pedocs
Open Access Erziehungswissenschaften

DIPF

Mildner, Dorothea; Hochweber, Jan; Frey, Andreas

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

Jude, Nina [Hrsg.]; Klieme, Eckhard [Hrsg.]: PISA 2009 - Impulse für die Schul- und Unterrichtsforschung. Weinheim u.a. : Beltz 2013, S. 151-171. - (Zeitschrift für Pädagogik, Beiheft; 59)

urn:nbn:de:0111-opus-78256

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 genützte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet
zur Gebührenzahlung an die VGWort, 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 bildungs-
standardbezogene 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

Dorothea Mildner/Jan Hochweber/Andreas Frey

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

1. Einleitung

Das *Programme for International Student Assessment* (PISA) wurde im Jahr 2009 zum vierten Mal durchgeführt. Die bei den ersten PISA-Erhebungen zentralen querschnittlichen Betrachtungen können somit inzwischen durch differenzierte Trendanalysen in ihrer Aussagekraft bedeutend erweitert werden. Bei PISA 2009 bildeten Trendanalysen bereits einen zentralen Bestandteil der regulären PISA-Berichterstattung (vgl. Klieme, Jude, Baumert & Prenzel, 2010; OECD, 2010).

Trendanalysen verfolgen zwei grundlegende Ziele. Das erste Ziel besteht im sogenannten *Benchmarking*. Beim Benchmarking werden Veränderungen der Kompetenzstände von Schülerinnen und Schülern zwischen den an PISA teilnehmenden Staaten (beziehungsweise Bildungssystemen) verglichen. Hierbei kann beispielsweise analysiert werden, inwieweit Veränderungen der mittleren Kompetenzen und gegebenenfalls deren Streuungen in verschiedenen Staaten gleich oder unterschiedlich ausfallen. Allerdings sind Trendanalysen für einzelne Teilnehmerstaaten nicht erst durch den Vergleich mit anderen Staaten interessant. Das zweite generelle Ziel, das *Monitoring*, betrachtet deshalb Veränderungen von Kompetenzverteilungen innerhalb eines Staates über die Zeit. Hierbei werden die Veränderungen häufig mit als relevant erachteten Bedingungsfaktoren für den Kompetenzerwerb in Beziehung gesetzt. So können Trendanalysen zum Beispiel Hinweise auf den Erfolg von Reformmaßnahmen im Bildungsbereich geben. Zudem gestatten die bei PISA verwendeten Kompetenzskalen, dass Messwerte kriteriumsorientiert und nicht nur normorientiert interpretiert werden können. Während bei normorientierten Testwertinterpretationen nur relative Vergleiche mit Referenzgruppen wie anderen Staaten möglich sind, erlauben kriteriumsorientierte Testwertinterpretationen Aussagen darüber, was Schülerinnen und Schüler mit einer gegebenen Wahrscheinlichkeit inhaltlich wissen und können (vgl. Herzberg & Frey, 2011). Damit gewinnen Trendaussagen auch jenseits des Staatenvergleichs an Aussagekraft, etwa dahingehend, inwieweit es einem Teilnehmerstaat gelungen ist, den Anteil sehr schwacher Leserinnen und Leser über einen gewissen Zeitraum hinweg zu reduzieren (z.B. Naumann, Artelt, Schneider & Stanat, 2010).

Eine nationale Perspektive auf Trends ist für Deutschland aus einem weiteren Grund von besonderem Interesse. In Deutschland wird bereits seit der ersten PISA-Erhebung im Jahr 2000 neben der altersbasierten Stichprobe von fünfzehnjährigen Schülerinnen und Schülern eine jahrgangsstufenbasierte Stichprobe von Neuntklässlerinnen und Neuntklässlern erhoben (vgl. Abschnitt 2). Da jahrgangsstufenbasierte Stichproben je-

weils nur in einigen Staaten untersucht wurden, bezog sich die internationale Trendberichterstattung bislang ausschließlich auf Trends für die fünfzehnjährigen Schülerinnen und Schüler. Eine Betrachtung von Trends für die Schülerinnen und Schüler der neunten Jahrgangsstufe ermöglicht darüber hinaus jedoch sehr interessante und neue Erkenntnisse. Es sind keineswegs stets parallele Kompetenzverläufe über die Zeit für Fünfzehnjährige und Neuntklässlerinnen und Neuntklässler zu erwarten. So könnten beispielsweise Veränderungen der Bildungsbeteiligung, des mittleren Alters von Neuntklässlerinnen und Neuntklässlern, der Anteile von Jugendlichen, die eine Klasse wiederholt haben, oder spezifische Veränderungen der Kompetenzen von Subpopulationen zu Differenzen zwischen den Trendverläufen von Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern führen. Derartige Differenzen sind für ein umfassendes Verständnis der Funktionsweise von Bildungssystemen hoch relevant.

Im vorliegenden Beitrag geht es entsprechend darum, die Trendberichterstattung in Deutschland, die sich bisher auf die Betrachtung der Trends bei den fünfzehnjährigen Schülerinnen und Schülern beschränkt, um die Betrachtung von Trends für die neunte Jahrgangsstufe zu erweitern. Dabei werden die Trends der Kompetenzen im Lesen, in der Mathematik und in den Naturwissenschaften zwischen PISA 2003 und PISA 2009 in den beiden Stichproben verglichen. Darüber hinaus soll exemplarisch versucht werden, mögliche Ursachen für unterschiedliche Trends in den Stichproben zu identifizieren. Dabei beschränken wir uns auf die Betrachtung der Veränderung in der Jahrgangsstufenverteilung fünfzehnjähriger Schülerinnen und Schüler.

Im Folgenden wird zunächst dargestellt, worin sich eine lebensalterbasierte Populationsdefinition – wie sie bei der internationalen PISA-Erhebung vorgenommen wird – und eine jahrgangsstufenbasierte Populationsdefinition – wie sie der nationalen Stichprobe der neunten Klassen in Deutschland zugrunde liegt – unterscheiden. Um zu prüfen, inwieweit die Jahrgangsstufenverteilung fünfzehnjähriger Schülerinnen und Schüler unterschiedliche Trends in den beiden PISA-Stichproben erklären könnte, wird die Veränderung dieser Verteilung zwischen PISA 2003 und PISA 2009 betrachtet. Auf die daran anschließende genauere Erläuterung der Zielsetzung dieses Beitrags folgt im Methodenteil eine Darstellung der analysierten PISA-Stichproben zwischen PISA 2003 und PISA 2009, der Datengrundlage sowie der statistischen Verfahren, die für die Vergleiche zwischen den Erhebungszeitpunkten eingesetzt wurden. Im Anschluss werden die Ergebnisse der Trendanalysen für die Fünfzehnjährigen und die Neuntklässlerinnen und Neuntklässler berichtet und mögliche Beziehungen zu Veränderungen in der Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen untersucht. Abschließend erfolgen ein Fazit der Ergebnisse sowie ein kurzer Ausblick.

