

Klieme, Eckhard; Vieluf, Svenja

Schulische Bildung im internationalen Vergleich. Ein Rahmenmodell für Kontextanalysen in PISA

Jude, Nina [Hrsg.]; Klieme, Eckhard [Hrsg.]: PISA 2009 - Impulse für die Schul- und Unterrichtsforschung. Weinheim u.a. : Beltz 2013, S. 229-246. - (Zeitschrift für Pädagogik, Beiheft; 59)

urn:nbn:de:0111-opus-78298



in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen / conditions of use

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

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

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

Kontakt / Contact:

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

Zeitschrift für Pädagogik · 59. Beiheft

PISA 2009 – Impulse für die Schul- und Unterrichtsforschung

Herausgegeben von
Nina Jude und Eckhard Klieme

BELTZ JUVENTA

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, bleiben dem Beltz-Verlag vorbehalten.

Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2013 Beltz Juventa · Weinheim und Basel

www.beltz.de · www.juventa.de

Herstellung: Lore Amann

Gesamtherstellung: Beltz Bad Langensalza GmbH, Bad Langensalza

E-Book

ISSN 0514-2717

Bestell-Nr. 443501

Inhaltsverzeichnis

Nina Jude/Eckhard Klieme

PISA 2009 – Impulse für die Schul- und Unterrichtsforschung.

Einleitung zum Beiheft 7

Schul- und Unterrichtsbedingungen

Brigitte Steinert/Jan Hochweber/Silke Hertel

Lesekompetenz und Lesefreude von Schülerinnen und Schülern und bildungsstandardbezogene Kompetenzüberzeugungen und Lerngelegenheiten in Schule und Unterricht 12

Silke Hertel/Simone Bruder/Nina Jude/Brigitte Steinert

Elternberatung an Schulen im Sekundarbereich. Schulische Rahmenbedingungen, Beratungsangebote der Lehrkräfte und Nutzung von Beratung durch die Eltern .. 40

Timo Ehmke

Soziale Disparitäten im Lesen und in Mathematik innerhalb von Schulklassen .. 63

Schülerinnen und Schüler mit Migrationshintergrund

Aileen Edele/Petra Stanat/Susanne Radmann/Michael Segeritz

Kulturelle Identität und Lesekompetenz von Jugendlichen aus zugewanderten Familien 84

*Robin Segerer/Alexandra Marx/Petra Stanat/Wolfgang Schneider/Thorsten Roick/
Peter Marx*

Determinanten der Lesekompetenz bei Jugendlichen nicht deutscher Herkunftssprache. Zur Bedeutung der Spracherwerbsreihenfolge im Falle von Mehrsprachigkeit 111

Trends und Veränderungen

Timo Ehmke/Eckhard Klieme/Petra Stanat

Veränderungen der Lesekompetenz von PISA 2000 nach PISA 2009. Die Rolle von Unterschieden in den Bildungswegen und in der Zusammensetzung der Schülerschaft 132

Dorothea Mildner/Jan Hochweber/Andreas Frey

Vergleichende Analysen der Kompetenzen von Fünfzehnjährigen und Neuntklässlern in den deutschen PISA-Erhebungen 2003 bis 2009 151

Linda Marie Bischof/Jan Hochweber/Johannes Hartig/Eckhard Klieme

Schulentwicklung im Verlauf eines Jahrzehnts – Erste Ergebnisse des PISA-Schulpanels 172

Grundlagen und konzeptionelle Ansätze

Nina Jude/Johannes Hartig/Stefan Schipolowski/Katrin Böhme/

Petra Stanat

Definition und Messung von Lesekompetenz. PISA und die Bildungsstandards .. 200

Eckhard Klieme/Svenja Vieluf

Schulische Bildung im internationalen Vergleich. Ein Rahmenmodell für Kontextanalysen in PISA 229

Eckhard Klieme/Svenja Vieluf

Schulische Bildung im internationalen Vergleich

Ein Rahmenmodell für Kontextanalysen in PISA

1. Eine Meta-Perspektive auf PISA

Das *Programme for International Student Assessment* (PISA) baut auf der mehr als fünfzigjährigen Tradition der international vergleichenden empirischen Bildungsforschung auf. 1958 schlossen sich Wissenschaftlerinnen und Wissenschaftler aus Europa und den USA in der *International Association for the Evaluation of Educational Achievement* (IEA) zusammen, um die qualitativ und historisch-systematisch argumentierende Komparatistik durch Schulleistungsstudien zu ergänzen. Seit den 1960er Jahren führt die IEA regelmäßig internationale Leistungsstudien durch, derzeit u.a. die Primarschulstudien TIMSS und PIRLS (in Deutschland IGLU genannt). Über die Ziele und Grenzen dieses Forschungsparadigmas wird in der IEA seit langem und immer wieder neu diskutiert. So unterscheiden Keeves und Lietz (2011) zwei Modi der international vergleichenden Schulleistungsforschung: einen deskriptiven, auf die fortlaufende Beobachtung von Systemmerkmalen und Ergebnissen bezogenen Modus, und einen erklärenden Modus, der die Frage in den Vordergrund stellt, inwieweit bildungs- und erziehungswissenschaftliche Befunde auf unterschiedliche kulturelle und systemische Kontexte übertragbar sind. Dem internationalen Vergleich kommt in den beiden Modi eine höchst unterschiedliche Funktion zu: Das sogenannte Bildungsmonitoring, einschließlich „Ranking“ und „Benchmarking“, einerseits, die Ausweitung des Erkenntnishorizonts durch eine interkulturelle und systemische Dimension andererseits.

Die Chance, aber auch die Problematik der internationalen Leistungsstudien besteht darin, dass die Untersuchungen beide Arten des Zugriffs zwanglos ermöglichen. Dementsprechend werden zwei Arten von „Produkten“ für Politik, Öffentlichkeit, Profession und Forschung bereitgestellt: (1) *Indikatoren*, mithilfe derer sich Struktur und Funktion sowie Produktivität und Chancengleichheit in Bildungssystemen beobachten lassen. (2) *Zuverlässige Daten*, die es Vertretern unterschiedlicher Disziplinen ermöglichen, Grundlagenforschung zu betreiben, aber auch bildungspolitisch relevante Fragen zu untersuchen, also *Wissen* zu schaffen über Faktoren, die die Qualität von Bildungssystemen beeinflussen.

In der Öffentlichkeit wurde und wird fast ausschließlich der erste Modus rezipiert. So führten bereits in den 1970er Jahren die im Rahmen des Six Subject Survey (Walker, 1976) ermittelten kritischen Befunde zum Leistungsstand im deutschen Bildungssystem zu Konflikten zwischen Kultusministerkonferenz und Wissenschaft und letztlich zum Ausscheiden Deutschlands aus den internationalen Vergleichen. Die TIMSS-Mittelstufenstudie, an der Mitte der 1990er Jahre auch Deutschland wieder teilnahm, bestätigte bezüglich des Leistungsvergleichs im Kern den zwanzig Jahre zuvor ermittelten

Befund (Baumert et al., 1997), gab aber – vor allem durch die angeschlossene Videostudie – zusätzliche Impulse im Sinne des zweiten Interpretationsmodus, d.h. Erkenntnisse und weiterführende Anregungen für eine fachdidaktisch, unterrichtswissenschaftlich und kulturvergleichend ausgerichtete Forschung, die über das Modellprogramm SINUS wiederum den mathematischen und naturwissenschaftlichen Unterricht in Deutschland nachhaltig beeinflusst hat. Eine ähnliche Rolle spielt in jüngster Zeit die IEA-Studie TEDS-M zu Leistungen und Einstellungen von Lehramtsstudierenden (Blömeke, Kaiser & Lehmann, 2010).

