

*Stefan T Hopmann / Gertrude Brinek / Martin Retzl (Hrsg.)*

**PISA zufolge PISA  
PISA According to PISA**

**Wien / Berlin (LIT-Verlag) 2007**  
ISBN 978-3-7000-0771-5 (Österreich)  
ISBN 978-3-8258-0946-1 (Deutschland)

Keine andere Bildungsforschung hat in den letzten Jahren soviel Furore gemacht wie PISA. In vielen Ländern, besonders aber in Österreich und Deutschland, scheinen viele in Politik und Öffentlichkeit anzunehmen, PISA sei eine unabhängige wissenschaftliche Bewertung der Leistungsfähigkeit bzw. der Schwächen nationaler Bildungssysteme und könne nachweisen, woran das jeweilige Schulwesen krankt. Dabei wird oft übersehen, dass nach dem Willen des Auftragsgebers OECD die PISA-Studien in erster Linie politischen und nicht wissenschaftlichen Zwecken dienen, und dass PISA ausdrücklich nicht darauf aufbaut, ob das, was gemessen wird, in der Schule überhaupt gelernt werden kann oder soll.

In den letzten Jahren ist vereinzelt wissenschaftliche Kritik am Design und an der Durchführung der PISA-Studien laut geworden. Bislang ist solche Kritik fast nie über den jeweiligen nationalen Rahmen hinaus gedrungen und konnte deshalb leicht als lokal begrenzter Einzelfall ohne Bedeutung für das Ganze abgetan werden. Eine umfassende Erwiderung auf diese Kritik hat es von Seiten der PISA-Forschung bis jetzt nicht gegeben, aber auch keine Richtigstellung der chronischen Irreführung der Öffentlichkeit.

Der vorliegende Band enthält wissenschaftliche Beiträge aus sieben europäischen Ländern, die unabhängig Leistungen und Grenzen des PISA-Projekts analysieren. Gefragt wird, ob PISA methodologisch den wissenschaftlichen Ansprüchen an international vergleichende Forschung gerecht wird. Untersucht werden unter anderem Stichproben, auf die sich PISA stützt, Fragen, die PISA stellt, sowie die Auswertung und schliesslich die Präsentation der Ergebnisse durch PISA selbst.

Ungeachtet der Verdienste der PISA-Forschung ist das Ergebnis ernüchternd. Es lässt sich in zwei Hauptpunkten zusammenfassen:

1. Was PISA misst, ist keineswegs repräsentativ für die jeweiligen Wissens- und Schulkulturen und ist dies in der Regel nicht einmal in den Fachgebieten, die es selbst ausdrücklich hervorhebt (wie Mathematik und Naturwissenschaften).
2. Die konkrete Durchführung ist mit so vielen Schwachstellen, Unsicherheiten und Fehlerquellen behaftet, dass die populären internationalen und nationalen (Schul-) Vergleiche wissenschaftlich unhaltbar sind.

Zusammen genommen ergibt sich, dass PISA im besten Fall eine interessante Jugendstudie über die mehr oder weniger zufällige Verteilung eines sehr testspezifischen Wissens in den beteiligten Ländern ist. Wissenschaftliche Vergleiche der Leistungsfähigkeit der jeweiligen Bildungssysteme oder gar einzelner Schulformen und Schulen lassen sich so nicht begründen.

**PISA hält nicht, was PISA verspricht – und kann das in seinem Rahmen auch nie leisten!**

Repräsentanten der PISA-Forschung in Österreich und Deutschland waren eingeladen, in diesem Band zur vorgebrachten Kritik Stellung zu nehmen. Leider hat dieser Einladung niemand Folge geleistet, vielleicht auch deshalb, weil selbst bei Widerlegung einzelner Punkte die öffentliche Aufmerksamkeit sichernden nationalen und internationalen Rankings nicht zu retten gewesen wären.

**Für Rückfragen: +43-650-7722602 oder stefan.hopmann@univie.ac.at**

**Im Einzelnen behandelt das Buch die folgenden Themen:**

**Stefan T. Hopmann, Gertrude Brinek (Österreich):**

**Introduction. PISA According to PISA – Does PISA keep what it promises?**

Seit Veröffentlichung der ersten PISA – Ergebnisse haben sich auch einige kritische Stimmen zur OECD-Studie zu Wort gemeldet. Allerdings blieb eine eingehende methodologische Debatte bisher aus. Das Buch sammelt daher Beiträge zur methodologischen Kritik an PISA aus den unterschiedlichsten Bereichen und Ländern. Als Ergebnis muss festgestellt werden, dass insbesondere die beliebten nationalen und internationalen PISA-Vergleiche mit so vielen Unsicherheiten und Fehlerquellen behaftet sind, dass sie wissenschaftlich gesehen so gut wie unbrauchbar sind. Abschliessend wird der Umgang der PISA-Konsortien mit Kritik diskutiert und mit dem Contergan-Skandal vor fünfzig Jahren verglichen.

**Antoine Bodin (Frankreich):**

**What does PISA really assess? What it doesn't? A French view**

Der Beitrag konzentriert sich auf die Analyse der externen Validität der Mathematik-Items. Indem PISA-Fragen französischen Assessment-Items gegenübergestellt werden, sollen Überschneidungen, kognitive Voraussetzungen und Unterschiede in den Kompetenzlevels veranschaulicht werden. Dabei zeigt sich, dass PISA und die französischen Tests zum Teil Unterschiedliches messen, und PISA damit nicht den dort angestrebten Zielen entspricht. Hervorgehoben wird aber auch der Beitrag von PISA zur Weiterentwicklung der Forschung.

**Wolfram Meyerhöfer (Deutschland):**

**Testfähigkeit – Was ist das?**

Das Problem der Testfähigkeit wird am Beispiel mathematischer Leistungstests erkundet. „Testfähigkeit“ beschreibt jene Kenntnisse, Fähigkeiten und Fertigkeiten, die in einem Test wie PISA miterfasst bzw. mitgemessen werden, die aber nicht unter den Begriff „mathematische Leistungsfähigkeit“ zu fassen sind. Anhand von Aufgaben aus TIMSS und PISA wird mit Hilfe von didaktischen und objektiv-hermeneutischen Aufgabeninterpretationen herausgearbeitet, welche empirischen Phänomene das Problem der Testfähigkeit ausmachen. Es zeigt sich, dass die für PISA erforderliche Testfähigkeit mit der in Lehrplan und Fachdidaktik gewünschten mathematischen Bildung wenig gemein hat.

**Jens Dolin (Dänemark):**

**PISA – an Example of the Use and Misuse of Large-Scale Comparative Tests**

Der Beitrag zeigt, dass die Entscheidung für oder gegen einen bestimmten technischen Aspekt oder ein bestimmtes Verfahren unvermeidlich Teil des zugrunde liegenden Wertesystems einer Evaluationsstudie ist. PISA ist also kein neutraler Test objektiver Fähigkeiten, sondern beinhaltet die Entscheidung für einen bestimmten, wenigstens in Dänemark sehr umstrittenen Typus von Bildung. Gleichzeitig weist der Autor darauf hin, dass PISA einerseits ein großes Potential in sich birgt, das es weiter zu entwickeln gilt, andererseits eine Studie dieser Art niemals fähig sein kann, ein Bildungssystem in seiner ganzen Fülle zu beschreiben.