2. Lebensalterbasierte und jahrgangsstufenbasierte Stichproben

Eine zentrale Aufgabe für eine internationale Schulleistungsstudie wie PISA ist es, verlässliche internationale Vergleiche von Kompetenzständen zu gewährleisten (Baumert & Stanat, 2006). Ein grundlegender Aspekt ist dabei die Definition der interessierenden

Zielpopulation, auf die die Ergebnisse generalisiert werden sollen. Angesichts der Organisation der meisten Bildungssysteme liegt es intuitiv nahe, eine jahrgangsstufenbasierte Definition der Zielpopulation vorzunehmen. Dabei werden in jedem Staat Schülerinnen und Schüler untersucht, die eine bestimmte Jahrgangsstufe besuchen. Allerdings führt gerade diese Vorgehensweise oftmals zu erheblichen Schwierigkeiten. International vergleichbare Jahrgangsstufen sind häufig schwierig zu identifizieren, da sich Staaten in der Art und dem Ausmaß an vorschulischer Bildung und Erziehung, im Schuleintrittsalter sowie in der institutionellen Struktur ihrer nationalen Bildungssysteme teilweise erheblich unterscheiden.

Folglich definieren internationale Schulleistungsvergleiche typischerweise ihre Zielpopulation über ein bestimmtes Lebensalter. Einige internationale Studien (z.B. Progress in International Reading Literacy Study – PIRLS; Third International Mathematics and Science Study – TIMSS) haben dabei ihre Zielpopulation auf Basis von Jahrgängen definiert, die eine maximale Abdeckung einer bestimmten Alterskohorte bieten. Ein Nachteil dieses Ansatzes liegt darin, dass geringe Variationen in der Altersverteilung zwischen den Jahrgangsstufen zur Auswahl unterschiedlicher Zieljahrgangsstufen in den Teilnehmerstaaten oder gar zwischen Bildungssystemen innerhalb von Staaten führen können. Kompetenzvergleiche können auch dadurch erschwert werden, wenn nicht alle Schülerinnen und Schüler des interessierenden Alters in den jahrgangsbasierten Stichproben repräsentiert sind und sich die nicht repräsentierten Schülerinnen und Schüler typischerweise in der nächsthöheren Jahrgangsstufe in einigen Staaten und in der nächstniedrigeren Jahrgangsstufe in anderen Staaten befinden. Dieser Umstand würde im ersten Fall Schülerinnen und Schüler auf potenziell höheren Kompetenzstufen, im zweiten Fall solche auf potenziell niedrigeren Kompetenzstufen ausschließen.

Um derartige Probleme zu vermeiden, wird bei PISA in Übereinstimmung mit dem zugrunde liegenden *Literacy*-Konzept[1] (z.B. Baumert, Stanat & Demmrich, 2001) eine lebensalterbasierte Definition der Zielpopulation genutzt. Es wird damit eine Definition verwendet, die in geringerem Ausmaß an die institutionellen Strukturen nationaler Bildungssysteme gebunden ist: Unabhängig von der besuchten Jahrgangsstufe oder der Art der Bildungseinrichtung werden Schülerinnen und Schüler untersucht, die zu Beginn des Testzeitraumes fünfzehn Jahre alt sind. Mit dieser Populationsdefinition macht PISA Aussagen über das Wissen und die Fähigkeiten einer Gruppe von Individuen, die innerhalb einer vergleichbaren Referenzperiode geboren wurden, die aber durchaus unterschiedliche Bildungserfahrungen innerhalb und außerhalb der Schule gemacht haben können. Bei PISA werden dieses Wissen und diese Fähigkeiten als Bildungserträge im Alter von fünfzehn Jahren interpretiert. In Abhängigkeit von den nationalen Regelungen der Staaten zum Schuleintritt, zur Schulwahl und zum Schulsystem können diese Schü-

1 Ausgangspunkt der Rahmenkonzeption von PISA ist eine Vorstellung von Grundbildung (engl. literacy), bei der es um die Fähigkeit geht, Kenntnisse und Fertigkeiten in einer Vielzahl von Situationen anzuwenden und zu analysieren, logisch zu denken und in effektiver Weise zu kommunizieren. Innerhalb der bei PISA untersuchten Domänen (Lesen, Mathematik, Naturwissenschaften) beruht Literacy auf bereichsspezifischem Wissen.

lerinnen und Schüler je nach Bildungssystem und Bildungszweig auf unterschiedliche Jahrgangsstufen verteilt sein. Solche Verschiedenheiten müssen berücksichtigt werden, wenn PISA-Ergebnisse zwischen Teilnehmerstaaten verglichen werden. Sind die Kompetenzwerte in den bei PISA untersuchten Bereichen Lesen, Mathematik und Naturwissenschaften in einem Staat signifikant höher als in einem anderen Staat, so kann daraus nicht automatisch geschlossen werden, dass die Schulen oder bestimmte Teile des Bildungssystems im ersten Staat effektiver sind als im zweiten. Es kann jedoch geschlossen werden, dass der kumulative Bildungsertrag bis zum Alter von fünfzehn Jahren im ersten Staat höher ist als im zweiten.

Bei einer altersbasierten Definition der Zielpopulation muss allerdings in Kauf genommen werden, dass die Möglichkeiten eingeschränkt sind, gefundene Unterschiede auf institutionelle Bedingungen zurückzuführen. So spielt die Zuordnung der Schülerinnen und Schüler zu Schulklassen oder Kursen – welche für die Institution Schule ein zentrales Charakteristikum und insbesondere für die Organisation von Unterricht von überragender Bedeutung ist – für die Stichprobengewinnung und letztlich für die Analysen im Rahmen der internationalen PISA-Erhebung keine Rolle. Um Staaten entgegenzukommen, die jahrgangsbasierte Ergebnisse im Rahmen nationaler Analysen anstreben, bietet das Design von PISA eine Option zur nationalen Stichprobenerweiterung, welche die altersbasierte Stichprobe um eine jahrgangsbasierte Stichprobe ergänzt. Von dieser Möglichkeit wurde in Deutschland bereits bei PISA 2000 Gebrauch gemacht. Allerdings wurden dort keine vollständigen Klassen als Stichprobeneinheiten gezogen, sondern zusätzlich zu den bereits in der Stichprobe der Fünfzehnjährigen enthaltenen Schülerinnen und Schülern aus neunten Klassen zehn weitere nicht fünfzehnjährige Neuntklässlerinnen und Neuntklässler zufällig ausgewählt, um so für diese Jahrgangsstufe auch die übrigen in ihr enthaltenen Altersgruppen zu repräsentieren (vgl. Sibberns & Baumert, 2001, S. 511). Seit PISA 2003 wird an den für den internationalen Vergleich ausgewählten Schulen die international verbindliche Stichprobe der Fünfzehnjährigen so ergänzt, dass dort zusätzlich jeweils zwei komplette Klassen der neunten Jahrgangsstufe erhoben werden. Die Ziehung der nationalen Stichproben der neunten Jahrgangsstufe wird somit erst ab PISA 2003 in identischer Weise vorgenommen.

3. Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen

Die wiederholte Erhebung von Kompetenzen bei Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern ermöglicht es, Unterschiede in den Trendverläufen zwischen beiden Stichproben zu beschreiben. Darüber hinaus können diese Unterschiede mit Merkmalen der Stichproben in Beziehung gesetzt werden, was im vorliegenden Beitrag exemplarisch für die Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen geschehen soll. Voraussetzungen für einen Erklärungsbeitrag der Jahrgangsstufenverteilung wären, dass (a) sich diese Verteilung im betrachteten Zeitraum in der Zielpopulation der Fünfzehnjährigen wesentlich verändert hat und (b) für den Fall einer solchen

Veränderung bedeutsame Unterschiede in den Kompetenzen der Fünfzehnjährigen zwischen den PISA-Erhebungen zu erwarten sind.

Zumindest zwei Entwicklungen im deutschen Bildungssystem sprechen dafür, dass der Besuch höherer Jahrgangsstufen durch die fünfzehnjährigen Schülerinnen und Schüler seit dem Beginn der PISA-Erhebungen zugenommen hat (Klieme et al., 2010). So hat in den Alterskohorten, aus denen sich die Teilnehmerinnen und Teilnehmer der PISA-Erhebungen rekrutierten, der Anteil vorzeitiger Einschulungen erheblich zugenommen, während der Anteil von Zurückstellungen bei der Einschulung kontinuierlich abgenommen hat. Auch die Anzahl der Klassenwiederholungen ging gemäß der amtlichen Schulstatistik, wenn auch nur im Primarbereich, in dem für die PISA-Teilnehmerinnen und -Teilnehmer relevanten Zeitraum von 1995 bis 2004 (vgl. Klieme et al., 2010) leicht zurück. Die prozentualen Anteile der fünfzehnjährigen Schülerinnen und Schüler in den verschiedenen Jahrgangsstufen in den PISA-Erhebungen 2003 bis 2009 werden in Tabelle 1 berichtet.[2] Ebenfalls dargestellt sind die Differenzen der Prozentwerte zwischen den einzelnen PISA-Erhebungen.

Jahrgangs-stufe	PISA 2003	PISA 2006	PISA 2009	Diff 06–03	Diff 09–06	Diff 09–03
7	1.7	1.6	1.2	–0.1	–0.3	–0.5
8	15.0	12.3	11.0	–2.6	–1.4	–4.0
9	59.9	56.5	54.8	–3.4	–1.8	–5.2
10	23.2	29.3	32.5	6.0	3.3	9.3
11	0.1	0.3	0.4	0.2	0.2	0.3
12			< 0.1		< 0.1	< 0.1

Tab. 1: Prozentuale Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen in den PISA-Erhebungen 2003 bis 2009 sowie die Differenzen (Diff) der Prozentwerte zwischen den PISA-Erhebungen (gewichtet), Werte gerundet

Die prozentualen Veränderungen stützen die erwartete deutliche Verschiebung hin zum Besuch höherer Jahrgangsstufen; die Anteile in den Jahrgangsstufen 8 und 9 nehmen zusammen um gut neun Prozent ab, der Anteil in Jahrgangsstufe 10 um nahezu den glei-

2 Es ließe sich einwenden, dass die Ermittlung der Verteilung basierend auf den PISA-Stichproben weniger aussagekräftig sei als der Rückgriff auf die Daten der amtlichen Schulstatistik. Allerdings beruhen Stichprobenziehung und Berechnung der Gewichte der Schülerinnen und Schüler bei PISA auf den jeweils aktuellen zum Zeitpunkt der Stichprobenziehung verfügbaren Angaben der Statistischen Landesämter sowie des Statistischen Bundesamts und der Landesschulämter (Carstensen, Frey, Walter & Knoll, 2007). Entsprechend sind basierend auf den PISA-Stichproben zuverlässige Aussagen über die Gesamtheit der Zielpopulation der fünfzehnjährigen Schülerinnen und Schüler in Deutschland möglich.

chen Wert zu. Die Tendenz ist dabei sowohl zwischen PISA 2003 und PISA 2006 als auch zwischen PISA 2006 und PISA 2009 ersichtlich, wobei die Differenzen im erstgenannten Zeitraum jeweils im Betrag höher ausfallen.

Sind nun angesichts der beschriebenen Veränderungen in der Jahrgangsstufenverteilung bedeutsame positive Veränderungen in den Kompetenzen der Fünfzehnjährigen zu erwarten? Die jeweils besuchte Jahrgangsstufe ist von entscheidender Bedeutung für die Lernzeit, welche Schülerinnen und Schüler in ihrer Schullaufbahn kumulieren. Die Lernzeit, die ihnen zur Verfügung steht, gehört wiederum zu den zentralen Faktoren des individuellen Lernerfolgs (Helmke, 2010; Seidel & Shavelson, 2007). Entsprechend sollten sich Veränderungen in der Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen auch in den mittleren Kompetenzen widerspiegeln. Weiteren Aufschluss geben hier Studien, die den Lernzuwachs innerhalb einzelner Schuljahre am Ende der Sekundarstufe I quantifizieren. In der Studie PISA-I-Plus wurde die Stichprobe der Fünfzehnjährigen aus PISA 2003 ein Jahr später (in der zehnten Jahrgangsstufe) noch einmal mit dem Schwerpunkt Mathematik getestet, um Aufschlüsse über die Kompetenzentwicklung im Verlauf eines Schuljahres zu ermöglichen (Prenzel et al., 2006). Für den grundbildungsorientierten Mathematiktest wurde zwischen der neunten und zehnten Jahrgangsstufe ein Kompetenzzuwachs von Cohens $d = 0.33$ ermittelt, was als deutliche Verbesserung eingestuft werden kann (Ehmke, Blum, Neubrand, Jordan & Ulfig, 2006). Anders stellen sich die Befunde im Bereich des Lesens dar: Während sich in der Hamburger Längsschnittstudie Aspekte der Lernausgangslage und der Lernentwicklung (LAU) zwischen dem Beginn der siebten und dem Ende der achten Jahrgangsstufe jährliche Kompetenzzuwächse von etwa einer Viertel Standardabweichung ergaben (Lehmann, Peek, Gänsfuß & Husfeldt, 2002), resultierte im Lesekompetenztest in der Studie Deutsch Englisch Schülerleistungen International (DESI) kein signifikanter Zuwachs im Laufe der neunten Jahrgangsstufe (Gailberger & Willenberg, 2008).