Ende der 1990er Jahre begann die *Organisation für wirtschaftliche Zusammenarbeit und Entwicklung* (OECD) diese Forschungstradition zu nutzen, um ihr System statistischer Indikatoren zu erweitern. PISA erhebt seitdem alle drei Jahre in den Mitgliedsstaaten der OECD und in mittlerweile mehr als 30 Partnerstaaten Kompetenzen von 15-jährigen Schülerinnen und Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften sowie zusätzliche fächerübergreifende Kompetenzen wie etwa das Problemlösen (PISA 2003 mit schriftlichen Aufgaben, PISA 2012 mit computer-simulierten Szenarien und PISA 2015 mit Aufgaben zur Interaktion mit einem simulierten Partner). PISA liefert zuverlässige Informationen zur Beobachtung und für den Vergleich von Bildungssystemen. Mithilfe der Ergebnisse können Strukturen von Bildungssystemen beschrieben sowie deren Funktionieren und (soweit international messbar) deren Ergebnisse beobachtet werden. Beispielsweise beschreibt PISA Schülerleistungen, aber auch nicht-kognitive *Outcomes* (z.B. Motivation), Bildungswege von Schülerinnen und Schülern, Merkmale von Schulen wie Schulklima und Lehrkooperation sowie Aspekte der Chancengleichheit im Bildungswesen (z.B. die geschlechtsspezifische Bildungsbeteiligung und die Koppelung zwischen sozioökonomischem Status und Schülerleistungen). Damit stellt die Studie „Transparenz über Rahmenbedingungen, Prozesse und Ergebnisse des Bildungssystems her, nutzt dabei verschiedene Perspektiven und dient somit der öffentlichen Verantwortung für das Bildungsgeschehen“ (Klieme, 2011).

Anders als die Studien der IEA ist PISA sogar selbst eine Instanz, die zum Gegenstand gesellschafts- und erziehungswissenschaftlicher Forschung wurde. Eine internationale Autorengruppe beschreibt PISA als „wissensbasiertes Regulationsinstrument im bildungspolitischen Raum“ (Ozga, 2012, S. 169; Übersetzung durch die Autoren), dessen Wirkung darin bestehe, dass Daten in Wissen und schließlich in normative politische Argumente transformiert würden, die Problemlagen definieren oder sogar Empfehlungen für Problemlösungen geben. Für nationale Bildungspolitik bedeutet dies – je nach Sichtweise – neue Möglichkeiten zur Analyse der Bildungswirklichkeit und eine Ausweitung der politischen Steuerungsmittel (so Ozga, 2008), oder die Entmachtung durch eine „transnationale Koalition aus Forschern, Managern und Unternehmensberatern“ (Münch, 2009, S. 30). Tatsächlich werden die PISA-Studien gern als Legitimationsinstanz für das „borrowing“, d.h. die Übertragung bildungspolitischer Konzepte eines „erfolgreichen“ Landes auf andere Staaten, verwendet – seien es Marktmechanismen, Standards und Evaluationen, Lehrerbildung oder Ganztags- und Einheitsschul-

system. Diese Kritik trifft PISA härter als IEA-Studien, nicht zuletzt weil mit der OECD eine mächtige supranationale Institution der Auftraggeber ist.

Auch PISA ist aber mehr als ein Vergleichstest zur Identifizierung „erfolgreicher“ Bildungssysteme im Sinne des Leistungsrankings. Neben den kognitiven Kompetenzen werden sogenannte nicht-kognitive Outcomes erhoben, also erzieherische Wirkungen im Sinne von Fend (2006) wie etwa Lernmotivation, fachbezogene Einstellungen, kooperative vs. kompetitive Orientierungen und Verhaltensweisen im Bereich Schulschwänzen. Zunehmend wichtiger ist über die Erhebungen hinweg – auch auf internationaler Ebene – die Kontextualisierung der Testergebnisse geworden, die auf Befragungen von Schülerinnen und Schülern, Schulleitungen, Eltern und bei PISA 2015 erstmals auch Lehrkräften beruht. Hier werden Faktoren abgefragt, von denen aus der Schul- und Unterrichtsforschung bekannt ist, dass sie einen Effekt auf schulische Lernergebnisse, Einstellungen und Verhaltensweisen haben. Wichtige Ausgangsbedingungen der schulischen Arbeit sind beispielsweise die familiäre Lernumgebung sowie die Ausstattung und Finanzierung der Schule. Zu den erhobenen Prozessmerkmalen des Lehr- und Lerngeschehens gehören unter anderem individuelle Lernstrategien, die Unterrichtsqualität und das Schulklima. Dabei wird berücksichtigt, dass diese Merkmale auf verschiedenen Ebenen von Schulsystemen (Schüler-, Klassen-,¹ Schul- und Systemebene) anzusiedeln sind. PISA stellt somit eine hoch komplexe Datenbasis bereit, die – entsprechend der oben dargestellten Dualität der Schulleistungsstudien – sowohl für Zwecke des Bildungsmonitorings genutzt werden kann als auch für wissenschaftliche Untersuchungen mit allgemeinen, nationalen und international vergleichenden Fragestellungen.

Diese wissenschaftlichen Untersuchungen zielen, abstrakt gesprochen, auf Muster von Zusammenhängen zwischen Ausgangsbedingungen, Prozessmerkmalen und verschiedenen Lernergebnissen. Der erste Analyseschritt ist hierbei die Zerlegung der Varianz von Schülerleistungen und anderen Merkmalen auf die Schüler-, die Schul- und die Systemebene; sie liefert Hinweise, inwieweit bildungsbezogene Prozesse und Ergebnisse tatsächlich auf der Ebene der Einzelschule oder des nationalen Schulsystems geprägt werden. Auf allen drei Ebenen werden sodann Variablen einbezogen, um deren direkte und indirekte Zusammenhänge mit Schülerleistungen zu bestimmen. Eine große, repräsentative Stichprobe und differenzierte statistische Modelle gewährleisten eine genaue Absicherung der Ergebnisse, sowohl innerhalb als auch zwischen den teilnehmenden Staaten.

1 PISA kann Merkmale der Klassenebene wie beispielsweise Unterrichtspraktiken erheben und international vergleichend beschreiben. In Bedingungsanalysen kann die Klassenebene allerdings nur in einzelnen Teilnahmestaaten (z.B. Deutschland) im Rahmen nationaler Ergänzungen untersucht werden, da auf internationaler Ebene eine altersbasierte Stichprobenziehung vorgenommen wird, also Fünfzehnjährige quer über Jahrgänge und Klassen hinweg einbezogen werden.

Das Design der Fragebögen ist zunehmend professionalisiert worden und stellt inzwischen einen eigenständigen Kernbereich des internationalen PISA-Projekts dar. Während die beteiligten Wissenschaftler im ersten Zyklus der PISA-Studien vornehmlich damit befasst waren, Tests und dazu passende Fragebögen für den jeweils im Vordergrund stehenden Leistungsbereich zu entwickeln (Lesen bei PISA 2000, Mathematik im Jahr 2003, Naturwissenschaften in 2006), wird im derzeitigen zweiten Durchgang (PISA 2009, 2012, 2015) an einem integrierten Design zur Erfassung von Ausgangsbedingungen, Prozessmerkmalen und erzieherischen Wirkungen gearbeitet, das in aktuellen Theorien und Methoden der empirischen Bildungsforschung verankert ist. Hierzu wurden gesonderte Expertengruppen² eingesetzt, die entsprechende Rahmenkonzepte formulieren sollten. Neben dem Fragebogen-Design dient das jeweilige Rahmenkonzept auch der Vergewisserung über Fragestellungen, Aussagemöglichkeiten und Grenzen der Studien. Insofern sind gerade diese Konzepte wichtige Dokumente der Selbstreflexion der an PISA beteiligten Wissenschaftler.