**Markus Puchhammer (Österreich):**

**Language-Based Item Analysis – Problems in Intercultural Comparisons**

Puchhammer unterzieht die verfügbaren Mathematik-Testaufgaben einer quantitativen Analyse, die sprachliche Schwierigkeiten bei PISA zu Tage bringt. Unter anderem wird festgestellt, dass unterschiedliche Übersetzungen der Testitems verschiedene Längen der Texte zur Folge hat, was allein erhebliche Ergebnisunterschiede verursachen kann. Auch die Schwierigkeit der übersetzten Begriffe wird analysiert. Es stellt sich heraus, dass manche Wörter in anderen Sprachen schwieriger zu erfassen sind, d.h. der Schwierigkeitsgrad der Aufgaben ist nicht für jede Sprache und jede Wissens-

kultur gleich. Damit scheitert der Anspruch von PISA, faire internationale und damit sprachübergreifende Vergleiche zu ermöglichen.

**SJ Prais (England):**

**England: Poor Survey Response and No Sampling of Teaching Groups**

Die PISA – Ergebnisse von 2003 hatten für England keine Aussagekraft, da die vorgegebenen Teilnahmequoten verfehlt wurden. Der Autor kritisiert an PISA und TIMSS, dass sie unkoordiniert durchgeführt werden, einander beachtlich überschneiden und überdies kostspielig und unwirtschaftlich sind. Da SchülerInnen mit geringem Leistungsniveau eher dazu tendieren, den Testungen fernzubleiben, werden Testergebnisse zudem tendenziell nach oben verzerrt. Neben der Kritik an der Repräsentativität des Samples wird auch bemängelt, dass in der Zusammenstellung des Samples Klassen oder Unterrichtsgruppen inadäquat gewichtet wurden.

**Bernadette Hörmann (Österreich):**

**Disappearing Students. PISA and Students With Disabilities**

Die Autorin bringt das Problem der systematischen Marginalisierung von Kindern mit Behinderung zur Sprache, das auch im Kontext von PISA tragend wird. Die Tests ergeben ein verzerrtes Bild, wenn Kinder mit Behinderungen (gleich welcher Art) ungleich berücksichtigt werden. Es wird dabei gezeigt, wie wenig dieses Problem in Europa und auch in Österreich überhaupt beachtet wird, obwohl es durchaus Möglichkeiten und Wege gäbe, Kinder mit speziellen Bedürfnissen in die Testung mit einzubeziehen und ihre Leistungen adäquat zu messen.

**Peter Allerup (Dänemark):**

**Identification of Group Differences Using PISA Scales - Considering Effects of Inhomogeneous Items**

Der Autor beschäftigt sich mit der Inhomogenität der Fragen in PISA. Es wird beispielsweise gezeigt, dass die relative Schwierigkeit der Mathematik - Testitems für Mädchen und Buben nicht gleich ist, d.h. dass die PISA – Skalen nicht homogen und direkte Vergleiche damit irreführend sind. Auch in Bezug auf die kulturellen Hintergrund (Ethnizität) der getesteten Kinder wird festgestellt, dass die PISA-Skalen bestehende Unterschiede nicht korrekt messen können bzw. vermutlich weit überschätzen. Deswegen sind Folgerungen über sozial bedingte Leistungsunterschiede auf Grundlage von PISA höchst anfechtbar.

**Sven Sjøberg (Norwegen):**

**PISA and "Real Life Challenges": Mission impossible?**

Der Beitrag zeigt, dass der Anspruch von PISA, „Real-Life-Skills und Kompetenzen in authentischen Kontexten“ zu testen, fragwürdig ist. Ein Test kann nie besser sein als die Items, aus denen er sich zusammensetzt. Demzufolge wäre es nötig, die Items der Testung und deren Anwendung genau darauf hin zu überprüfen, was aber weder durch PISA selbst belegt ist, noch anderen durch die Geheimhaltung der Testaufgaben möglich ist. Ob PISA überhaupt lebensweltlich relevante Kenntnisse misst, ist also weder belegt, noch nachprüfbar. Die Aufgaben, die bekannt geworden sind, lassen daran zweifeln. Eine weitere Kritik betrifft die PISA zugrunde liegende Annahme, dass die teilnehmenden Schülerinnen und Schüler ihr Bestes im Test zeigen. Studien zufolge sind Motivation und Testwilligkeit in verschiedenen Ländern äusserst ungleich. Ein erheblicher Teil der Leistungsunterschiede könnte also daran liegen, dass PISA in manchen Ländern für viele Befragte keinen „authentischen“ Wert hat.

**Gjert Langfeldt (Norwegen):**

**PISA – Undressing the Truth or Dressing up a Will to Govern?**

Hier werden die Testgütekriterien Validität und Reliabilität von PISA genauer in den Blick genommen. Dabei wird zunächst die Qualität des Samples von PISA kritisiert (z.B. Zusammensetzung der Populationen, Auswahl der teilnehmenden Schulen etc.). Die

Schwachstellen in der Validität der Studie liegen in der systematischen Verleugnung der Diversität der Kinder (etwa der kulturabhängigen Verzerrungen der Testergebnisse) und in der Herkunft der Testitems: Sie sind nicht repräsentativ für die unterschiedlichen Wissenskulturen, sondern der weitaus grösste Teil stammt aus englischsprachigen Ländern. So gesehen ist wenig überraschend, dass vier der sechs besten Teilnahmestaaten englischsprachige Länder sind. In beiden Bereichen entspricht PISA nicht den üblichen Standards vergleichender Forschung. Das entscheidende Problem liegt aber darin, dass diese Hintergründe für Laien nicht sichtbar sind, und daher -guten Glaubens - auf Grundlage von PISA etwa politische Entscheidungen getroffen werden.

**Joachim Wuttke (Deutschland):**

**Uncertainties and Bias in PISA**

Der Autor unterzieht PISA einer statistischen Kritik. Die Kritikpunkte betreffen unter anderem das Sampling (etwa die in den Teilnehmerstaaten sehr unterschiedliche Einbeziehung von Kindern mit Lernbehinderungen, die unterschiedlichen bzw. teils unzureichenden Teilnahmequoten, die kulturellen Unterschiede bezüglich Antwortstrategien u.v.m.) und die Auswahl und Form der Items (u.a. das fragwürdige Auswahlverfahren für die Testitems, die Verwendung des Rasch-Modells). Im Ergebnis sind für den Autor die von PISA vorgelegten Vergleiche methodisch unhaltbar. Freilich könnte nach Meinung des Autors selbst eine Behebung der angeführten Mängel die testtechnischen Grenzen der PISA-Studie nicht aufheben.