4. Zielsetzung

Im vorliegenden Beitrag wird untersucht, wie sich die mittleren Kompetenzen der fünfzehnjährigen Schülerinnen und Schüler sowie der Schülerinnen und Schüler in neunten Jahrgangsstufen in den Domänen Lesen, Mathematik und Naturwissenschaften zwischen PISA 2003 und PISA 2009 in Deutschland verändert haben. Bei allen Analysen beschränken wir uns auf die Betrachtung von Trends in den Kompetenzmittelwerten der Schülerinnen und Schüler. Weiterhin beschränkt sich unsere Darstellung auf die PISA-Erhebungen der Jahre 2003 bis 2009 im Lesen und in der Mathematik sowie der Jahre 2006 und 2009 in den Naturwissenschaften. Dies geschieht aus zwei Gründen: Erstens wird erst seit PISA 2003 bei der nationalen Stichprobenerweiterung um die Schülerinnen und Schüler aus neunten Klassen ein vergleichbares Vorgehen angewandt (vgl. Abschnitt 2). Zweitens erfolgt für den Bereich Naturwissenschaften auch international erst beginnend mit PISA 2006 eine Trendberichterstattung, da erst zu diesem Zeitpunkt, als die Naturwissenschaften den Schwerpunktbereich von PISA bildeten, die ausdifferen-

zierte Konzeption für die Erfassung naturwissenschaftlicher Kompetenz vorlag (OECD, 2006).

Über die reine Deskription von Trends in den Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler hinaus stellt sich die Frage, welche Ursachen möglichen Unterschieden zwischen den Trends in den Stichproben zugrunde liegen. Als Erklärungsfaktoren bieten sich hierfür insbesondere Veränderungen in strukturellen Merkmalen des Bildungssystems an, welches sich in den zurückliegenden Jahren seit PISA 2000 in verschiedener Hinsicht bedeutend gewandelt hat (siehe Klieme et al., 2010). Wir werden uns hier auf einen potenziell zentralen Erklärungsfaktor, die Veränderung in der Verteilung fünfzehnjähriger Schülerinnen und Schüler auf die Jahrgangsstufen, beschränken und gehen dabei von folgenden Annahmen aus: Die mit dem zunehmenden Besuch höherer Jahrgangsstufen (vgl. Abschnitt 3) einhergehende längere Beschulung sollte sich insgesamt positiv auf die Kompetenzmittelwerte der Fünfzehnjährigen auswirken. Die von den Neuntklässlerinnen und Neuntklässlern besuchte Jahrgangsstufe ist dagegen definitionsgemäß konstant, sodass – bei Berücksichtigung dieses einen Erklärungsfaktors – von einer günstigeren Veränderung der Kompetenzmittelwerte bei den Fünfzehnjährigen verglichen mit den Neuntklässlerinnen und Neuntklässlern auszugehen ist. Aus den in Abschnitt 3 angeführten Studien zum Lernzuwachs lässt sich ableiten, dass positive Veränderungen in den Kompetenzmittelwerten der Fünfzehnjährigen im Lesen vor allem dann zu erwarten sind, wenn ein höherer Anteil von Schülerinnen und Schülern die Jahrgangsstufe 7 beziehungsweise 8 bereits absolviert hat. Abnehmende Anteile in Jahrgangsstufe 9 zugunsten zunehmender Anteile in Jahrgangsstufe 10 sollten weniger bedeutsam sein, da sich im Lesen für Jahrgangsstufe 9 keine Lernzuwächse nachweisen ließen. In der Mathematik hingegen resultierten deutliche Lernzuwächse auch für die neunte Jahrgangsstufe; entsprechend sollten eine Abnahme des Anteils der Fünfzehnjährigen in der neunten Jahrgangsstufe und eine Zunahme des Anteils in der zehnten Jahrgangsstufe mit positiven Veränderungen in den Kompetenzmittelwerten einhergehen. Betrachtet man die Veränderungen in der Jahrgangsstufenverteilung zwischen PISA 2003 und PISA 2009 (vgl. Abschnitt 3), so ergibt sich angesichts der Verringerung der Anteile in den Jahrgangsstufen 8 und 9 und der Zunahme des Anteils in Jahrgangsstufe 10 die Erwartung eines vergleichsweise deutlichen Kompetenzzuwachses im Bereich Mathematik und eines geringer ausgeprägten Kompetenzzuwachses im Bereich Lesen. Für die Naturwissenschaften lassen sich aus bisherigen Studien keine eindeutigen Vorhersagen ableiten. Allerdings zeigen sich zwischen den relevanten PISA-Erhebungen 2006 und 2009 nur vergleichsweise geringe Veränderungen hin zum Besuch höherer Jahrgangsstufen. Entsprechend wäre im Bereich Naturwissenschaften eine allenfalls schwache Verbesserung des Kompetenzmittelwerts zu erwarten.

5. Methode

Im Folgenden werden zunächst die analysierten Stichproben aus den PISA-Erhebungen 2003 bis 2009 dargestellt, bevor die statistischen Verfahren kurz erläutert werden, die für die Vergleiche zwischen den Erhebungszeitpunkten eingesetzt wurden.

5.1 Stichprobenbeschreibung

Insgesamt wurden sechs Stichproben analysiert. Für die PISA-Erhebungen 2003, 2006 und 2009 wurden jeweils die internationale Stichprobe der Fünfzehnjährigen sowie die nationale Stichprobenerweiterung der neunten Jahrgangsstufe in Deutschland einbezogen.[3] Nachfolgend werden die analysierten Stichproben einzeln beschrieben.

PISA 2003

Internationale Stichprobe der Fünfzehnjährigen

Zwischen März und Mai 2003 nahmen am internationalen Vergleich 216 Schulen in Deutschland teil. Innerhalb dieser Schulen wurden Zufallsstichproben von fünfzehnjährigen Schülerinnen und Schülern gezogen. An den Sonder- und Förderschulen sowie an den Beruflichen Schulen umfasste die Stichprobe alle Fünfzehnjährigen, in den anderen Schulen wurden jeweils 25 Fünfzehnjährige in die Stichprobe einbezogen. Insgesamt umfasst die für den internationalen Vergleich im Rahmen von PISA 2003 gezogene Stichprobe 4660 fünfzehnjährige Schülerinnen und Schüler.

Nationale Stichprobe der neunten Jahrgangsstufe

Bei PISA 2003 wurden in den am internationalen Vergleich teilnehmenden allgemeinbildenden Schulen zusätzlich zwei vollständige neunte Klassen pro Schule zufällig gezogen. Die Schülerinnen und Schüler der Stichprobe bearbeiteten am ersten Testtag – wie die Fünfzehnjährigen – den internationalen PISA-Test und an einem zweiten Testtag zusätzliche nationale (stärker curricular orientierte) Tests für Mathematik und Naturwissenschaften sowie einen ergänzenden nationalen Fragebogen. Die nationale PISA-2003-Stichprobe umfasst insgesamt 8559 Neuntklässlerinnen und Neuntklässler an 198 Schulen. Die Differenz in der Anzahl der Schulen zur internationalen Stichprobe ergibt sich durch die teilnehmenden neun Sonder- und Förderschulen sowie neun Beruflichen Schulen.