Ziel des hier vorliegenden Aufsatzes ist es, diese Selbstreflexion für die deutsche Diskussion zugänglich zu machen. Die nachfolgenden Abschnitte beruhen wesentlich auf dem internationalen Rahmenkonzept für PISA 2012, das von den Autoren unter Beteiligung der internationalen Expertengruppe verfasst wurde (Klieme, Backhoff-Escudero et al., 2010). Im vorliegenden Aufsatz gehen wir jedoch vertiefend auf Forschungsliteratur ein und integrieren bereits Elemente des Rahmenkonzeptes für PISA 2015.

Nachfolgend (Abschnitt 2) wird zunächst am Beispiel PISA 2012 das Design der Studien dargestellt. Im Vordergrund stehen die in den Schüler-, Eltern- und Schulfragebogen ermittelten Merkmale, wobei auch erläutert werden soll, auf welcher theoretischen Basis diese Merkmale ausgewählt und operationalisiert werden. Anschließend (Abschnitt 3) werden aus der Nutzerperspektive, d.h. vor allem aus der Sicht von Bildungspolitik und -administration, Funktionen und Ziele der PISA-Studien erläutert. Es wird deutlich, dass PISA in der Tat an erster Stelle dem Bildungsmonitoring dient. Das internationale Rahmenkonzept geht aber auch auf die wissenschaftlichen Aussagemöglichkeiten und Grenzen der PISA-Studien ein, hier dargestellt in Abschnitt 4. Die Ausführungen zeigen insgesamt, inwieweit PISA anschlussfähig ist an die empirisch-pädagogische Forschung und andere Bereiche der Bildungsforschung.

2. Das PISA-Design und seine wissenschaftlichen Grundlagen

Um die verschiedenen Komponenten des Untersuchungsdesigns für internationale Leistungsvergleiche zu systematisieren, hat die *International Association for the Evaluation of Educational Achievement* bereits in den 1960er Jahren das sogenannte Kontext-Input-Prozess-Outcome-Modell (Purves, 1987) entwickelt, das sich ganz ähnlich in vielen

2 Die internationale Expertengruppe wurde bzw. wird bei PISA 2009 von Jaap Scheerens (Universität Twente, Niederlande), bei PISA 2012 und PISA 2015 von Eckhard Klieme (DIPF) geleitet.

aktuellen Arbeiten der empirischen Bildungsforschung findet. In diesem Modell werden Hintergrundvariablen und strukturelle Bedingungen (Inputs) unterschieden von Aktivitäten und Maßnahmen und deren Qualität (Prozesse) sowie den Leistungen und nicht-kognitiven (mit Fend: erzieherischen) Ergebnissen. Zum zweiten werden Handlungsebenen abgegrenzt, vom Bildungs- bzw. Schulsystem über die Ebene der Organisation (Einzelschule) und die Klassenebene zu Lernvoraussetzungen, -aktivitäten und Ergebnissen auf der Schülerebene. Die übergeordneten Ebenen werden als Kontext angesehen, der die jeweils untergeordneten Ebenen beeinflusst.

Eine aktuelle Version dieser im Englischen „CIPO-Modell“ genannten Matrix stellt Tabelle 1 dar. Diese Darstellung wurde dem internationalen Rahmenkonzept für PISA 2012 entnommen (Klieme, Backhoff-Escudero et al., 2010). Bei den in Tabelle 1 aufgeführten Merkmalen handelt es sich demnach um Konstrukte, die im Test- und Fragebogendesign der PISA-Erhebung 2012 international umgesetzt wurden. Gegenüber früheren Erhebungswellen konnte vor allem der Schülerfragebogen wesentlich erweitert werden, weil nun auch hier – wie schon seit PISA 2000 im Leistungstest – das Prinzip des sogenannten „Matrix-Designs“ eingesetzt wurde, bei dem die Schülerinnen und Schüler teilweise unterschiedliche, aber einander überlappende Mengen von Fragen bzw. Aufgaben bearbeiten. Dieses Prinzip erlaubte es im Jahr 2012, die Zahl der einbezogenen Konstrukte wesentlich zu vergrößern.

Die erste Spalte der Tabelle benennt die drei Ebenen, die in den PISA-Fragebögen explizit angesprochen sind: die Schüler-, die Klassen- und die Schulebene. Daneben werden die jeweiligen bei PISA erfassten Prozessmerkmale dargestellt sowie ggfs. Input- und Outcome-Merkmale. Prozessmerkmale sind alle Aspekte der Qualität und der Quantität des Lehrens und Lernens; darüber hinaus fallen professionelle Aktivitäten von Lehrerinnen und Lehrern und Schulleiterinnen und Schulleitern in diese Kategorie. Kognitive wie nicht-kognitive Ergebnisse (Outcomes) werden bei PISA ausschließlich auf der Ebene der Schülerinnen und Schüler durch Tests und Fragebögen erfasst; auf der Schul- und der Systemebene werden dann entsprechende Mittelwerte betrachtet. „Inputs“ werden bei Schülern und deren Eltern sowie auf der Schulebene (bei den Schulleitungen) erfragt, teilweise auch aus den Angaben der Schüler und Eltern aggregiert. Die Input-Variablen betreffen größtenteils den sozialen und familiären Hintergrund, aber auch Strukturmerkmale auf der Schulebene, zum Beispiel die Größe der Schule oder die Art der Finanzierung. Lehr- und Lernprozesse in der Klasse werden in den verschiedenen Fragebögen angesprochen und deshalb bei Tabelle 1 in einer gesonderten Zeile dargestellt; statistisch lässt die Stichprobenziehung der internationalen Studie (anders als die im vorliegenden Band ausgewertete deutsche Zusatzerhebung) jedoch nicht zu, einzelne Klassen zu identifizieren, sodass auch die Klassenmerkmale aggregiert auf der Schul- bzw. Systemebene ausgewertet werden.

<p>Schülerinnen und Schüler</p>	<p>Input</p> <p>Geschlecht, Alter, sozio-ökonomischer Status, Migrationshintergrund, Klassenstufe, Bildungskarriere, Klassenwiederholungen, Noten</p> <p>kulturelle Ressourcen, familiäre Lernumwelt und Unterstützung</p> <p>Erfahrung, Einstellungen und Fähigkeiten bzgl. Informations- und Kommunikationstechnologie</p> <p>Offenheit, Ausdauer und Engagement beim Problemlösen, Problemlösestile</p>	<p>Prozesse</p> <p>Lernzeit (inklusive Hausaufgaben und Nachhilfe)</p> <p>Teilnahme an außerunterrichtlichen Aktivitäten</p> <p>Unpünktlichkeit, Schulschwänzen</p> <p>schulisches Engagement</p> <p>schulbezogene und fachspezifische subjektive Normen,</p> <p>wahrgenommene Kontrolle des Lernprozesses,</p> <p>Attributions-Stil</p> <p>Akkulturationsstrategien</p>	<p>Outcomes</p> <p>Kompetenzen in Mathematik, Lesen, Naturwissenschaften, Problemlösen</p> <p>metakognitive Strategien</p> <p>Einstellung zum Fach, fachspezifische Motivation, fachspezifische Lernintentionen, fachspezifisches Selbstkonzept und Selbstwirksamkeit</p> <p>allgemeine schulbezogene Einstellungen, allgemeine Lernmotivation, Bildungsaspirationen</p>
<p>Klassen</p>	<p>Prozesse</p> <p>Unterrichtszeit, Leistungsgruppierungen, Leistungsbeurteilung und Feedback</p> <p>implementiertes Curriculum, Aufgaben, lernbezogene Aktivitäten, Nutzung von Lernmaterial</p> <p>Unterrichtsqualität: Disziplin/ Klassenklima, Unterstützung, Strukturierung, Schülerorientierung, kognitive Aktivierung</p>		
<p>Schulen</p>	<p>Input</p> <p>Schulform, Schulgröße, soziale und ethnische Komposition, Größe des Standorts</p> <p>Status (öffentlich versus privat), personelle und materielle Ausstattung, Wettbewerb mit anderen Schulen, mittlere Klassengröße</p> <p>Elternengagement, Leistungsdruck der Eltern</p>	<p>Prozesse</p> <p>Lehrerbildung/ Lehrerfortbildung</p> <p>Aufnahmekriterien der Schule, Bildungsgänge, Kursangebot/Curriculum, Leistungsdifferenzierung</p> <p>Lernzeit in den jeweiligen Fächern/ intendiertes Curriculum, Angebot an extracurricularen Aktivitäten, Angebote der Berufsberatung und -information</p> <p>geteilte Normen, Leitungsstil, Arbeitsmoral der Lehrkräfte und Lehrerkooperation</p> <p>Schulentwicklung, Schulevaluation, Rechenschaftslegung, Schulautonomie, externe Kooperationen</p> <p>Schulklima, Schüler-Lehrer-Beziehungen, unterstützende Lernumgebung, Regelungen zum Umgang mit Schulschwänzen</p>	