**Rolf V. Olsen (Norwegen):**

**Large-Scale International Comparative Achievement Studies in Education:  
Their Primary Purposes and Beyond**

Hier plädiert der Autor für eine Ausweitung und Ausdehnung der PISA-Studie, um das bestehende Konzept sinnvoll ausschöpfen und davon ausreichend profitieren zu können. Es wird vorgeschlagen, die umfangreichen und aufwändig erhobenen Daten der Studie ausführlichen Sekundäranalysen zu unterziehen, um in diesem Rahmen einerseits zusätzliche Erhebungen durchführen zu können und andererseits den gesamten Forschungszugang zu erweitern. Zu diesem Zweck werden sechs konkrete Beispiele vorgeschlagen.

**Michael Uljens (Finnland):**

**The Hidden Curriculum of PISA -**

**The Promotion of Neo-liberal Policy by Educational Assessment**

Das Phänomen PISA wird als ein Beispiel für einen grundsätzlichen, nahezu alle europäischen Staaten betreffenden Trend zur Transformation der Bildungspolitik angesehen. Für Finnland im Speziellen bedeutete dies, dass PISA in etwa dem entsprach, was die Wendung zur sogenannten Bildungspolitik „der dritten Republik“ ohnedies anstrebte. Es wird im weiteren auf die auch in Finnland verbreitete Kritik vor allem an den Intentionen, theoretischen Fundierungen und methodologischen Aspekten des PISA-Projektes verwiesen, aber auch gezeigt, wie diese bei den finnischen Erklärungsversuchen für den unerwarteten Erfolg Finnlands bei PISA unberücksichtigt blieb.

**Thomas Jahnke (Deutschland):**

**Deutsche Pisa-Folgen**

In diesem Beitrag werden die Folgen von PISA in Deutschland, vor allem in Bezug auf die Beschlüsse der deutschen Kultusministerkonferenz zum Bildungsmonitoring und zu den nationalen Bildungsstandards in Mathematik thematisiert. Unter Einbezug amerikanischer Forschung wird die Einführung eines umfassenden Bildungsmonitorings in Deutschland als Gefahr der Etablierung einer „Teaching-to-the-Test“-Kultur kritisiert. Zudem werden beispielhaft einzelne, am PISA-Vorbild orientierte Aufgaben der Bildungsstandardtests einer Analyse unterzogen. Es zeigt sich, dass der Großteil der Aufgaben didaktisch unsachgemäß und zum Teil direkt fehlerhaft konstruiert ist. Allerdings sei fraglich, so der Autor, ob selbst eine fundamentale methodologische Kritik die

Geltungsmacht von PISA in Deutschland und deren Einfluss auf die gegenwärtige Bildungspolitik erschüttern könne.

**Dominik Bozkurt, Gertrude Brinek, Martin Retzl (Österreich):  
PISA in Österreich: Mediale Reaktionen, öffentliche Bewertungen  
und politische Konsequenzen**

Mithilfe einer Medienanalyse werden die öffentlichen und politischen Reaktionen in Österreich auf die PISA-Ergebnisse aus den Jahren 2000 und 2003 dargestellt. Dabei stellt sich heraus, dass sich die Reaktionen aus den beiden Jahren sowohl inhaltlich als auch quantitativ erheblich voneinander unterscheiden. Außerdem zeigt sich, dass ungeachtet dessen, was PISA eigentlich zu leisten vermag, höchst widersprüchliche Schlussfolgerungen und politischen Absichten mit PISA untermauert werden. Es scheint so, als könne jeder aus Politik, Wirtschaft, Kultur, Bildung etc. seine Überzeugungen in der PISA-Studie bestätigt finden.

**Stefan T. Hopmann (Österreich):  
Epilogue: No Child, No School, No State Left Behind:  
Comparative Research in the Age of Accountability**

Im Schlusskapitel wird die Frage gestellt, warum und unter welchen Bedingungen das Projekt PISA Erfolg haben konnte. Wie kommt es, dass in manchen Ländern die gesamte Bildungspolitik in den „schiefen“ Schatten von PISA gerät, während PISA andernorts nur eine Stimme unter vielen ist? Es zeigt sich, dass es für die Leistungen und Schwächen des PISA-Projektes gute historische und aktuelle Gründe gibt. PISA hat – wie es scheint – den meisten Erfolg mit dem, wozu es am wenigsten taugt, nämlich mit seinen methodisch unhaltbaren nationalen und internationalen Vergleichen der Leistungsfähigkeit der jeweiligen Bildungssysteme.

Speziell Allerup, Langfeldt, Wuttke und Prais zeigen in ihren Beiträgen durch eindrucksvolle statistische Analysen, dass das PISA-Sampling durch bspw.

- unterschiedliche teils unzureichende Teilnahmequoten,
- die Auswahl der teilnehmenden Schulen,
- die Zusammensetzung der Populationen,
- die unterschiedliche Einbeziehung von Kindern mit Lernbehinderungen,
- die kulturellen Unterschiede bezüglich Antwortstrategien,
- die unzureichende Repräsentativität oder inadäquate Gewichtung von Klassen und Unterrichtsgruppen

erhebliche Schwachstellen aufweist.

Durch weitere methodologische Analysen erläutern die genannten Autoren, dass die Auswahl und Form der Testitems wie 1) deren Herkunft (größtenteils aus dem angelsächsischen Sprachraum) oder 2) die systematische Verleugnung der Diversität der Kinder durch kulturabhängige Verzerrungen der Testergebnisse die Validität der Studie stark beeinträchtigen. Es werden dabei die üblichen Standards vergleichender Forschung nicht eingehalten. Dies ist jedoch für Laien nicht ersichtlich, wodurch - guten Glaubens – auf Grundlage von PISA politische Entscheidungen getroffen werden.

Ein weiterer wichtiger Aspekt, den vor allem Allerup thematisiert, ist, dass die Schwierigkeit der Mathematik-Testitems für Mädchen und Buben aber auch für Jugendliche unterschiedlichen kulturellen Hintergrunds unterschiedlich ist – die Items also inhomogen sind. Leistungsunterschiede sind oft dadurch zu erklären - Leistungsvergleiche daher oft irreführend und alle Folgerungen über Ursachen der Leistungsunterschiede vor dem Hintergrund dieser Tatsache gehen über reine Spekulation kaum hinaus.

Warum sind nun die genannten Schwachstellen so erheblich und bedeutend?

Weil eindrucksvoll belegt wird, dass die von PISA selbst gesteckten Ziele und Ansprüche wie:

- Rückmeldung über die Qualität des Bildungssystems und der Schulen zu geben
- einen Vergleich der Leistungsfähigkeit der Bildungssysteme (league tables) zu ermöglichen
- eine Grundlage für die Politik darzustellen, aus der sinnvolle Maßnahmen und Reformen für das Schulsystem abgeleitet werden können
- Kommunikationsfähigkeit oder die notwendigen Fähigkeiten für lebenslanges Lernen zu messen

durch das von PISA verwendete Werkzeug, die angewandte Methodik niemals erreicht und nicht eingehalten werden können und somit PISA auch methodisch nie halten kann, was es verspricht.