PISA 2006

Internationale Stichprobe der Fünfzehnjährigen

In Deutschland wurden zwischen April und Mai 2006 für den internationalen Vergleich Schülerinnen und Schüler aus 226 Schulen getestet. Den internationalen Vorgaben folgend wurden pro Schule 25 fünfzehnjährige Schülerinnen und Schüler aller Jahrgangs-

3 Anders als die PISA-Stichproben der Fünfzehnjährigen enthalten die PISA-Stichproben der Neuntklässlerinnen und Neuntklässler keine Schülerinnen und Schüler aus Sonder- und Förderschulen sowie Beruflichen Schulen. Da bei der Trendberichterstattung im Rahmen von PISA jeweils die gesamten Stichproben der Fünfzehnjährigen zugrunde gelegt werden, wird auch im Folgenden die gesamte Stichprobe der Fünfzehnjährigen inklusive dieser Schülerinnen und Schüler herangezogen (Anteile der Schülerinnen und Schüler in Sonder- und Förderschulen sowie Beruflichen Schulen: PISA 2003: 6.3%, PISA 2006: 7.1%, PISA 2009: 7.7%).

stufen zufällig gezogen. Die Stichprobe umfasst 4891 fünfzehnjährige Schülerinnen und Schüler.

Nationale Stichprobe der neunten Jahrgangsstufe
Die internationale Stichprobe wurde bei PISA 2006 in Deutschland durch die zufällige Ziehung von zwei kompletten Klassen der neunten Jahrgangsstufe je für den internationalen Vergleich gezogener Schule ergänzt. Die Schülerinnen und Schüler bearbeiteten am ersten Testtag den internationalen PISA-Test und an einem zweiten Testtag Mathematikaufgaben zur Überprüfung des Erreichens der Bildungsstandards in Mathematik für den Mittleren Schulabschluss (KMK, 2004a). Die Stichprobe umfasst 9577 Schülerinnen und Schüler an 203 Schulen. Die Differenz in der Anzahl der Schulen zur internationalen Stichprobe ergibt sich durch die teilnehmenden Sonder- und Förderschulen sowie Beruflichen Schulen.

PISA 2009
Internationale Stichprobe der Fünfzehnjährigen
PISA 2009 wurde zwischen April und Mai 2009 an Schulen aller Schularten in Deutschland durchgeführt. Die Stichprobe umfasste gemäß den internationalen Vorgaben pro Schule 25 fünfzehnjährige Schülerinnen und Schüler aller Jahrgangsstufen an insgesamt 226 Schulen in Deutschland. In die Stichprobe gehen Daten von 4979 Schülerinnen und Schülern ein.

Nationale Stichprobe der neunten Jahrgangsstufe
Darüber hinaus wurden an den an PISA 2009 teilnehmenden allgemeinbildenden Schulen pro Schule zusätzlich zwei komplette neunte Klassen getestet. Diese nahmen am ersten Testtag am internationalen PISA-Test teil und an einem zweiten Testtag an den Tests des ersten Ländervergleichs zur Überprüfung der Bildungsstandards für den Mittleren Schulabschluss in den Fächern Deutsch, Englisch und Französisch (KMK, 2004b, c; Köller, Knigge & Tesch, 2010). Die Stichprobe umfasst 9461 Schülerinnen und Schüler an 202 Schulen. Die Differenz in der Anzahl der Schulen zur internationalen Stichprobe ergibt sich durch die teilnehmenden 13 Sonder- und Förderschulen sowie elf Beruflichen Schulen.

5.2 Datengrundlage

Die Datengrundlage der analysierten Stichproben (vgl. Abschnitt 5.1) bilden im Falle der Fünfzehnjährigen Kompetenzwerte für Deutschland aus der internationalen PISA-Skalierung, die auch Grundlage der Berichtlegung sind (z.B. Klieme et al., 2010; OECD, 2010). Bei den Stichproben der neunten Jahrgangsstufe beruhen die Kompetenzwerte auf nationalen Skalierungen, die unter Federführung der jeweils mit der Projektkoordination in Deutschland beauftragten Institute durchgeführt wurden. Für PISA 2003 und

PISA 2006 war dies das Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) in Kiel und für PISA 2009 das Deutsche Institut für Internationale Pädagogische Forschung (DIPF) in Frankfurt am Main. Das Vorgehen bei der nationalen Skalierung orientiert sich am internationalen Vorgehen (OECD, 2011) und unterscheidet sich davon im Wesentlichen durch die Berücksichtigung weiterer Variablen aus dem Fragebogen für Schülerinnen und Schüler, die als zusätzliche Hintergrundinformationen in die Skalierung mit einfließen (siehe z.B. Carstensen et al., 2007). Die nationale Skalierung bei PISA 2009 wurde zur Gewährleistung von Vergleichbarkeit mit den früheren PISA-Erhebungen in enger Abstimmung zwischen DIPF und IPN durchgeführt. Eine Vergleichbarkeit der verwendeten Kompetenzdaten zwischen den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe ist ebenfalls gegeben, da bei der nationalen Skalierung (a) das internationale Hintergrundmodell übernommen und im Weiteren ausschließlich ergänzt wurde und (b) die internationalen Itemschwierigkeitsparameter verwendet wurden. Zudem wurde dasselbe mehrdimensionale Skalierungsmodell wie in der internationalen Skalierung herangezogen.

In den im Folgenden durchgeführten Analysen wird mit den *Plausible Values* (PVs; von Davier, Gonzalez & Mislevy, 2009) als Schätzern der Personenfähigkeit gearbeitet, die in den internationalen und nationalen Skalierungen bei PISA 2003, PISA 2006 und PISA 2009 gewonnen wurden. Die PVs erlauben eine unverzerrte Schätzung der Kompetenzmittelwerte und Kompetenzstreuungen sowie der Beziehungen der Kompetenzen zu weiteren Variablen, die – wie die im vorliegenden Beitrag analysierten Variablen – im Hintergrundmodell berücksichtigt wurden. Infolge der Verwendung von PVs lagen in den Stichproben jeweils für alle Schülerinnen und Schüler Kompetenzwerte für alle Domänen (Lesen, Mathematik, Naturwissenschaften) vor, das heißt auch für jene Schülerinnen und Schüler, die aufgrund des Studiendesigns nicht alle Domänen bearbeitet haben.

Gebhardt und Adams (2007) diskutieren drei verschiedene Methoden der Trendskalierung mit PISA-Daten. Als „originale Trends“ werden Veränderungen in den Kompetenzindikatoren bezeichnet, die in den PISA-Erhebungszyklen zur querschnittlichen Berichterstattung konstruiert wurden. Diese originalen Trends basieren auf gemeinsamen Skalierungen der Aufgabenantworten aus den teilnehmenden OECD-Staaten und werden von der OECD im Rahmen ihrer Trendberichterstattung (z.B. OECD, 2010) und auch in der vorliegenden Arbeit verwendet. Unter „marginalen Trends“ werden Veränderungen verstanden, die auf eigenen Skalierungen der Aufgabenantworten für jeden Staat beruhen. Die Aufgabenschwierigkeiten werden für jeden Staat separat geschätzt und es resultiert eine staatenspezifische Kompetenzskala. Sind einzelne Aufgaben oder Aufgabengruppen in einem Staat schwerer oder leichter als in der Gesamtheit der teilnehmenden Staaten, werden diese Unterschiede in den staatenspezifischen Aufgabenschwierigkeitsparametern berücksichtigt. „Bedingte Trends“ sind marginale Trends, bei denen die Veränderungen zwischen den Erhebungszyklen zusätzlich um mögliche Unterschiede in den Stichprobenzusammensetzungen korrigiert werden. In Deutschland wurden im Rahmen von PISA 2006 für längsschnittliche Fragestellungen Ergebnisse einer eigens dafür vorgenommenen marginalen Skalierung berichtet (Carstensen, Prenzel & Baumert, 2008). Es konnte gezeigt werden, dass die originalen und margina-

len Trends für den Bereich Lesen in Deutschland voneinander abweichen und die originalen Trends die Kompetenzen der Schülerinnen und Schüler für PISA 2003 und PISA 2006 überschätzen. Dieser Befund wird in Hinblick auf die Ergebnisse dieses Beitrags in Abschnitt 7 noch einmal aufgegriffen.

