziehungen zwischen Variablen werden mittels Streudiagrammen dargestellt. Zum Beispiel zeigt Abb. 2.1a die Beziehung zwischen dem Ergebnis auf der Naturwissenschaftsskala und der Größe des Bruttonationalprodukts, und Abb. 6.10 illustriert die Überlegenheit der Mädchen beim Leseverständnistest.

- **Mehrebenenendiagramme**

Der Bericht enthält eine ganze Reihe von Analysen, bei denen verschiedene Modelle zum Einsatz kommen, um den Effekt verschiedener Kombinationen von Kontextvariablen auf die Testergebnisse zu untersuchen. Diese Analysen werden in mehreren Diagrammen abgebildet, aus denen die Veränderungen bei verschiedenen Faktorenkombinationen und das Ausmaß von deren Zusammenhang hervorgehen (vgl. die Kästen 5.1-5.9).

Die häufigste Form der Datenpräsentation ist also die alphabetische Liste. Das Problem liegt jedoch darin, dass diejenigen Tabellen, an denen die Medien das größte Interesse haben, nicht so angezeigt werden, sondern als Ranglisten erscheinen. Eine solche „Leistungstabellen“-Form scheint vom PISA-Konsortium befördert zu werden.

Das ist irreführend, denn der Bericht ist im Umgang mit Daten an und für sich sehr sorgfältig, und eine faire Lektüre seines ganzen Inhalts würde deutlich machen, dass die pauschale Rangfolge der Länder bei den drei Haupttests zwar interessant ist, aber nur einen kleinen Ausschnitt eines größeren Gesamtbildes darstellt, bei dem die Testergebnisse in Abhängigkeit von verschiedenen Umständen, politischen Verhältnissen u.a. dargestellt werden. So ist es z.B. die Hauptabsicht von Abbildung 2.11b auf S. 56f., die multiple Vergleiche von Durchschnittsleistungen auf der Naturwissenschafts-Skala anstellt, zu zeigen, dass – obwohl die Daten in einer Rangliste präsentiert werden – statistisch signifikante Unterschiede nur zwischen den Leistungen einiger weniger Länder vorliegen. So haben zwar Finnland und Kirgisien signifikant unterschiedliche Ergebnisse im Vergleich mit den anderen Ländern (das eine Land in positiver, das andere in negativer Hinsicht), aber alle anderen Länder weisen durchmischte Ergebnisse auf. Im Vergleich mit Dänemark haben beispielsweise 25 Länder signifikant niedrigere Ergebnisse, 20 Länder haben signifikant höhere, und 11 unterscheiden sich nicht signifikant.

Abb. 6.8a auf S. 296f. zeigt Entsprechendes für das Leseverständnis. Die Länder sind vom höchsten Leistungsstand (Korea, 556) bis zum niedrigsten (Kirgisien, 285) angeordnet. In diesem Fall haben sowohl Korea als auch Kirgisien Ergebnisse, die sich von allen anderen Ländern signifikant unterscheiden. Im Vergleich dazu ist das Durchschnittsergebnis für Kroatien signifikant höher als das von 24 Ländern und signifikant niedriger als das von 24 anderen Ländern – und sieben Länder haben keine signifikant bedeutenden Unterschiede.

Auch Abbildung 6.20a auf S. 316f. zeigt die Ergebnisse für Mathematik in einer Rangliste. Im Unterschied dazu listet Tabelle 2.13, die die Gesamtergebnisse für Naturwissenschaften jeder ihrer sieben Einzelskalen gegenüberstellt, die Länder alphabetisch auf. Es wäre ja auch hier möglich gewesen, die Ranglistenform beizubehalten, wurde aber klugerweise nicht gemacht.

Alphabetische Anordnung bei allen Tabellen

Ist es wirklich sinnvoll, bei allen Länderlisten eine alphabetische Reihenfolge zu verwenden? Ein solches Verfahren würde zwar nicht unbedingt die Medien davon abhalten, ihre eigenen Ranglisten zu basteln, aber es würde ein Signal geben, dass das PISA-Konsortium einen solchen Ansatz prinzipiell nicht gutheißt und ihn für unangemessen hält, um solche komplexen Daten darzustellen.