## **Inhalt/Table of content**

Zu diesem Buch	1
Vorwort <i>Richard Olechowski</i>	5
Introduction: PISA According to PISA – Does PISA Keep What It Promises? <i>Stefan T. Hopmann/Gertrude Brinek</i>	9
What Does PISA Really Assess? What Does It Not? A French View <i>Antoine Bodin</i>	21
Testfähigkeit – Was ist das? <i>Wolfram Meyerhöfer</i>	57
PISA – An Example of the Use and Misuse of Large-Scale Comparative Tests <i>Jens Dolin</i>	93
Language-Based Item Analysis – Problems in Intercultural Comparisons <i>Markus Puchhammer</i>	127
England: Poor Survey Response and No Sampling of Teaching Groups <i>S J Prais</i>	139
Disappearing Students PISA and Students With Disabilities <i>Bernadette Hörmann</i>	157
Identification of Group Differences Using PISA Scales – Considering Effects of Inhomogeneous Items <i>Peter Allerup</i>	175
PISA and “Real Life Challenges”: Mission Impossible? <i>Svein Sjøberg</i>	203

PISA – Undressing the Truth or Dressing Up a Will to Govern? <i>Gjert Langfeldt</i>	225
Uncertainties and Bias in PISA <i>Joachim Wuttke</i>	241
Large-Scale International Comparative Achievement Studies in Education: Their Primary Purposes and Beyond <i>Rolf V. Olsen</i>	265
The Hidden Curriculum of PISA – The Promotion of Neo-Liberal Policy By Educational Assessment <i>Michael Uljens</i>	295
Deutsche Pisa-Folgen <i>Thomas Jahnke</i>	305
PISA in Österreich: Mediale Reaktionen, öffentliche Bewertungen und politische Konsequenzen <i>Dominik Bozkurt, Gertrude Brinek, Martin Retzl</i>	321
Epilogue: No Child, No School, No State Left Behind: Comparative Research in the Age of Accountability <i>Stefan T. Hopmann</i>	363

## **Zu diesem Buch**

„PISA zufolge PISA“ (PISA According to PISA) war Thema eines Symposiums, das im März 2007 an der Universität Wien von der Forschungseinheit für Schul- und Bildungsforschung des Instituts für Bildungswissenschaft durchgeführt wurde. Zu dieser Veranstaltung hatten wir einige Kritiker der vorliegenden PISA-Studien eingeladen, aber auch Vertreter des österreichischen PISA-Konsortiums, die jedoch leider kurzfristig absagen mussten. Unser Ziel war eine Versachlichung der Debatte über PISA, weg vom politisch-ideologischen Streit über PISA und hin zu einer Analyse der methodologischen Voraussetzungen und Folgen des PISA-Projektes, genauer gesagt der Frage: Ob PISA bei gegebenem Design halten kann, was es in seinen Analysen und Berichten zu erklären verspricht. Bei dieser Veranstaltung wurde die Frage gestellt, ob und wie PISA international als wissenschaftliches Vorhaben diskutiert wird. Wir haben dies zum Anlass genommen, den vorliegenden Band zu gestalten.

Trotz seiner enormen Breitenwirkung hat PISA in der vergleichenden Bildungsforschung bislang kaum internationale, wohl aber eine Reihe von nationalen Nachfragen ausgelöst (vgl. Hopmann & Brinek in diesem Band). Für diesen Band haben sich nun erstmals länderübergreifend achtzehn ForscherInnen aus sieben Ländern (Dänemark, Deutschland, England, Finnland, Frankreich, Norwegen und Österreich) zusammengefunden, um PISA kritisch von allen Seiten unter die Lupe zu nehmen, und zwar den gesamten Forschungsprozess vom Design und den Erhebungsinstrumenten über die Durchführung und Datenanalyse bis hin zur öffentlichen Präsentation der Daten. Berücksichtigt wurden dabei alle einschlägig relevanten wissenschaftlichen Zugänge zu PISA: empirische Bildungsforschung, Forschungsmethodologie, Statistik, die allgemeine und die einschlägigen Fachdidaktiken. Fast alle Beteiligten verfügen über langjährige Erfahrungen in der vergleichenden Bildungsforschung oder verwandten Unternehmungen, einige waren zumindest zeitweise auch direkt an PISA-Forschungen beteiligt.

Bei allem Respekt für das grossartige Engagement der OECD und der nationalen PISA-Konsortien, fällt das Ergebnis sehr ernüchternd aus: PISA hält nicht annähernd, was PISA verspricht, und kann das mit den angewandten

Mitteln auch gar nicht leisten! Das PISA-Projekt ist offenkundig mit so vielen Schwachstellen und Fehlerquellen belastet, dass sich zumindest die populärsten Endprodukte, die internationalen Vergleichstabellen sowie die meisten nationalen Zusatzanalysen zu Schulen und Schulstrukturen, Unterricht, Schulleistungen und Problemen wie Migration, sozialer Hintergrund, Geschlecht usw., in den bisher praktizierten Formen wissenschaftlich schlicht nicht aufrecht erhalten lassen. Sie überspannen bei weitem die Tragfähigkeit des gewählten Designs und dessen theoretische und methodische Grundlagen. Wer auf dieser Grundlage über Schulstrukturen, Lehrpläne, nationale Tests oder die zukünftige Lehrerbildung befinden will, ist nicht gut beraten.

Damit hört PISA nicht auf, eines der wichtigsten und ertragreichsten Projekte der vergleichenden Forschung der Gegenwart zu sein. Einzelne Beiträge in diesem Band weisen dazu ausdrücklich auf künftige Möglichkeiten der PISA-Forschung hin. Nur scheint dringend geboten, die zum Teil bei solcher Forschung unvermeidlichen Grenzen der Geltung und Zuverlässigkeit weitaus deutlicher auszuweisen und dafür zu sorgen, dass nicht auch künftig PISA für Beweislasten in Anspruch genommen wird, die es auf wissenschaftlich vertretbare Weise nicht schultern kann. Man kann fast sagen, es gilt das Gute an PISA und die interessierte Öffentlichkeit gegen den methodologisch haltlosen Überschwang einiger der an PISA Beteiligten in Schutz zu nehmen. Sonst droht die Gefahr, dass eines Tages die Bildungsverwaltungen, die Schulen und Schulleitungen, die Lehrkräfte und die Schülerinnen und Schüler nicht nur des stetigen Missbrauchs ihrer Daten überdrüssig werden, sondern gleich alle vergleichbaren Maßnahmen und Forschungsvorhaben in Bausch und Bogen ablehnen oder gar – wie das mit staatlichen Tests in einigen Ländern (u.a. in den USA, Chile, Norwegen) schon passiert ist – boykottieren oder durch mutwilliges Antwortverhalten beschädigen. Dies würde der vergleichenden Bildungsforschung als Ganzes einen nachhaltigen Schaden zufügen, und die Sorge, dass es dazu am Ende der PISA-Begeisterung kommen kann, begründet unser Engagement.