Die Zielpopulation der internationalen PISA-Stichprobe in Deutschland bilden alle Schülerinnen und Schüler, die zum Zeitpunkt der Erhebung fünfzehn Jahre alt sind. Die Auswahl der Schulen und der teilnehmenden Schülerinnen und Schüler erfolgt im Rahmen einer zweifach stratifizierten Stichprobenziehung (z.B. Sibberns & Baumert, 2001). Die Wahrscheinlichkeit, in die Stichprobe zu gelangen, ist als Folge dieses Vorgehens nicht für alle Schülerinnen und Schüler gleich. Um repräsentative Aussagen über die Grundgesamtheit der Fünfzehnjährigen zu ermöglichen, werden für alle Schülerinnen und Schüler Gewichte ermittelt (vgl. OECD, 2011). Diese werden zentral durch das internationale PISA-Konsortium berechnet und durch das deutsche Konsortium verifiziert. Nach einem vergleichbaren Vorgehen werden auch für die nationalen Stichproben der neunten Jahrgangsstufe Gewichte entwickelt, die es ermöglichen, auf die Zielpopulation der Neuntklässlerinnen und Neuntklässler in Deutschland zu schließen. Für die im Weiteren dargestellten Analysen werden die finalen Gewichte der Schülerinnen und Schüler (*final student weights*) verwendet (eine Ausnahme wird in Abschnitt 6.2 erläutert).

5.3 Analysen

Bei den nachstehend berichteten Analysen beschränken wir uns auf die Betrachtung von Trends in den Kompetenzmittelwerten der Schülerinnen und Schüler. Mittelwerte und zugehörige Standardfehler werden gemäß dem bei PISA international üblichen Vorgehen errechnet (vgl. OECD, 2011). Unterschiede in den Kompetenzmittelwerten zwischen den verschiedenen Erhebungszeitpunkten werden anhand der *z*-Verteilung auf Signifikanz geprüft. Dabei muss zusätzlich zum Standardfehler der Kompetenzmittelwerte, der sich aus dem Standardmessfehler und dem Sampling-Fehler zusammensetzt, der *Linking-Fehler* berücksichtigt werden. Um die Kompetenzverteilungen verschiedener PISA-Erhebungen aufeinander beziehen zu können, werden die Itemschwierigkeiten für die einzelnen Kompetenzbereiche jeweils an der Berichtsskala jener Erhebung verankert, bei der diese Kompetenz erstmals den inhaltlichen Schwerpunkt darstellte (*Linking*). Das Ausmaß der Variation in den Schwierigkeiten der für das Linking verwendeten Items zwischen den Erhebungen wird durch den Linking-Fehler quantifiziert. Der Linking-Fehler reflektiert somit die statistische Unsicherheit, die auf die Verbindung von Tests verschiedener Erhebungszeitpunkte zurückgeht.

Der Standardfehler der Mittelwertedifferenz zweier Erhebungszeitpunkte berechnet sich nach folgender Formel (exemplarisch für den Vergleich der beiden Erhebungszeitpunkte PISA 2003 und PISA 2009; vgl. OECD, 2011, S. 146):

$$SE = \sqrt{\sigma^2_{\hat{\mu}_{2003}} + \sigma^2_{\hat{\mu}_{2009}} + \sigma^2_{\hat{\mu}_{Linking\text{-}Fehler}}}$$

Die Durchführung der berichteten Signifikanztests zur Beurteilung von Mittelwertunterschieden zwischen je zwei Erhebungszeitpunkten erfolgte unter Verwendung der international gültigen Linking-Fehler. Da die bei PISA administrierten Tests für die Stichprobe der Fünfzehnjährigen und die Stichprobe der neunten Jahrgangsstufe jeweils identisch sind und die Skalierung der Stichprobe der neunten Jahrgangsstufe auf Grundlage der international für die Fünfzehnjährigen geschätzten Itemschwierigkeiten (vgl. Abschnitt 5.2) durchgeführt wird, sind für beide Stichproben dieselben Linking-Fehler zu verwenden. Tabelle 2 enthält die Linking-Fehler für die PISA-Erhebungen 2003 bis 2009.

	Linking-Fehler auf der PISA-Skala
Gesamtskala Lesen 2003/2006	5.307
Gesamtskala Lesen 2003/2009	4.088
Gesamtskala Lesen 2006/2009	4.069
Gesamtskala Mathematik 2003/2006	1.382
Gesamtskala Mathematik 2003/2009	1.990
Gesamtskala Mathematik 2006/2009	1.333
Gesamtskala Naturwissenschaften 2006/2009	2.566

Tab. 2: Linking-Fehler für die PISA-Erhebungen 2003 bis 2009

6. Ergebnisse

Zunächst werden die Veränderungen in den mittleren Kompetenzen in den Domänen Lesen und Mathematik zwischen PISA 2003 und PISA 2009 sowie in den Naturwissenschaften zwischen PISA 2006 und PISA 2009 in den Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler dargestellt. Anschließend werden Veränderungen in der Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen berichtet und mit den Trends in Beziehung gesetzt.

6.1 Veränderungen der mittleren Kompetenzen in den Stichproben

Lesekompetenz

Die Verläufe der mittleren Lesekompetenz über die Erhebungszeitpunkte (siehe Abbildung 1) unterscheiden sich zwischen den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe nicht auffällig voneinander. Die mittlere Lesekompetenz der Fünfzehnjährigen in Deutschland lag bei PISA 2003 bei 491 Punkten (SE = 3.4 Punkte), bei PISA 2006 bei 495 Punkten (SE = 4.4 Punkte) und bei PISA 2009 bei 497 Punkten

(*SE* = 2.7 Punkte). Die mittlere Lesekompetenz der Fünfzehnjährigen stieg demzufolge über die Erhebungszeitpunkte zumindest deskriptiv kontinuierlich leicht an.

In der Stichprobe der neunten Jahrgangsstufe lag die mittlere Lesekompetenz bei PISA 2003 bei 492 Punkten (*SE* = 1.9 Punkte), bei PISA 2006 wurden 498 Punkte (*SE* = 3.5 Punkte) und bei PISA 2009 494 Punkte (*SE* = 2.2 Punkte) erreicht. Die mittlere Lesekompetenz in der neunten Jahrgangsstufe erreichte also deskriptiv bei PISA 2006 ihre höchste Ausprägung innerhalb des analysierten Zeitraums.