Diese prinzipielle Haltung müsste im Vorwort des nächsten PISA-Berichts deutlich gemacht werden. Zurzeit stellt das Vorwort das allgemeine Grundprinzip für international vergleichende Tests dar, enthält eine kurze Geschichte des Projekts, einen Hinweis auf den innovatorischen Aspekt und eine Liste derjenigen Personen, die für Anlage und Durchführung verantwortlich zeichnen. Es könnte aber auch wichtige „Warnungen vor Risiken und Nebenwirkungen“ beinhalten, die der Gebrauch der Daten mit sich bringen kann, und kurz darlegen, weshalb Ranglisten für unangebracht gehalten werden und wie ihre Verwendung vom eigentlichen Sinn der PISA-Studie wegführt. Ähnliche Hinweise könnten an anderen – gut ausgewählten – Stellen wiederholt werden, so bei Kasten 1.1 „Key Features von Pisa 2006“ und in der Leseanweisung, die sich gegenwärtig auf S. 26 befindet.

Außerdem sollten angemessene und deutlich erkennbare „Warnungen vor Risiken und Nebenwirkungen“ in allen Presseerklärungen enthalten sein.

Einige Tabellen und Diagramme könnten so umgestaltet werden, dass sie zu einem tieferen Verständnis der Daten einladen. Zum Beispiel:

- Gruppierung in diskreten Kategorien (z.B. die Verwendung von Vierteln oder die Einteilung in Spitzen-, Mittel- und Enddrittel oder Halbstandardabweichungen oder andere sich anbietende Kategorien);
- Durchführung von mehr Messungen, z.B. des Unterschiedes zwischen dem 95. und dem 5. Perzentilwert bei verschiedenen Berechnungen;
- mehr Verknüpfungen zwischen den Messergebnissen und kontextuellen und beschreibenden Daten.

PISA ist ein bemerkenswertes Projekt. Es wäre höchst unglücklich, wenn es – durch die Verwendung von Tabellen mit nationalen Ergebnissen in Ranglistenform – mit dem oberflächlichen – und oftmals irreführenden – Ansatz einiger Medien assoziiert würde. Wie bereits erwähnt, wird die Mehrheit der PISA-Tabellen zurzeit in alphabetischer Reihenfolge präsentiert; eine entsprechende Veränderung der wenigen anderen würde nicht nur zu größerer Konsistenz, sondern auch zu einer angemesseneren Form der Datenpräsentation führen.

Empfehlungen und Schlussfolgerungen

PISA ist ein Riesenerfolg der OECD. Es bietet Regierungen, Wissenschaftler/innen und Wähler/innen reiche Information über ihr Bildungssystem. Es ist außerdem zum Ausgangspunkt für zahlreiche nationale Forschungsprojekte geworden.

Das Programm könnte durch folgende Veränderungen in Zukunft noch besser werden:

- eine Verschiebung der Testziele von PISA von „Schnappschüssen“ des jeweiligen nationalen Leistungsstandes hin zu einer detailreicheren Interpretation der Stärken und Schwächen der einzelnen Länder im Hinblick auf die Förderung lebenslangen Lernens ihrer Bevölkerung;
- eine Rückbesinnung auf die Frage, wie Schulen und Bildungssysteme ihre Leistungen steigern und für mehr Bildungsgerechtigkeit sorgen können;
- eine Ausweitung des Testbereichs, beispielsweise durch Einbeziehung der Kenntnisse in Geographie und Sozialkunde;
- die Beteiligung von Lehrkräften bei Design und Entwicklung der nächsten PISA-Runden;
- die Einbeziehung von Lehrerinformationen, um die Daten stärker zu kontextualisieren;
- eine Ausweitung der Methodologie durch Implementierung einer Längsschnittkomponente;
- die Umgestaltung der wenigen Tabellen, die derzeit noch nach Rangfolge geordnet sind, in eine alphabetische Reihenfolge.

Solche Veränderungen bei PISA würden viele Vorteile haben. Dazu zählen größere Unterstützung – und Nutzung – der PISA-Daten durch Lehrkräfte und die Entwicklung eines fachkundigeren öffentlichen Verständnisses für das Bildungssystem, das sich nicht mehr so leicht von sensationellen Meldungen beeindrucken ließe. Ein solches Verständnis würde zu einer breiteren Vorstellung von Bildung beitragen und die Aufmerksamkeit auf Themen wie Bildungsqualität und Bildungsgerechtigkeit lenken; daraus könnten viele Konsequenzen sowohl in der Bildungspolitik als auch in der Schulpraxis gezogen werden. Dafür lohnt sich der Einsatz!