Tab. 1: Einordnung bei PISA 2012 erfasster Merkmale in eine Taxonomie der Ergebnisse („Outcomes“) von Bildungsprozessen und ihrer Bedingungen auf verschiedenen Ebenen des Bildungssystems

Auf die kognitiven Ergebnisvariablen, also die Leistungstests, die überwiegend im Zentrum der PISA-Rezeption stehen, soll an dieser Stelle nicht im Detail eingegangen werden (vgl. dazu etwa Klieme, Artelt et al., 2010). Aus erziehungswissenschaftlicher Sicht ist allerdings bemerkenswert, dass das Kompetenzkonzept der Studie letztlich auf die Pädagogische Anthropologie von Heinrich Roth (1971) zurückgeführt werden kann. Ausgehend von Mündigkeit als oberstem Ziel einer der Emanzipation verpflichteten Erziehung führt Roth den Kompetenzbegriff ein, um Mündigkeit auszdifferenzieren, und zwar in einem dreifachen Sinn, nämlich als Selbstkompetenz, Sachkompetenz und Sozialkompetenz (Roth, 1971, S. 180). Die in PISA erhobenen Grunddimensionen der Lese-, Mathematik- und Naturwissenschaft-Kompetenz sind als Sachkompetenzen im Sinne von Roth zu verstehen. Darüber hinaus erfasst PISA auch breitere kognitive Fähigkeiten, wie zum Beispiel Lernstrategien und Problemlösekompetenz. Ferner werden auch Aspekte der Selbstkompetenz sowie teilweise der Sozialkompetenz in den Kontextfragebögen berücksichtigt, so etwa Engagement, gemeinsame Werte und Überzeugungen, Motivation zu lernen und zu kooperieren und die Fähigkeit zur Selbstregulation eigener Lernprozesse. Die Bezüge zur Diskussion um Kompetenzen, wie sie Heinrich Roth in die Erziehungswissenschaft eingeführt hat, sind nicht zufällig: Über Publikationen von Fend und Weinert fanden sie Eingang in das OECD-Projekt „Definition and Selection of Competences“ sowie schließlich in die PISA-Studie (vgl. Klieme & Hartig, 2007).

Das CIPO-Modell ist zunächst nichts anderes als eine Taxonomie, die hilft, Konstrukte bzw. (auf der Ebene der Operationalisierungen) Messvariablen zu systematisieren. Häufig wird das Modell aber auch dazu genutzt, das Zusammenspiel von Faktoren auf verschiedenen Ebenen zu beschreiben. Man findet in solchen Darstellungen oft allerhand Pfeile, die andeuten sollen, dass „höhere“ auf „niedrigere“ Ebenen sowie Inputs auf Prozesse einwirken, die wiederum zu Outcomes führen. Dass eine solche einfache „Produktionsfunktion“, wie sie etwa Walberg (1986) in seinem einflussreichen Aufsatz darstellte, zu einfach ist, um die komplizierten Zusammenhänge, Wirkungen und Wechselwirkungen zu verstehen, ist aus erziehungswissenschaftlicher Sicht offensichtlich.

Schon die Unterscheidung zwischen Input, Prozess und Outcomes ist keineswegs eindeutig. Manche Prozessmerkmale (z.B. Lernstrategien) können ebenso gut als Inputs oder als Outcomes behandelt werden, abhängig von der theoretischen Perspektive, dem Forschungsdesign oder praktischen Überlegungen. In einem dynamischen Modell der Schuleffektivität (vgl. Creemers & Kyriakides, 2008) können zudem Outcomes als Inputs für die weitere Entwicklung angesehen werden. Angst vor Mathematik kann zum Beispiel als ein (negatives) Outcome schulischer Bildung angesehen werden, aber auch als ein Inputmerkmal, das einen Effekt auf weitere Prozess- und Outcome-Variablen, wie etwa Hausaufgabenaktivitäten, hat. Diese Überlegungen zeigen, dass das CIPO-Modell zwar ein nützliches Ordnungsschema ist, aber keine ausreichende theoretische Basis für systematische Wirkungsanalysen.

Um die Konstrukte und Messgrößen, die in das Design eingehen, auszuwählen und die daraufhin erhobenen Daten angemessen zu analysieren, nutzt PISA Wissen aus der

aktuellen Bildungsforschung. Zwei Forschungsfelder sind dabei von besonderer Bedeutung: die Schulwirkungsforschung sowie fachspezifische Lehr-Lern-Forschung.

Das zentrale Ziel der Schulwirkungsforschung ist es, „Faktoren im Bereich des Unterrichts, des Curriculums und der Lernumgebung auf verschiedenen Ebenen, wie der Klassen- und der Schulebene, [zu identifizieren], die – unter Berücksichtigung von Hintergrundmerkmalen wie der Fähigkeit, dem sozioökonomischen Status und der vorherigen Leistung – direkt oder indirekt Unterschiede zwischen Schülern bezüglich der Outcomes erklären können“ (Creemers & Kyriakides, 2008, S. 12; Übersetzung durch die Autoren). Im Sinne kognitiver Modelle des Lernens und Lehrens werden diese Faktoren nicht als direkt wirksame Maßnahmen gesehen, sondern als soziale Interaktionen und Lerngelegenheiten, die Schülerinnen und Schüler – ausgehend von ihren individuellen Fähigkeiten und Vorkenntnissen – durch eigene Lernaktivitäten für den Kompetenzerwerb nutzen. Die Schulwirkungsforschung kann mittlerweile auf einen beträchtlichen Wissensschatz zurückgreifen (z.B. Creemers & Kyriakides, 2008; Scheerens, 2000; Teddlie & Reynolds, 2000). Es ist vor allem das Verdienst von Jaap Scheerens, dieses Wissen systematisch für das Design der PISA-Fragebögen erschlossen zu haben.