Die größten Differenzen in der Lesekompetenz finden sich mit jeweils sechs Punkten[4] bei den Fünfzehnjährigen zwischen PISA 2009 und PISA 2003 und für die neunte Jahrgangsstufe zwischen PISA 2006 und PISA 2003. Diese Differenzen sind jedoch – ebenso wie alle anderen Differenzen zwischen den Erhebungszeitpunkten in den beiden Stichproben – statistisch nicht bedeutsam ($p \geq .16$). Damit lässt sich die deskriptiv leicht positive Veränderung bei den Fünfzehnjährigen und den Neuntklässlerinnen und Neuntklässlern über den beobachteten Zeitraum von sechs Jahren nicht gegen den Zufall absichern.

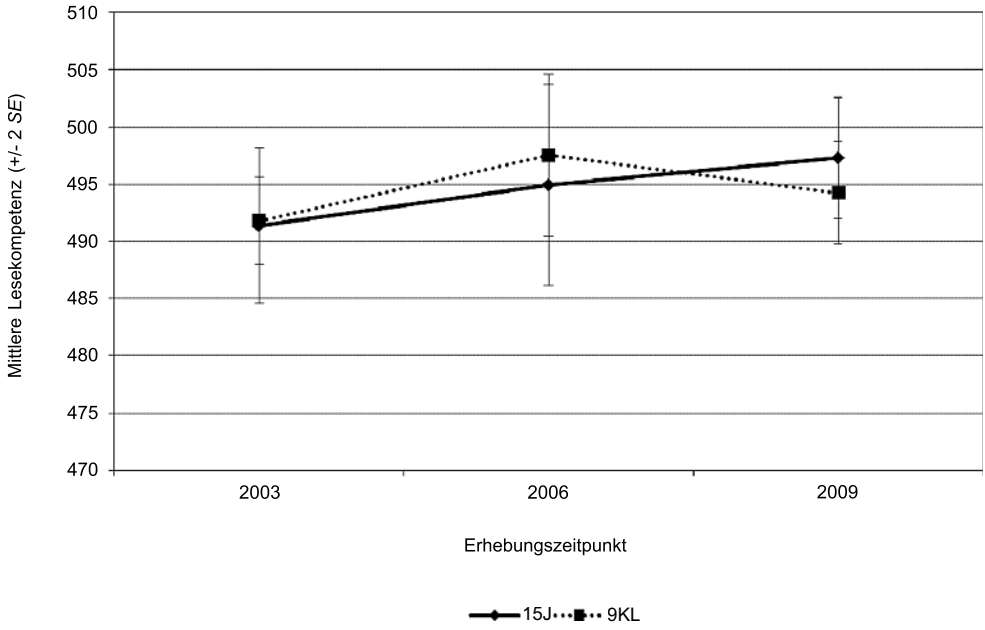

Abb. 1: Mittlere Lesekompetenz in Deutschland von PISA 2003 bis PISA 2009

4 In der Ergebnisdarstellung werden ganzzahlig gerundete Werte angegeben. Es können im Folgenden auf Rundung zurückzuführende vermeintliche Inkonsistenzen bei der Darstellung von Differenzen vorkommen. Alle Berechnungen wurden mit Dezimalstellen durchgeführt, erst für die Ergebnisdarstellung wurde gerundet.

Mathematische Kompetenz

Die Veränderungen der mathematischen Kompetenz in den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe über die Erhebungszeitpunkte sind in Abbildung 2 dargestellt. Die mittlere mathematische Kompetenz der Fünfzehnjährigen in Deutschland lag bei PISA 2003 bei 503 Punkten (*SE* = 3.3 Punkte), bei PISA 2006 bei 504 Punkten (*SE* = 3.9 Punkte) und bei PISA 2009 bei 513 Punkten (*SE* = 2.9 Punkte). Die mittlere mathematische Kompetenz der Fünfzehnjährigen veränderte sich zwischen PISA 2003 und PISA 2006 also kaum, stieg jedoch in den drei darauf folgenden Jahren deutlicher an (siehe auch Frey, Heinze, Mildner, Hochweber & Asseburg, 2010).

In der Stichprobe der neunten Jahrgangsstufe lag die mittlere mathematische Kompetenz bei PISA 2003 bei 504 Punkten (*SE* = 1.9 Punkte). Bei PISA 2006 wurden 501 Punkte (*SE* = 3.2 Punkte) erzielt, bei PISA 2009 507 Punkte (*SE* = 2.4 Punkte). Wie bei den Fünfzehnjährigen erreichte die mittlere mathematische Kompetenz der Neuntklässlerinnen und Neuntklässler deskriptiv ihre höchste Ausprägung bei PISA 2009, bewegte sich aber über alle Erhebungszeitpunkte auf einem sehr ähnlichen Niveau.

Die größten Differenzen finden sich bei den Fünfzehnjährigen mit zehn Punkten zwischen PISA 2009 und PISA 2003 und für die neunte Jahrgangsstufe mit sechs Punkten zwischen PISA 2009 und PISA 2006. Der Anstieg der mathematischen Kompetenz der Fünfzehnjährigen um zehn Punkte zwischen PISA 2003 und PISA 2009 ist statistisch bedeutsam ($p = .04$). Die Effektstärke dieser Differenz beträgt Cohens $d = 0.10$, was nach gängigen Kriterien der psychologischen Forschung als eher geringer Unterschied anzusehen ist (vgl. Bortz & Döring, 2006, S. 606), aber gemessen an den Befun-

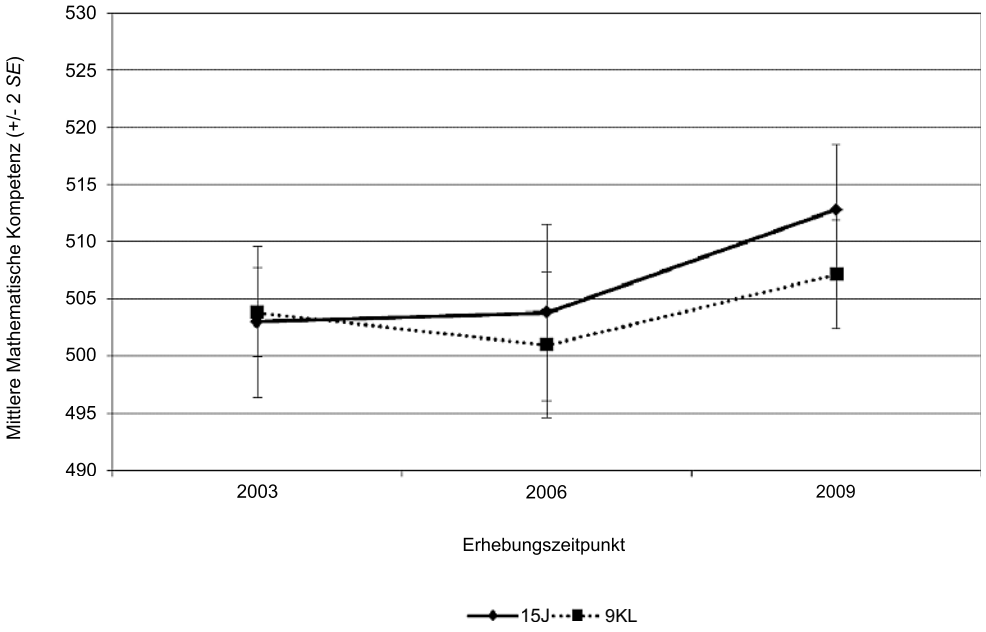

Abb. 2: Mittlere Mathematische Kompetenz in Deutschland von PISA 2003 bis PISA 2009