Natürlich waren auch unserem Vorhaben deutliche Grenzen gesetzt:

- Zum einen war eine direkte Re-Analyse von PISA-Originaldaten, PISA-Fragen etc. nur im begrenzten Masse möglich, nur dort wo einzelnen Datenbestände zugänglich waren. Zusätzlich haben wir fast die gesamte Literatur zur PISA-Methodologie und ihren Implikationen ausgewertet (vgl. die Literaturangaben in den einzelnen Kapiteln). PISA lässt eine unabhängige Überprüfung der vollständigen Datensätze einschliesslich aller Unterlagen bislang jedoch nicht zu. Es mag also sein, dass sich bei einer entsprechenden

Nachlese – wenn sie eines Tages möglich sein wird – einzelne Ergebnisse unserer Metaanalysen anders darstellen, als es uns nachzuprüfen möglich war. Allerdings haben sich so viele kritische Einwände in unseren Untersuchungen ergeben, dass die Widerlegung eines halben oder ganzen Dutzends von ihnen an den Kernaussagen dieses Bandes nichts ändern würde. In jeder Phase des PISA-Projektes gibt es zahlreiche Designentscheidungen und -probleme, die für sich allein genommen ausreichen, einen erheblichen Teil der gegenwärtig üblichen Darstellung und Nutzung der PISA-Ergebnisse für wissenschaftlich nicht tragfähig zu halten.

- Zum zweiten war uns nicht daran gelegen, mit einer Stimme zu reden. Nicht nur, dass auch PISA-Kritiker Unterschiedliches kritikwürdig finden, und deshalb verschiedene Zugänge und Argumentationen wählen. Wir wollten die ganze Bandbreite der zur Zeit in Europa zugänglichen Kritik präsentieren und niemanden ausschliessen, nur weil der eine oder die andere eventuell einzelne Punkte oder Schlussfolgerungen nicht teilen. Wir hatten auch das deutsche und das österreichische PISA-Konsortium mehrfach zur Mitwirkung eingeladen. Leider ist diese nicht zustande gekommen. Zum Glück waren einige andere mit PISA-Erfahrung dennoch bereit, an unserem Vorhaben teilzunehmen. Damit gelingt uns aber nur teilweise, das gesamte Für und Wider der Diskussion widerzuspiegeln. Wir zweifeln jedoch nicht daran, dass die PISA-Konsortien genügend andere Möglichkeiten haben, sich aktiv an der Debatte zu beteiligen.

Ein solches Vorhaben wie das vorliegende kann ohne vielfältige Hilfe nicht gelingen. Das seinerzeitige Österreichische Bundesministerium für Bildung, Wissenschaft und Kultur (BMBWK), die Österreichische Gesellschaft für Bildungsforschung sowie der norwegische Forschungsverbund „Achieving School Accountability in Practice (ASAP)“, zu dessen Veröffentlichungen auch dieser Band zählt, haben das Symposium und die Arbeiten am vorliegenden Band großzügig unterstützt. Nicht zu vergessen ist die Hilfe durch die Sekretariate in Wien (Patricia Stuhr) und in Kristiansand (Inger Linn Nystad Baade, Karen Beth Lee Hansen), die uns durch Geduld und Sprachfertigkeit ein Gelingen ermöglichten. Schliesslich möchten wir Richard Olechowski und dem LIT-Verlag für die Aufnahme in die Reihe „Schulpädagogik und Pädagogische Psychologie“ herzlich danken.

Stefan T. Hopmann, Gertrude Brinek, Martin Retzl  
Wien, im September 2007

Wissenschaft lebt von der Diskussion. Aus diesem Grund möchten wir Sie herzlich einladen, an unserem Online-Diskussionsforum teilzunehmen. Posten Sie Ihre Meinung, Kritik und Anregungen zum Buch! Nähere Informationen dazu sind auf folgender Homepage verfügbar:

<http://institut.erz.univie.ac.at/home/fe2/>.

Wir freuen uns auf eine anregende Diskussion!

## **Introduction:**

# **PISA According to PISA – Does PISA Keep What It Promises?**

Stefan T. Hopmann/Gertrude Brinek

*Austria: University of Vienna*

For the time being, PISA is the most successful enterprise in comparative education. Every time a new PISA wave rolls in, or an additional analysis appears, governments fear the results, newspapers fill column after column, and the public demands answers to the claimed failings in their country's school system. Of course, such a tremendous impact evokes discussions and criticism. On the one side are those:

- who blame PISA for not covering the whole breath of education or schooling (e.g. Fuchs 2003; Ladenthin 2004; Kraus 2005; Herrmann 2005; Dohn 2007; adding to the PISA frame: Benner 2002),
- who point to the fact that PISA is run by private companies (“PISA Incorporated”) looking for a share of the ever-growing testing market (see e.g. Bracey 2005; Flitner 2006; Lohmann 2006), or
- who depict PISA as a New Public Government outlet of the most neo-liberal kind (see e.g. Lohmann 2001; Huisken 2005; Klausnitzer 2006).

On the other side are those who praise PISA for giving us the best data base ever available for comparative research, for developing new tools of research, and for PISA's creative analysis of its data sets (for many examples see Pekrun 2003; Roeder 2003; Weigel 2004; Stack 2006; Olsen in this volume).

## **PISA According to PISA**

However, surprisingly, and in spite of its public impact, PISA has not lead to thorough methodological debates within the comparative research community,

at least not internationally. There have been some critiques pointing to design or analytic short-comings in some of the participating countries (e.g. Bonnet 2002; Romainville 2002; Nash 2003; Prais 2003, 2004; Goldstein 2004; Allerup 2005, 2006; Bodin 2005; Bottani & Virgnaud 2005; Gaeth 2005; Olsen 2005; Jahnke & Meyerhöfer 2006; Neuwirth, Ponocny & Grossmann 2006; Grisay & Monseur 2007). There has been some fundamental, and highly contested criticism of the methodological soundness of PISA's research as a whole (Jahnke & Meyerhöfer 2006; especially Wuttke 2006; rebuttal by Prenzel & Walter 2006)<sup>1</sup>. However, none of this has led to an international debate on the validity claims of PISA *outside* the PISA community itself. It seems as if the overwhelming success of the approach has led to any attempt to discuss PISA's design, data collection and analysis methodologically looking petty-minded and irreverent. The strategy of PISA itself in not giving access to the full database, including all the questionnaires, contributes to this problem.

The present volume on "PISA According to PISA" is probably the first independent international approach to discuss the methodological merits and shortcomings of PISA in relation to the validity and reliability claims PISA itself puts forward. Our aim is not to add to the debate for or against PISA. Most of us believe that PISA is an important milestone in the history of our field. But we do question if some basic elements of PISA are done well enough to carry the weight of, e.g., comparative league tables or of in-depth analyses of weaknesses of educational systems. We ask if other, and better, uses of the PISA data base are warranted, and if PISA-as-a-public-event should come under much more independent scrutiny – if only to avoid its misuse to validate claims and policies which cannot be legitimately derived from PISA.