Da PISA fachliche Kompetenzen beschreiben und erklären will, muss zudem auf fachspezifische Lehr-Lern-Forschung zurückgegriffen werden. Das Rahmenkonzept für PISA 2012 bezieht sich auf moderne Konzeptionen für den Mathematikunterricht wie „Lernen im Kontext“ (Schoenfeld, 2006) und „Diskussions-basierten Unterricht“ (Stein, Engle, Smith & Hughes, 2008), mit mathematischem Problemlösen und Modellieren (Blum & Leiss, 2005) als Kernaktivitäten. Die „derzeitige Vision eines wirksamen Mathematikunterrichts“, wie Stein und Koautoren (2008, S. 315) es nennen, sieht folgendermaßen aus: „Den Schülern werden vermehrt realistische und komplexe mathematische Probleme vorgelegt; sie unterstützen sich wechselseitig beim Lösen der Problemstellungen und diskutieren schließlich ihre Strategien und Lösungen in der Klasse unter Anleitung des Lehrers“ (Übersetzung durch die Autoren). Um diese Vision des Mathematikunterrichts abbilden, aber auch vom herkömmlichen Lösen standardisierter Aufgaben abgrenzen zu können, verwendet PISA 2012 gleich drei parallele, einander ergänzende Ansätze: Anhand exemplarischer Aufgabenstellungen werden Schüler und Schülerinnen gefragt, welche Lerngelegenheiten (opportunities to learn; Schmidt & Maier, 2009) sie aus ihrem Unterricht kennen. Zum zweiten werden Basisdimensionen der Unterrichtsqualität im Sinne von Klieme, Pauli und Reusser (2009) erfasst, und zum dritten wird die Häufigkeit der Anwendung lehrerzentrierter, schülerorientierter und individuell fördernder Unterrichtspraktiken ermittelt. Damit macht PISA 2012 einen innovativen, breiten Anlauf zu einer international vergleichenden Unterrichtsforschung, die in den Rahmen der Schulleistungsstudie mit Schwerpunkt Mathematik eingebaut ist.

3. Produkte der PISA-Studie im Rahmen des Bildungsmonitorings

Aus Sicht von Bildungspolitik und -administration kann PISA als ein Element der sogenannten neuen Steuerung der öffentlichen Verwaltung (New Public Management) be-

griffen werden, die Transparenz, Evaluation und Rechenschaftslegung eine große Bedeutung zumisst, bei einem gleichzeitigen Anstieg von Autonomie und Dezentralisierung (z.B. Böttcher, Bos, Döbert & Holtappels, 2008). Eine dauerhafte, datengestützte Beobachtung der Entwicklung von Bildungssystemen, wie sie PISA vornimmt, kann für die Evaluation und Rechenschaftslegung auf Systemebene genutzt werden. Die OECD selbst nutzt die Ergebnisse in ihren „Indikatoren für Bildungssysteme“. Deren Veröffentlichung regt öffentliche Debatten an, orientiert Bildungspolitik und liefert Grundlagen für Entscheidungen über Maßnahmen und Strategien in Bildungssystemen.

Dementsprechend entscheiden politische Fragestellungen über die Auswahl der Indikatoren. Bildungspolitik beschäftigt sich vor allem mit Ressourcen, Strukturen und Regeln für Prozessabläufe in Schulsystemen (z.B. Bildungsgänge, Curricula und Standards, personelle und materielle Ausstattung), mit den Bildungswegen und -ergebnissen (etwa dem Niveau und Profil der Schülerleistungen) und schließlich auch mit Fragen der Chancengleichheit (z.B. wie Ressourcen und Zugangschancen verteilt werden). Entsprechend reflektieren die in PISA gebildeten Indikatoren diese grundlegenden Themen.

Eine Vielzahl von Indikatoren, die auf den PISA-Fragebögen basieren, werden in den verschiedenen Ausgaben des OECD-Berichts „Bildung auf einen Blick“ dargestellt (vgl. zuletzt OECD, 2011). Neben der reinen Dokumentation von Leistungsergebnissen waren dies in der Vergangenheit:

- Der Zusammenhang von Geschlecht, sozialer Herkunft und Migrationshintergrund mit der Schülerleistung (nahezu jährlich, von 2002-A10 bis 2011-A5);
- Einstellungen von Schülern wie Bildungsaspiration, Lernmotivation und Interesse, Selbstregulation, Leseverhalten und Zugehörigkeitsgefühl zur Schule sowie deren Zusammenhänge mit individuellen Voraussetzungen einerseits, Leistungen andererseits (2003-A9 & A10, 2004-A8, 2007-A4 & A5, 2011-A6);
- Profile von Schülerinnen und Schülern, die herausragende Leistungen erbringen, unter anderem deren Einstellungen und Motivation (2009-A4 & A5); analog: Profile von Risikoschülern (2006-A6);
- Zusammenhänge zwischen Outcomes und Ressourcen (2007 & 2008-B7), vor allem der Klassengröße (2008-D2);
- Institutionelle Differenzierung im Bildungssystem – verstanden als horizontale und vertikale Differenzierung – und deren Zusammenhang mit Schülerleistungen (2002-A7, 2006-A7);
- Vergleich zwischen öffentlichen und privaten Schulen (2005-D5);
- Outcomes von berufsbezogenen, verglichen mit allgemeinen Bildungsgängen (2007 & 2008-C1);
- Die Nutzung von Evaluation und Leistungsbeurteilung in Bildungssystemen (2008-D5);
- Schulklima und Schulqualität aus der Perspektive von Eltern (2008-A5) bzw. Schulleitungen (2002-D5);
- Chancenverteilung bezüglich des Zugangs zu qualifizierter Bildung (2007-A4 &

A7, 2008-A3 & A7, 2011-D6) oder technologiegestützten Lerngelegenheiten (2002-D3 & D4) in Abhängigkeit vom sozialen und familiären Hintergrund der Schülerinnen und Schüler.

Ob auf PISA beruhende Indikatoren für Politik und Bildungsadministration informativ und nützlich sind, hängt nicht zuletzt von der Breite und Aussagefähigkeit des Designs ab. Ein Mittel, die Angemessenheit des Designs und der darauf beruhenden Indikatoren zu prüfen, ist ein Abgleich mit der aktuellen Forschungsliteratur, etwa dem „Handbook of Education Policy Research“ (Sykes, Schneider & Plank, 2009). In seinen 76 Kapiteln werden verschiedenste Aspekte dessen behandelt, was man im Deutschen wohl als „Steuerung des Bildungswesens“ übersetzen würde – jedoch ohne den mit dem Steuerungsbegriff konnotierten Anspruch auf Durchgriff „von oben“. Der Abschnitt des Buches, der sich mit „Policy implications of resources, management and organization“ befasst, beinhaltet acht „steuerungsrelevante“ Themen wie Klassengröße, Chancengerechtigkeit, Schulwahl und Ressourcen; für alle diese Themen kann PISA dank des Fragebogen-Designs Kennwerte liefern. Dasselbe gilt für den Abschnitt „Teaching and learning policy“ mit Themen wie Lerngelegenheiten und Qualifikation der Lehrkräfte sowie den Abschnitt „Educational access and differentiation“ zu Bildungswegen – einschliesslich früher Bildung –, ethnischer Diskriminierung oder dem „Schattensystem“ von Nachhilfe und privaten Zusatzschulen. Praktisch alle Themen des Handbuchs können durch PISA-Indikatoren langfristig, international vergleichend beobachtet werden, mit Ausnahme der Fragen zu Entscheidungsprozessen in der Bildungspolitik, z.B. zur Rolle der Lehrgewerkschaften.