Literaturangaben

- Blum, A./Goldstein, H./Guerin-Pace, F. (2001): International Adult Literacy Survey (IALS): An Analysis of International Comparisons of Adult Literacy. In: *Assessment in Education* 8, H. 2, S. 225-246.
- Bracey, G.W. (2005): Research: Put out over PISA – Assessing for Assessment. In: *Phi Delta Kappan* 86, H. 10, S. 797f.
- Brown, P./Halsey, A.H./Lauder, H./Stuart Wells, A. (1997): The Transformation of Education and Society: An Introduction. In: Dies. (Hrsg.): *Education – Culture – Economy – Society*. Oxford: University Press.
- Burkhardt, H. (2005): Assessment Instruments and Learning Goals: PISA, TIMSS and Other Tests. Vortrag beim Royal Statistical Society Meeting in London.
- Department for Children, Schools and Families (2005): Seminar am 21. März 2005 im Church House, London.
- Education International (2007): Testing, Ranking, Reforming: Impact of PISA 2006 on The Education Policy Debate. Online-Publikation.
- Education International (2009): PISA: Is Testing Dangerous? URL: <http://www.ei-ie.org/en/article>; Zugriffsdatum: 24.08.2009.
- Goldstein, H. (2004): International Comparisons of Student Attainment: Some Issues Arising from the PISA Study. In: *Assessment in Education* 11, H. 3, S. 319-330.
- Grek, S. (2009): Governing by Numbers: The PISA ‘Effect’ in Europe. In: *Journal of Education Policy* 24, H. 2, S. 23-37.
- Hopmann, S./Brinek, G./Retzl, M. (2007): PISA zufolge PISA. Hält PISA, was es verspricht? Wien: Lit-Verlag. URL: <http://bildungswissenschaft.univie.ac.at/fe2/projekte/pisa-zufolge-pisa>; Zugriffsdatum: 24.08.2009.
- Hörmann, B. (2007): Disappearing Students: PISA and Students with Disability. In: Hopmann, S./Brinek, G./Retzl, M.: *PISA zufolge PISA. Hält PISA, was es verspricht?* Wien: Lit-Verlag, S. 157-174.
- Huisken, F. (2005): Der „Pisa-Schock“ und seine Bewältigung. – Wieviel Dummheit braucht/verträgt die Republik? Hamburg: VSA-Verlag.
- Jahnke, T. (2007): Deutsche Pisa-Folgen. In: Hopmann, S./Brinek, G./Retzl, M.: *PISA zufolge PISA. Hält PISA, was es verspricht?* Wien: Lit-Verlag, S. 305-320.
- Mansell, W. (2007): *Education by Numbers*. London: Politico’s.
- Marshall, B. (2004): Testing, Testing, Testing. In: Wragg, T. (Hrsg.): *Letters to the Prime Minister*. London: New Vision Group.
- Micklewright, J./Schnepf, S.V. (2006): Response Bias in England in PISA 2006 and 2003. Department for Education and Skills: Research Report 771. Online-Artikel.
- Mortimore, P. (1998): *The Road to Improvement. Reflection on School Effectiveness*. Lisse: Swets & Zeitlinger.
- Mortimore, P. (2008): Lies, Damned Lies and Sats-Based League Tables. In: *Education Guardian* vom 05.08.2008.
- Mortimore, P. (2009): School League Tables and Testing. An Interview by John Graham. In: *Professional Voice* 6, H. 3, S. 57-61.
- OECD (2007a): *No More Failures: Ten Steps to Equity in Education*. Paris: OECD.
- OECD (2007b): *PISA 2006: Science Competencies for Tomorrow’s World*. Paris: OECD.
- PISA (2006a): Technical Report. URL: www.oecd.org/dataoecd/0/47/42025182.pdf; Zugriffsdatum: 24.08.2009.
- PISA (2006b): *Science Competencies for Tomorrow’s World. Vol. 1: Analysis*. URL: www.oecd.org/dataoecd/15/13/39725224.pdf; Zugriffsdatum: 24.08.2009.
- Prais, S. (2003): Cautions on OECD’s Recent Educational Survey (PISA). In: *Oxford Review of Education* 29, H. 2, S. 139-163.

- Prais, S. (2004): Rejoinder to OECD's Response. In: Oxford Review of Education 30, H. 4, S. 569-573.
- Smithers, A. (2004): England's Education: What Can Be Learned by Comparing Countries? University of Liverpool: Centre for Education and Employment Research.
- Wuttke, J. (2006): Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung. In: Jahnke, T./Meyerhöfer, W. (Hrsg.): PISA & Co.: Kritik eines Programms. Hildesheim/Berlin: Franzbecker, S. 101-1