The volume seeks to follow – as much as possible – the whole PISA research process from the design and sampling, the data collection and analysis, through to the data presentation and impact. Our aim is not to give an overview of the different national PISA debates, rather to discuss general issues of construction and use. The contributors come from seven countries and from all walks of educational research, including specialists in empirical research methodology, statistical data analysis, general and subject matter didactics, and educational policy analysis. We include contributors who are or

---

<sup>1</sup> The editors of the above mentioned "PISA & Co." volume (Jahnke & Meyerhöfer 2006) are working on a new and revised edition of that book, including an explicit discussion of the response they got to the first edition. This book will be available by late 2007.

have themselves been involved in PISA or similar projects (see the bios at the end of the book).

To highlight just a few core issues:

- *Antoine Bodin* (IREM de Besançon – Université de Franche-Comté) shows from a French perspective how much the PISA – assessment is embedded in a certain understanding of (school) knowledge, which doesn't fit all.
- *Wolfram Meyerhöfer* (Universität Potsdam) continues this argument by an in-depth-analysis of what PISA really asks for in its questionnaires, showing how little this is in touch with a comprehensive concept of “Bildung” or even current didactics.
- *Jens Dolin* (Syddansk Universitet) adds similar arguments from a Danish perspective, underlining how much PISA's conceptualization of knowledge is at risk to misrepresent what is taught and learned in schools.
- *Markus Puchhammer* (Technikum Wien) shows – using the published example questions – how translation problems may affect results to a degree making comparisons guesswork.
- *S J Prais* (National Institute of Economic and Social Research London) uses the example of England to demonstrate serious flaws in the response rates and sampling, which necessarily lead to biased results.
- *Bernadette Hörmann* (Universität Wien) points to the systematic marginalization of special needs students by PISA and to how little there has been done to deal with their role within the PISA approach at least in Austria.
- *Peter Allerup* (Århus Universitet) elaborates a similar issue by showing from Denmark to what degree PISA's much acclaimed analysis of the impact of gender, migration and similar factors depends on but a few, highly problematic items.
- *Svein Sjøberg* (Universitetet i Oslo) underlines how much both, PISA's design on the one hand, and the student response behavior on the other are culturally embedded, which may lead to a partial or complete mismatch.
- *Gjert Langfeldt* (Agder Universitet) questions the validity and reliability claims made by PISA, pointing to constructional constraints, methodological mishaps and the cultural bias embedded in the PISA design.
- *Joachim Wuttke* gives a comprehensive overview over recently voiced criticism of PISA's research conduct and the resulting bias and uncertainties, which put not at least its league tables and comparisons at random.
- *Rolf Olsen* (Universitetet i Oslo) outlines ways how PISA can overcome

some of its short-comings by broadening its approach and adding new research.

- *Michael Uljens* (Åbo Akademi) explains the Finnish PISA success by the fact that what PISA asks for had already gained a foothold in Finnish schooling before PISA came around.
- *Thomas Jahnke* (Universität Potsdam) elaborates from a German perspective how PISA fails to really assess what is or should be taught in schools, and how reliance on PISA can lead to an impoverished view on the curriculum.
- *Dominik Bozkurt, Gertrude Brinek* and *Martin Retzl* (Universität Wien) use the Austrian example to show how the public and political response to PISA unfolds irrespective of what PISA really can cover or prove.
- Finally, *Stefan T. Hopmann* (Universität Wien) puts both the PISA project and the PISA discourse in a comparative perspective, showing how much the design, use of and response to PISA is depending on the needs and traditions of those involved.

All in all, the contributions give a very varied picture of the PISA effort. No step in the research process seems to be without substantial problems, several steps do not meet rigorous scholarly standards. Some of us seem to believe that these are obstacles, which can be solved within the PISA frame (e.g. Allerup, Dolin, Olsen, Sjøberg), others tend to a conclusion that the PISA project is beyond repair (e.g. Langfeldt, Meyerhöfer, Wuttke) or so much embedded in a specific political purpose, that it rather should be considered as a type of research-based policy making, not as a scholarly undertaking (e.g. Hopmann, Jahnke, Uljens, Bozkurt/Brinek/Retzl).

Almost all of the chapters raise serious doubts concerning the theoretical and methodological standards applied within PISA, and particularly to its most prominent by-products, its national league tables or analyses of school systems. Without access to the full set of original data, it is difficult to come to final conclusions. However, from our viewpoint, a few points seem to be evident beyond any reasonable doubt:

- PISA is by design culturally biased and methodologically constrained to a degree which prohibits accurate representations of what actually is achieved in and by schools. Nor is there any proof that what it covers is a valid conceptualization of what every student should know.
- The product of most public value, the national league tables (cf. Steiner-Khamsi 2003), are based on so many weak links that they should be abandoned right away. If only a few of the methodological issues raised in this

volume are on target, the league tables depend on assumptions about their validity and reliability which are unattainable.

- The widely discussed by-products of PISA, such as the analyses of “good schools”, “good instruction” or of differences between school systems and on issues like gender, migration, or social background, go far beyond what a cautious approach to these data allows for. They are more often than not speculative, and would at least need a wider framing by additional research looking at the aspects, which PISA by design cannot cover or gets wrong.
- Any policy making based on these data (whether about school structures, standards or the curriculum) cannot be justified. The use and misuse of PISA data in such contexts – done with or without PISA researchers consent or cooperation – belongs solely to the sphere of policy making. Of course PISA researchers have the same right as every citizen to pronounce their political convictions in public. However they cannot do so claiming research as an unquestionable basis for their arguments.

This does not mean that there are no valuable lessons to be drawn from PISA. At least it is a very innovative comparative study on the uneven distribution of a peculiar kind of knowledge and abilities among young people in different countries. However, the use of PISA as research on schooling by the OECD, its members and some of the research groups connected to the effort goes far beyond what is scientific evidence or simply well done research. PISA is not according to PISA, when it comes to how it is produced and used in these cases.

### **PISA – The Contergan of Educational Research?**

Of course, we would have loved to add to this volume commentaries and criticism of what is presented here by members of the PISA consortium – because we believe in the necessity of broad and uninhibited scholarly exchange. However, repeated invitations to address these issues in open symposia, or to contribute to this volume, remained either unanswered or were turned down. The German PISA consortium went so far to make an official decision not to participate in this effort; others simply kept silent. Time and again we were told in public and at meetings that most of the methodological criticism published on PISA has been proven wrong, and that every possible weakness has been taken care of. However, we could not obtain a published justification for this

claim. Even an invitation to contribute a summary of the counterarguments to this volume was turned down.