Die besondere Chance eines auf PISA aufbauenden Bildungsmonitorings liegt im internationalen Vergleich der Indikatoren, der allerdings voraussetzt, dass die relevanten Indikatoren sich tatsächlich in international äquivalenter Weise messen lassen (Hui & Triandis, 1985; van de Vijver & Leung, 1997; van de Vijver, van Hemert & Poortinga, 2008). Die Vergleichbarkeit von Indikatoren soll durch das Einbeziehen internationaler Experten, durch nationale Adaptationen und durch einen Feldtest gewährleistet werden. Für Leistungstests und Fragebogenskalen lässt sich deren Vergleichbarkeit auch empirisch überprüfen mithilfe von Analysen differenzieller Itemfunktionen und Mehrgruppenfaktorenanalysen. Hier geht es vor allem um den Einfluss kulturspezifischer Werte, Normen und mentaler Schemata auf die Bedeutung von Konstrukten, die Interpretation einzelner Fragen sowie das Antwortverhalten (van de Vijver & Leung, 1997). Generell wird in PISA eine hohe interkulturelle Äquivalenz sowohl der Tests als auch der Fragebogenskalen erzielt (vgl. OECD, 2005). Gleichzeitig deuten aber verschiedene Sekundäranalysen darauf hin, dass ein Teil der internationalen Varianz in den Mittelwerten der Fragebogenskalen durch methodische Artefakte wie etwa kulturbedingte Unterschiede im Antwortverhalten erzeugt wird (vgl. van de Gaer, Grisay, Schulz & Gebhardt, 2012). Aufschlussreichere Informationen für das Monitoring nicht-kognitiver Prozesse und Outcomes können insofern Trendanalysen innerhalb eines Landes erbringen, da diese von kulturellen Einflüssen auf die Definition und Messung der Konstrukte nicht betroffen sind. Deshalb ist es für die Nutzung der PISA-Daten zum Systemmoni-

von Bildungsaspirationen (Dupriez, Monseur, van Campenhoudt & Lafontaine, 2012) oder Moderationseffekte kultureller Wertorientierungen auf den Zusammenhang zwischen schulischen Prozessfaktoren, wie der Lehrerunterstützung, und Schülerleistungen (Vieluf, 2012).

Die Breite der wissenschaftlichen Erkenntnisse, die mit PISA verbunden sind, kann exemplarisch anhand von Vorträgen bei der ersten PISA-Forschungs-Konferenz (Prenzel, Kobarg, Schöps & Rönnebeck, 2013)³ dargestellt werden. Während einige der Studien auf international vergleichenden Analysen basieren, beschreiben andere nur Ergebnisse für einzelne Staaten.

- In Staaten mit niedrigem Leseengagement ist die Chancengleichheit geringer als in Staaten, in denen Schülerinnen und Schüler mehr lesen und mehr Freude am Lesen berichten (Baye, Monseur & Lafontaine, 2009).
- Bei der Förderung naturwissenschaftlicher Kompetenz und des naturwissenschaftlichen Interesses von Schülerinnen und Schülern scheint ein fokussiertes Muster naturwissenschaftlichen Lernens erfolgversprechender als eine breite Vielfalt naturwissenschaftlicher Aktivitäten (Kobarg, Seidel, Prenzel, McCrae & Walker, 2009).
- Das Interesse von Schülerinnen und Schülern an einer Karriere im Bereich der Naturwissenschaften hängt zusammen mit dem Ausmaß, in dem Schulen auf solche Karrieren vorbereiten (Lie & Kjaernsli, 2009).
- In Schweizer Schulen korrelieren Unterrichtszeit und Mathematikleistung signifikant – außer in Gymnasien (Angelone, Moser & Ramseier, 2009).
- Während japanische Schülerinnen und Schüler großes Interesse an entdeckendem Lernen zeigen, ist der Oberstufenunterricht in Japan nicht auf diese Interessen zugeschnitten. Dieses Ergebnis könnte möglicherweise zur Erklärung des vergleichsweise geringen Interesses an Naturwissenschaften in Japan beitragen (Yasushi, 2009).
- In Südkorea finden sich Zusammenhänge der Schülerleistungen mit den Ressourcen einer Schule sowie mit dem Ausmaß, in dem die Schule Schülerinnen und Schüler ermutigt, ihr volles Potential zu nutzen (Kaplan, 2009a).

Für die wissenschaftliche Grundlagenforschung von besonderer Bedeutung sind Erkenntnisse zur Validität zentraler Konstrukte, wie etwa die Arbeit von Marsh, Hau, Artelt, Baumert & Peschar (2006) zur interkulturell äquivalenten Erfassung von Aspekten der Selbstregulation bei PISA 2000 oder die Analysen von Drechsel, Carstensen und Prenzel (2011) zur Dimensionalität des Interesses an Naturwissenschaften, basierend auf PISA 2006.

Wenngleich PISA einen Beitrag dazu leisten kann, Wissen über Bildungspolitik und -praxis zu kreieren, müssen dabei wichtige Grenzen beachtet werden. Eine Einschränkung ist, dass PISA den Ertrag von Bildungsprozessen misst, also Fähigkeiten und Kompetenzen, die seit der frühen Kindheit beim Durchlaufen verschiedener Schulen

3 Zusammenfassungen aller Präsentationen, die im Folgenden zitiert werden, sind erhältlich auf <http://www.pisaresconf09.org>

bis zum Alter von 15 Jahren akkumuliert worden sind. Damit kann PISA nicht ermitteln, wie viel Lernen tatsächlich in der von den Schülerinnen und Schülern mit 15 Jahren besuchten Schule der Sekundarstufe stattgefunden hat. Nur wenn das Leistungsniveau zum Zeitpunkt des Eintritts in die aktuelle Schule erfasst worden wäre, könnte der Fortschritt (*value-added*) in der Leistung mit den bildungsbezogenen Erfahrungen in der spezifischen Schule in Bezug gesetzt werden. Solche Messungen der Leistung zu einem Zeitpunkt vor der PISA-Erhebung stehen jedoch nicht zur Verfügung. Auch der Einfluss von Lehrerinnen und Lehrern auf Schülerleistungen kann anhand von PISA nicht beurteilt werden – zumindest nicht mit dem Design, das während des ersten Jahrzehnts auf internationaler Ebene implementiert worden ist (zu den deutschen Erweiterungen vgl. Klieme, Artelt et al., 2010). Der Grund hierfür ist, dass lediglich eine Zufallsstichprobe Fünfzehnjähriger in jeder Schule gezogen wird, anstelle von vollständigen Klassen, was die Erhebung von Informationen über Unterrichtsstrategien und Unterrichtsqualität in einer bestimmten Klasse ausschließt. Lediglich Deutschland untersucht regelmäßig auch vollständige Klassen (vgl. die Beiträge des vorliegenden Bandes zu Befunden aus der Jahrgangsstufe 9). Schließlich muss berücksichtigt werden, dass etwa in einem Fünftel der Staaten eine Mehrheit der Schülerinnen und Schüler erst kurz vor der Erhebung auf eine weiterführende Schule gewechselt ist. Dies erschwert einmal mehr Schlüsse über Schuleffekte in den entsprechenden Staaten.

Weiterhin muss bei der Interpretation der Ergebnisse berücksichtigt werden, dass es generell ausgesprochen schwierig ist, auf der Basis einer Survey-Studie mit Leistungsdaten zu belegen, dass eine bestimmte Maßnahme oder Praxis einen direkten Einfluss oder eine tatsächliche Wirkung auf Schülerleistungen hat (Gustafsson, 2007; Kaplan, 2009b). Wird etwa beobachtet, dass Schulen höhere Schülerleistungen aufzuweisen haben, wenn sie ihre Schulevaluationsdaten veröffentlichen, so erlaubt diese querschnittlich ermittelte Korrelation keine Schlüsse darüber, ob die Veröffentlichung zu einer höheren Leistung beiträgt oder ob es vielmehr vor allem erfolgreiche Schulen sind, die ihre Ergebnisse veröffentlichen. Wenn die OECD aus dem beobachteten Zusammenhang die politische Schlussfolgerung zieht, externes Monitoring könne den Leistungsstand einer Schule verbessern (OECD, 2007, S. 276), liegt also eine wissenschaftlich unzulässige Interpretation vor. Auch wenn komplexe statistische Verfahren zur Analyse von Kausaleffekten eingesetzt werden (z.B. *Matching*-Verfahren; siehe dazu Heckman, 2005; Kaplan, 2009b; Schneider, Carnoy, Kilpatrick, Schmidt & Shavelson, 2007), würde eine solche Interpretation voraussetzen, dass alle Variablen, die potentiell mit der Auswahl des Treatments sowie mit dem interessierenden Ergebnis zusammenhängen, ebenfalls erfasst worden sind. Selbst im Rahmen einer so groß angelegten Untersuchung wie PISA ist dies schwerlich umzusetzen.

Der Wert von PISA ergibt sich nicht zuletzt aus dem kontinuierlichen Zusammenspiel mit anderen Untersuchungen. Auf der einen Seite kann mithilfe von Längsschnittstudien, experimentellen Untersuchungen und Interventionsstudien überprüft werden, inwieweit in PISA beobachtete Korrelationen tatsächlich auf einem Kausalzusammenhang beruhen. Auf der anderen Seite können auf der Grundlage der aktuellen Forschungsliteratur jeweils jene Variablen als Indikatoren für das Bildungsmonitoring aus-

gewählt werden, deren Bedeutung für Bildungsqualität und -effizienz in vorherigen Untersuchungen, im Idealfall in Längsschnitt- oder Interventionsstudien, belegt worden ist.

Das Konzept der Wirksamkeit wird hier in dem technischen Sinne verwendet, der Kausalitätsanalysen in der ökonomischen und psychologischen Forschung zugrunde liegt, ausgehend von den wissenschaftstheoretischen und statistischen Ansätzen bei Donald Rubin (1974). In der Perspektive eines akteurszentrierten Institutionalismus, wie ihn – aufbauend auf internationaler soziologischer Forschung – Helmut Fend (2006) in seiner Neuen Theorie der Schule eingeführt hat, wären solche Ursache-Wirkungs-Gefüge dahingehend zu hinterfragen, wie Regelungen und Vorgaben, einschließlich systematischer „Interventionen“, auf allen beteiligten Ebenen rekontextualisiert werden. Hierzu kann eine quantitative Studie wie PISA nur begrenzt beitragen. Immerhin lässt sich im Rahmen des mehrperspektivischen Fragebogen-Designs systematisch untersuchen, wie bestimmte Phänomene wie etwa Unterrichtspraktiken, Kooperation und Partizipation von unterschiedlichen Akteursgruppen an einer Schule wahrgenommen werden. In den Planungen und Vorstudien für PISA 2015 werden solche Ansätze derzeit erprobt.

Auf lange Sicht wird die Verfügbarkeit von Trend-Daten der größte Nutzen der PISA-Datenbank sein. Querschnittsstudien tragen nur wenig zum Verständnis der Entstehung von Outcomes bei, weil es praktisch unmöglich ist, alle relevanten Variablen abzudecken – inklusive Informationen über die Schülerleistung bei Eintritt in die Sekundarstufe. Demgegenüber erlauben Veränderungen in Outcomes auf der Schul- oder Schülerebene, aber auch solche auf der Ebene von Staaten (vgl. Gustafsson, 2007; Hanushek & Wößmann, 2010), substantiellere Schlussfolgerungen – vor allem wenn sie mit Veränderungen von Inputs und Prozessen in Beziehung gesetzt werden können. Um dieses Ziel zu erreichen, muss PISA einen Variablensatz definieren, der über die Erhebungszyklen konstant gehalten wird. Es wäre außerdem aufschlussreich, wenn im Rahmen von PISA zunehmend auch längsschnittliche Designs auf Schul- und Schülerebene implementiert würden, wie es in Deutschland mehrfach geschehen ist (vgl. etwa Bischof et al., in diesem Band).

5. Fazit

Der vorliegende Beitrag hatte zum Ziel, die Motive, Aussagemöglichkeiten und Grenzen der PISA-Studie sowie ihren praktischen und theoretischen Hintergrund zusammenfassend darzustellen. Die Darlegungen spiegeln den internationalen Arbeitsstand bei der Entwicklung der Rahmenkonzeption, die den Kontext der Leistungsmessungen abdeckt, also die relevanten Input-, Prozess- und Outcome-Merkmale auf Schüler-, Klassen- und Schulebene.

Eine zentrale Aufgabe von PISA ist es, Indikatoren für das international vergleichende Monitoring von Bildungssystemen bereitzustellen. Gleichzeitig dient die Studie aber dem Aufbau von Wissen darüber, *warum* spezifische Merkmalskonstellationen in verschiedenen Bildungssystemen vorgefunden werden. Die Auswahl der Merkmale

ist von politischen Interessen geleitet, beruht aber andererseits auf Erkenntnissen der empirisch-pädagogischen Forschung und benachbarter Disziplinen, insbesondere der Schuleffektivitätsforschung und der Lehr-Lern-Forschung in den Testdomänen (Lesen, Mathematik, Naturwissenschaften und Problemlösen).

Bildungsmonitoring und Wirkungsanalyse: Diese beiden Modi der Gestaltung und Interpretation charakterisieren internationale Schulstudien schon seit mehr als einem halben Jahrhundert. Ihre Koexistenz führt immer wieder zu Missverständnissen und Kurzschlüssen, sowohl bei Autoren als auch bei Kritikern des Paradigmas. Um Überinterpretationen und unangemessene Bedeutungszuschreibungen zu vermeiden, ist es von zentraler Bedeutung, die Aussagemöglichkeiten und Einschränkungen der Studien zu bestimmen. Beispielsweise jene, dass sich Wirkungsfragen aufgrund einer querschnittlichen Survey-Studie kaum beantworten lassen.

Im vorliegenden Aufsatz wurde versucht, die Chancen und Grenzen von PISA als Instrument des Bildungsmonitorings und als wissenschaftliche Datenquelle zu beleuchten unter Hinweis auf Literatur, die entweder in PISA aufgegriffen wird oder umgekehrt PISA-Daten für Forschungszwecke verwendet. Der Aufsatz kann daher als Dokumentation der Selbstreflexion von Wissenschaftlern angesehen werden, die bei PISA engagiert sind. Eine externe Bewertung der Studie, die in der deutschen Erziehungswissenschaft an prominenter Stelle rezipiert wurde, hat Münch (2009) vorgelegt. Er charakterisiert PISA als „Benchmarking der OECD-Mitgliedstaaten in der Erfüllung global definierter Bildungsstandards“ unter der „Leitidee der Bildung als Humankapital“ (S. 31 und 33) und bewertet den Erkenntnisgewinn als „eher bescheiden“ (S. 220). Es sei dem Leser überlassen, ein eigenes Fazit zu bilden.

Literatur

- Angelone, D., Moser, U., & Ramseier, E. (2009). *Instruction time and performance. Analyses of the importance of instruction time for competencies in Science and Mathematics based on Swiss PISA 2006 Data* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/context/room3/Angelone_Moser_Ramseier.pdf
 [20.12.2012].
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O., & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Baye, A., Monseur, C., & Lafontaine, D. (2009). *Institutional and socioeconomic factors influencing the reading engagement* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/content/room1/Baye_Monseur_Lafontaine_2.pdf
 [20.12.2012].
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blum, W., & Leiss, D. (2005). How Do Students and Teachers Deal with Mathematical Modeling Problems? The Example “Sugarloaf”. *ICTMA 12 Proceedings*, 222-231.