As sad as this is, it was no surprise. In the preparation of this volume we exchanged quite a few notes on how the national debates around PISA unfold in ‘our’ countries. What emerged was a picture not unlike that seen in the behaviour of large companies when they encounter a potential scandal, e.g. pharmaceutical companies dealing with ill-conceived drugs (like Chemie Grünenthal in the famous Contergan/Thalomid case or other scandals; cf. Kirk 1999; Luhmann 2000; Schulz 2001) where the strategy is one of an “issue framing” (cf. Entman 1993; Sniderman & Theriault 2004). To take just the most recent German example:

- If some critique is voiced in public, the first response seems to be silence. Or as the leader of the German consortium, Manfred Prenzel, puts it in case of this book: One doesn’t want to provide “a forum for unproven allegations” (as an answer to the invitation to participate in this book by mail 2007-05-09, which was turned down by a “unanimous” decision of the German PISA consortium confirmed by a mail 2007-05-21). He wrote this before knowing the authors and titles of all but one of the chapters contained in this volume.
- If that is not enough, the next step is often to raise doubts about the motives and the abilities of those who are critical of the enterprise. For instance, when asked about the recently published volume on *PISA & Co.* (Jahnke & Meyerhöfer 2006), Olaf Köller, as the head of the German National Institute for Educational Progress, suggested that (1) these critics were unqualified to discuss PISA (even though they included many leading members of the mathematics didactic research in Germany) and (2) they were probably driven by envy or other non-scholarly motives (Köller 2006a; Kerstan 2006).
- The next step seems to acknowledge some problems, but to insist that they are very limited in nature and scope, not affecting the overall picture. Alternatively, it is pointed out that these problems are well known within large-scale survey research of the kind like PISA, and even unavoidable when working comparatively (e.g. Köller 2006b). Of course that claim does not reduce the impact of these problems on the validity of the results.
- Finally, there is the statement that the criticism does not contain anything new, and nothing that has not been dealt with within the PISA research itself – and often this claim is accompanied by references to opaque technical

reports, that only insiders can understand, or to unpublished papers or reports (e.g. Prenzel & Walter 2006; Schleicher 2006).

What does *not* happen is what is normally considered to be “good science”: open debate on the pros and cons of the arguments. If one understands PISA as an economic enterprise, in line with the abovementioned pharmaceutical companies, this is quite reasonable. Ignoring, silencing, or simply marginalizing a critic does less harm to the brand than a public argument. A public rebuttal carries the risk that some customers would not be totally convinced (“semper aliquid haeret”). It is only necessary to take firmer steps when criticism finally becomes so public that it cannot be ignored by customers and buyers. But the first move is still to discredit the critics and their supporters as being uninformed, ill-equipped, or simply following a personal agenda. The final move rests on the claim that there is other research, which proves the critics wrong – although for a variety of reasons the data-sets on which these conclusions are based cannot be made available. By using such techniques, companies can hold the realistic expectation that even proven deficiencies will not harm sales substantially and over time.

Of course, the comparison of PISA and Contergan can be seen as over-reaching: Thalomid did lead to thousands of severely disabled newborns, whereas PISA only does harm to children’s education in the worst case. Additionally, the Grünenthal company directly advertised the medication for purposes with high risk, whereas the PISA consortium can argue that it is up to the people to believe or not to believe in what PISA tells. But other similarities are striking: PISA has a large “market share” to defend: most of public money spent on educational research nowadays is being put into PISA and similar approaches (the standards and testing business); many chairs in education have turned to related topics and issues, thus providing a significant market for collaborators in the field. This is all too big and too seductive to be put at risk just because of a few other scholars who do not support the whole enterprise or the way it is done.

The readers of this volume should expect similar responses to what is said here. But don’t worry: Nobody is going to pull PISA into courts of law because of its flaws – as was the case with the pharmaceutical companies. No other court than the one of public reasoning is available, but with Kant we do believe, that this is the strongest court of all.

Discussion is an essential part of science. Therefore we invite you to take

part in our discussion forum on the Internet and to post your opinion and critique concerning the book. Find more information at

<http://institut.erz.univie.ac.at/home/fe2/>.

We are looking forward to an inspiring discussion!

## References

- Aktionsrat Bildung: Bildungsgerechtigkeit. Jahresgutachten 2007. (ed. Vereinigung der Bayerischen Wirtschaft e.V.; online: [http://www.aktionsrat-bildung.de/fileadmin/Dokumente/Bildungsgerechtigkeit\\_Jahresgutachten\\_2007\\_-\\_Aktionsrat\\_Bildung.pdf](http://www.aktionsrat-bildung.de/fileadmin/Dokumente/Bildungsgerechtigkeit_Jahresgutachten_2007_-_Aktionsrat_Bildung.pdf) retr. 2007/07/07).
- Allerup, P.: PISA præstationer-målinger med skæve målestokke?. In: Dansk Pædagogisk Tidsskrift 2005-1, 68-81
- Allerup, P.: PISA 2000's læseskala – vurdering af psykometriske egenskaber for elever med dansk og ikke-dansk sproglig baggrund. (Rockwool Fondens Forskningsenhed og Syddansk Universitetsforlag) Odense 2006.
- Allerup, P.: Identification of Group Differences Using PISA Scales – Considering Effects of Inhomogeneous Items. In this volume.
- Benner, D.: Die Struktur der Allgemeinbildung im Kerncurriculum moderner Bildungssysteme. Ein Vorschlag zur bildungstheoretischen Rahmung von PISA. In: Zeitschrift für Pädagogik 48-2002-1, 68-90.
- Bodin, A.: What does PISA really assess? What it doesn't A French view. Report prepared for Joint Finnish-French conference "Teaching mathematics: Beyond the PISA survey", Paris 2005.
- Bodin, A.: What does PISA really assess? What it doesn't? A French view. In this volume.
- Bonnet, G.: Reflections in a Critical Eye: On the Pitfalls of International Assessment. In: Assessment in Education 2002-9, 387-400.
- Bottani, N. & Virgnaud, P.: La France et les evaluations internationales. Paris 2005. online: <http://lesrapports.ladocumentationfrancaise.fr/BRP/054000359/0000.pdf> (retr. 2007/07/07).
- Bracey, G.W.: Research: Put Our Over PISA. In: Phi Delta Kappan 86-2005-10, 797.
- Dohn, N.B.: Knowledge and Skills for PISA – Assessing the Assessment. In: Journal of Philosophy of Education. 41-2007-1, 1-16.
- Flitner, E.: Pädagogische Wertschöpfung. Zur Rationalisierung von Schulsystemen durch public-private-partnerships am Beispiel von PISA. In: Oelk-