- Böttcher, W., Bos, W., Döbert, H., & Holtappels, H. G. (Hrsg.) (2008). *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive*. Münster: Waxmann.
- Coll, R. K., Dahsah, C., & Faikhamta, C. (2010). The Influence of Educational Context on Science Learning: A Cross-National Analysis of PISA. *Research in Science & Technological Education*, 28(1), 3-24.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice, and Theory in Contemporary Schools*. London/New York: Routledge.
- Drechsel, B., Carstensen, C., & Prenzel, M. (2011). The Role of Content and Context in PISA Interest Scales: A Study of the Embedded Interest Items in the PISA 2006 Science Assessment. *International Journal of Science Education*, 33(1), 73-95.
- Dupriez, V., Monseur, C., van Campenhoudt, M., & Lafontaine, D. (2012). Social Inequalities of Post-secondary Educational Aspirations: Influence of Social Background, School Composition and Institutional Context. *European Educational Research Journal*, 11(4), 504-519.
- Fend, H. (2006). *Neue Theorie der Schule*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gustafsson, J.-E. (2007). Understanding Causal Influences on Educational Achievement through Analysis of Differences over Time within Countries. In T. Loveless (Hrsg.), *Lessons Learned: What International Assessments Tell Us about Math Achievement* (S. 37-63). Washington: The Brookings Institution.
- Hanushek, E. A., & Wößmann, L. (2010). *The Economics of International Differences in Educational Achievement* (IZA Discussion Paper No. 4925). Bonn: IZA.
- Heckman, J. J. (2005). The Scientific Model of Causality. *Sociological Methodology*, 35, 1-97.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in Cross-cultural Psychology: A Review and Comparison of Strategies. *Journal of Cross-Cultural Psychology*, 16, 131-152.
- Kaplan, D. (2009a). *Advances in Multilevel Latent Variable Models for PISA Data* (paper presented at the PISA Research Conference, Kiel, 14-16 September 2009). http://www.pisaresconf09.org/user_uploads/files/methodology/room1/Kaplan.pdf [20.12.2012].
- Kaplan, D. (2009b). Causal Inference in Non-experimental Educational Policy Research. In D. N. Plank, W. E. Schmidt & G. Sykes (Hrsg.), *AERA Handbook on Education Policy Research* (S. 139-153). Washington: American Educational Research Association.
- Keeves, J. P., & Lietz, P. (2011). The Relationship of IEA to Some Developments in Educational Research Methodology and Measurement during the Years from 1962 to 1992. In C. Papanastasiou, T. Plomp & E. C. Papanastasiou (Hrsg.), *IEA 1958-2008: 50 Years of Experiences and Memories* (S. 217-252). Nicosia: Cultural Center of the Kykkos Monastery.
- Klieme, E. (2011). Bildung unter undemokratischem Druck? Anmerkungen zur Kritik der PISA-Studie. In S. Aufenanger, S. F. Hamburger, L. Ludwig & R. Tippelt (Hrsg.), *Bildung in der Demokratie. Beiträge zum 22. Kongress der DGfE*, Band 2 (S. 289-302). Opladen/Farmington Hills: Barbara Budrich.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E., Backhoff-Escudero, E., Blum, W., Buckley, J., Hong, Y.-Y., Kaplan, D., Levin, H. M., & Vieluf, S. (2010). *Designing PISA as a Sustainable Data Base for Educational Policy and Research. The PISA 2012 Context Questionnaire Framework*. Paris: OECD.
- Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im empirischen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (Zeitschrift für Erziehungswissenschaft, Sonderheft 8, S. 11-29).
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Classrooms. In T. Janik & T. Seidel (Hrsg.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom* (S. 137-160). Münster: Waxmann.

- Kobarg, M., Seidel, T., Prenzel, M., McCrae, B., & Walker, M. (2009). *Patterns of Science Teaching and Learning in an International Comparison* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/context/room3/Kobarg_etal.pdf [20.12.2012].
- Lie, S., & Kjaernsli, M. (2009). *Predictors for Students' Choice of Science Careers* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/context/room3/Lie_Kjaernsli.pdf [20.12.2012].
- Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's Brief Self-report Measure of Educational Psychology's Most Useful Affective Constructs: Cross-cultural, Psychometric Comparisons Across 25 Countries. *International Journal of Testing*, 6, 311-360.
- Münch, R. (2009). *Globale Eliten, lokale Autoritäten: Bildung und Wissenschaft unter dem Regime von PISA*, McKinsey & Co. Frankfurt a.M.: Suhrkamp.
- OECD (Hrsg.) (2005). *PISA 2003 Technical Report*. Paris: OECD.
- OECD (Hrsg.) (2007). *PISA 2006 – Science Competencies for Tomorrow's World*. Paris: OECD.
- OECD (Hrsg.) (2011). *Education at a Glance*. Paris: OECD.
- Ozga, J. (2008). Governing Knowledge: Research Steering and Research Quality. *European Educational Research Journal*, 7(3), 261-272.
- Ozga, J. (2012). Assessing PISA. *European Educational Research Journal*, 11(2), 166-171.
- Prenzel, M., Kobarg, M., Schöps, K., & Rönnebeck, S. (2013). *Research on PISA. Research Outcomes of the PISA Research Conference 2009*. New York: Springer.
- Purves, A. C. (1987). The Evolution of the IEA: A Memoir. *Comparative Education Review*, 31(1), 10-28.
- Roth, H. (1971). *Pädagogische Anthropologie. Band 2: Entwicklung und Erziehung*. Hannover: Schroedel.
- Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66, 688-701.
- Scheerens, J. (2000). *Improving School Effectiveness* (Fundamentals of Educational Planning Series, IIEP, Vol. 68). Paris: UNESCO.
- Schmidt, W. H., & Maier, A. (2009). Opportunity to Learn. In G. Sykes, B. L. Schneider & D. N. Plank (Hrsg.), *Handbook on Education Policy Research* (S. 541-549). New York: Routledge.
- Schneider, B. L., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating Causal Effects: Using Experimental and Observational Designs*. Washington: American Educational Research Association.
- Schoenfeld, A. (2006). Mathematics Teaching and Learning. In P. A. Alexander & I. H. Winne (Hrsg.), *Second Handbook of Educational Psychology* (S. 479-510). London: Routledge.
- Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating Productive Mathematical Discussions: Five Practices for Helping Teachers Move Beyond Show and Tell. *Mathematical Thinking and Learning*, 10(4), 313-340.
- Sykes, G., Schneider, B. L., & Plank, D. N. (2009). *Handbook of Education Policy Research*. New York: Routledge.
- Teddle, C., & Reynolds, D. (Hrsg.) (2000). *The International Handbook of School Effectiveness Research*. New York: Routledge.
- van de Gaer, E., Grisay, A., Schulz, W., & Gebhardt, E. (2012). The Reference Group Effect: An Explanation of the Paradoxical Relationship Between Academic Achievement and Self-Confidence Across Countries. *Journal of Cross-Cultural Psychology*, 43(8), 1205-1228.
- van de Vijver, F. J. R., & Leung, K. (1997). *Methods and Data Analysis for Cross-cultural Research*. Newbury Park: Sage.
- van de Vijver, F. J. R., van Hemert, D. A., & Poortinga, Y. H. (Hrsg.) (2008). *Multilevel Analysis of Individuals and Cultures*. New York: Erlbaum.

- Vieluf, S. (2012). *Links between Student Achievement and Teacher Support in Different Education Systems* (Vortrag bei der Arbeitstagung Methoden des interkulturellen Vergleichs, unveröffentlichtes Manuskript). Frankfurt a.M.: DIPF.
- Walberg, H. J. (1986). Syntheses of Research on Teaching. In M. C. Wittrock (Hrsg.), *Handbook of Research on Teaching* (3. Aufl., S. 214-229). New York: Macmillan.
- Walker, D. F. (1976). Toward Comprehension of Curricular Realities. *Review of Research in Education*, 4, 268-308.
- Yasushi, O. (2009). *Comparison of Attitudes toward Science between Grade 9 and 10 Japanese Students by Using the PISA Questions and Its Implications on Science Teaching in Japan* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009). http://www.pisaresconf09.org/user_uploads/files/content/room2/OGURA_Yasushi.pdf [20.12.2012].

Anschrift des Autors/der Autorin

Prof. Dr. Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: klieme@dipf.de

Dipl.-Psych. Svenja Vieluf, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: vieluf@dipf.de