- ers J. et al. (eds.): *Rationalisierung und Bildung bei Max Weber.* ) Bad Heilbrunn (Klinkhardt 2006, 245-266.
- Fuchs, H.-W.: *Auf dem Wege zu einem neue Weltcurriculum? Zum Grundbildungskonzept von PISA und der Aufgabenzuweisung an die Schule.* In: *Zeitschrift für Pädagogik* 49-2003-2, 161-179.
- Gaeth, F.: *PISA (Programme for International Student Assessment) Eine statistisch-methodische Evaluation.* Berlin (Freie Universität) 2005.
- Goldstein, H.: *International Comparisons of Student Attainment: Some Issues Arising from the PISA Study.* In: *Assessment in Education – Principles, Policy, and Practice* 11-2004-3, 319-330.
- Grisay, A. & Monseur, C.: *Measuring the Equivalence of Item Difficulty in the Various Versions of an International Test.* In: *Studies in Educational Evaluation* 33-2007-1, 69-86.
- Hermann, U.: *Fördern “Bildungsstandards” die allgemeine Schulbildung?* In: Rekus, J. (ed.): *Bildungsstandards, Kerncurricula und die Aufgabe der Schule.* Münster (Aschendorff ) 2005, 24-52.
- Herrmann U.: *PISA – Welche Konsequenzen für Schule und Unterricht kann man wirklich ziehen? Diskussionsbeitrag DIDACTA Hannover 2006.* FORUM BILDUNG (online: <http://forum-kritische-paedagogik.de/start/download.php?view.209>; retr. 2007/07/07).
- Hopmann, S.T.: *Restrained Teaching: The Common Core of Didaktik.* In: *European Educational Research Journal* 6-2007-2, 109-124.
- Huisken, F.: Der “PISA-Schock” und seine Bewältigung – Wieviel Dummheit braucht/verträgt die Republik?* Hamburg (VSA-Verlag) 2005.
- Jahnke, T. & Meyerhöfer, W. (eds.): *PISA & Co – Kritik eines Programms.* Hildesheim (Franzbecker) 2006.
- Kerstan, K.: *An PISA gescheitert.* In: *DIE ZEIT*, 16.11.2006 Nr. 47
- Kirk, B.: *Der Contergan-Fall: eine unvermeidbare Arzneimittelkatastrophe?* Stuttgart (Wissenschaftliche Verlagsgesellschaft) 1999.
- Klausnitzer, J.: *PISA – einige offene Fragen zur OECD Bildungspolitik.* 2006. Online: [http://www.links-netz.de/K\\_texte/K\\_klausnitzer\\_oecd.html](http://www.links-netz.de/K_texte/K_klausnitzer_oecd.html) (retr. 2007/07/07).
- Köller, O.: *Kritik an PISA unberechtigt.* Interview mit bildungsklick.de 2006a. online: <http://bildungsklick.de/a/50155/kritik-an-pisa-unberechtigigt> (retr. 2007/07/07).
- Köller, O.: *Stellungnahme zum Text von Joachim Wuttke: Fehler, Verzerrun-*

- gen, Unsicherheiten in der PISA-Auswertung. Press release of the National Institute for Educational Progress (IQB) 2006/11/14, (2006b).
- Kraus, J.: Der PISA Schwindel. Unsere Kinder sind besser als ihr Ruf. Wie Eltern und Schule Potentiale fördern können. Wien (Signum Verlag) 2005.
- Ladenthin, V.: Bildung als Aufgabe der Gesellschaft. In: *studia comenia et historica* 34-2004-71/72, 305-319.
- Lohmann, I.: After Neoliberalism. Können nationalstaatliche Bildungssysteme den ‚freien Markt‘ überleben? 2001 online: <http://www.erzwiss.uni-hamburg.de/Personal/Lohmann/AfterNeo.htm> (retr. 2007/07/07)
- Lohmann, I.: Was bedeutet eigentlich “Humankapital”? GEW Bezirksverband Lüneburg und Universität Lüneburg: Der brauchbare Mensch. Bildung statt Nützlichkeitswahn. Bildungstage 2007, online: <http://www.erzwiss.uni-hamburg.de/Personal/Lohmann/Publik/Humankapital.pdf> (retr. 2007/07/07)
- Luhmann, H.-J.: Die Contergan-Katastrophe revisited – Ein Lehrstück vom Beitrag der Wissenschaft zur gesellschaftlichen Blindheit. In: *Umweltmedizin in Forschung und Praxis*. 5-2000-5, 295-300.
- Nash, R.: Is the School Composition Effect Real? A Discussion with Evidence from the UK PISA Data. In: *School Effectiveness and School Improvement* 14-2003-4, 441-457.
- Neuwirth, E., Ponocny, I. & Grossmann, W. (eds.): PISA 2000 und PISA 2003. Graz (Leykam) 2006.
- Olsen, R.V.: Achievement Tests From an Item Perspective. An Exploration of Single Item Data form the PISA and TIMSS studies. Oslo (University of Oslo) 2005. Online at: [http://www.duo.uio.no/publ/realvag/2005/35342/Rolf\\_Olsen.pdf](http://www.duo.uio.no/publ/realvag/2005/35342/Rolf_Olsen.pdf) (retr. 2007/07/07)
- Pekrun, R.: Vergleichende Evaluationsstudien zu Schülerleistungen: Konsequenzen für die Bildungsforschung. In: *Zeitschrift für Pädagogik* 48-2002-1, 111-128.
- Prais S. J.: Cautions on OECD’s recent educational survey(PISA): Rejoinder to OECD’s response. In: *Oxford Review of Education* 30-2004-4
- Prais S.J.: Cautions on OECD’s Recent Educational Survey (PISA) In: *Oxford Review of Education* 29-2003-2, 139-163.
- Prenzel, M. & Walter, O.: Wie solide ist PISA? Oder Ist die Kritik von Joachim Wuttke begründet? Kiel (IPN) 2006 (two pages including a one page attachment!)
- Roeder, P.M.: TIMSS und PISA – Chancen eines neuen Anfangs in Bil-

- dungspolitik, -planung. -verwaltung und Unterricht. Endlich ein Schock mit Folgen? In: Zeitschrift für Pädagogik 49-2003-2, 180-197.
- Romainville, M.: On the Appropriate Use of PISA. In: La Revue Nouvelle 2002-3/4.
- Schleicher, A.: Interview mit der Frankfurter Rundschau. In: Frankfurter Rundschau vom 28.11.2006
- Schulz, J.: Management von Risiko- und Krisenkommunikation – zur Bestandserhaltung und Anschlussfähigkeit von Kommunikationssystemen. Berlin (Humboldt Universität) 2001.
- Stack, M.: Testing, Testing, Read All About It: Canadian Press Coverage of the PISA Results. In: Canadian Journal of Education, 29-2006-1, 49-69.
- Steiner-Khamsi, G.: The Politics of League Tables. ([http://www.sowi-onlinejournal.de/2003-1/tables\\_khamsi.htm](http://www.sowi-onlinejournal.de/2003-1/tables_khamsi.htm); retr. 2007/07/07).
- Weigel, T.M.: Die PISA-Studie im bildungspolitischen Diskurs. Eine Untersuchung der Reaktionen auf PISA in Deutschland und im Vereinigten Königreich. Diplomarbeit Trier (Universität) 2004.
- Wuttke, J. (2006): Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung.- In: Jahnke, T. & Meyerhöfer, W. (Hrsg): PISA & Co. Kritik eines Programms. Hildesheim, Berlin (Franzbecker), 101-154.
- Wuttke, J.: Uncertainties and Bias in PISA. In this volume.