den von PISA-I-Plus immerhin etwa einem Drittel des Lernzuwachses in der neunten Jahrgangsstufe im Bereich Mathematik entspricht (vgl. Abschnitt 3). Die einzelnen Veränderungen von PISA 2003 zu PISA 2006 (+1 Punkt) beziehungsweise von PISA 2006 zu PISA 2009 (+9 Punkte) sind hingegen nicht signifikant ($p \geq .07$). Für die Stichprobe der neunten Jahrgangsstufe finden sich keinerlei überzufällige Veränderungen zwischen den Erhebungszeitpunkten ($p \geq .14$).

Naturwissenschaftliche Kompetenz

Abbildung 3 zeigt die Veränderungen der naturwissenschaftlichen Kompetenz in den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe zwischen PISA 2006 und PISA 2009. Die mittlere naturwissenschaftliche Kompetenz der Fünfzehnjährigen in Deutschland lag bei PISA 2006 bei 516 Punkten (*SE* = 3.8 Punkte) und bei PISA 2009 bei 520 Punkten (*SE* = 2.8 Punkte). Die mittlere naturwissenschaftliche Kompetenz der Fünfzehnjährigen stieg demnach deskriptiv leicht an (siehe auch Rönnebeck, Schöps, Prenzel, Mildner & Hochweber, 2010).

In der Stichprobe der neunten Jahrgangsstufe lag die mittlere naturwissenschaftliche Kompetenz bei PISA 2006 bei 514 Punkten (*SE* = 3.4 Punkte), bei PISA 2009 wurden 518 Punkte (*SE* = 2.5 Punkte) erreicht. Die mittlere naturwissenschaftliche Kompetenz nahm damit im beobachteten Dreijahreszeitraum auch in der neunten Jahrgangsstufe deskriptiv leicht zu.

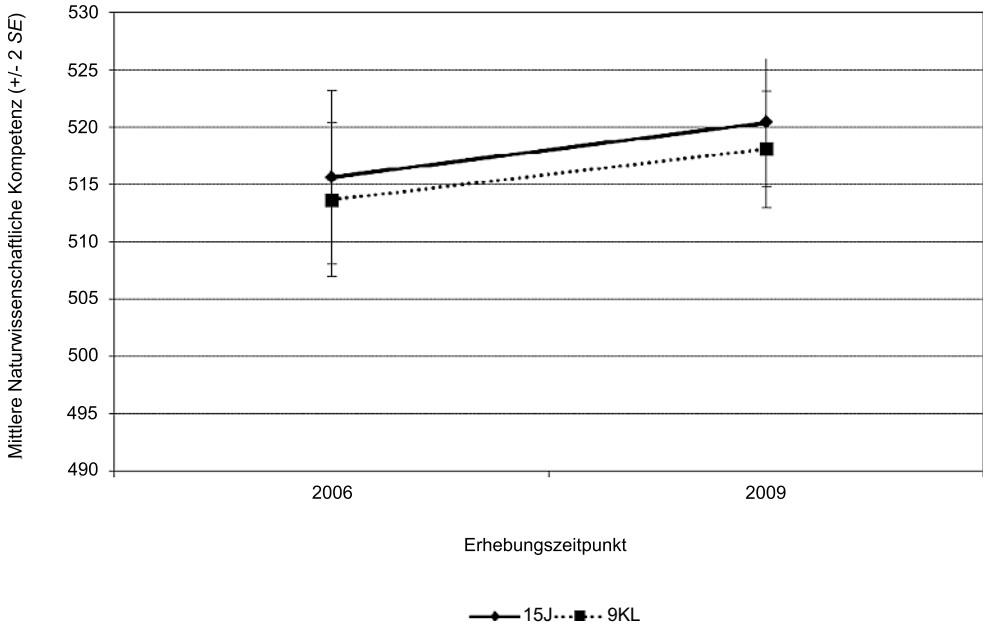

Abb. 3: Mittlere Naturwissenschaftliche Kompetenz in Deutschland von PISA 2006 bis PISA 2009

Die Verläufe sind für die beiden Stichproben nahezu parallel und auch das erreichte Punkteniveau ist jeweils sehr ähnlich (vgl. Abbildung 3). Zwischen PISA 2006 und PISA 2009 finden sich deskriptiv sowohl für die Fünfzehnjährigen als auch für die neunte Jahrgangsstufe Zuwächse in Höhe von vier Punkten in der naturwissenschaftlichen Kompetenz. Diese Zuwächse sind vergleichsweise gering und statistisch nicht signifikant ($p \geq .37$).

Zusammenfassung

Deskriptiv betrachtet erweisen sich die Verläufe der mittleren Kompetenzen über die beobachteten Erhebungszeitpunkte in den Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler als relativ ähnlich, wobei die Veränderung der mittleren Kompetenzen bei den Fünfzehnjährigen insgesamt etwas günstiger ausfällt. Allerdings lässt sich in beiden Stichproben keine der Veränderungen zwischen PISA 2006 und PISA 2009 gegen den Zufall absichern; ebenso statistisch nicht signifikant sind die Veränderungen zwischen PISA 2003 und PISA 2006, die auch deskriptiv in beiden Stichproben gering ausfallen. Die einzige signifikante Veränderung zeigt sich zwischen PISA 2003 und PISA 2009 in der mittleren mathematischen Kompetenz der Fünfzehnjährigen – ein positiver Trend, der sich für die Neuntklässlerinnen und Neuntklässler nicht nachweisen lässt.

6.2 Analysen zur Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen

Können die beobachteten Trends in den Kompetenzmittelwerten der Fünfzehnjährigen mit den Veränderungen in der Verteilung auf die Jahrgangsstufen in Beziehung gebracht werden? Es zeigt sich, dass die in Abschnitt 4 formulierten Annahmen im Wesentlichen zutreffen: Zwischen den PISA-Erhebungen 2003 und 2009 beziehungsweise 2006 und 2009 zeigen sich im Lesen und in den Naturwissenschaften deskriptiv nur geringe positive Veränderungen, die sich nicht statistisch absichern lassen. In der Mathematik resultiert dagegen zwischen PISA 2003 und PISA 2009 ein deutlicher sowie signifikanter mittlerer Kompetenzzuwachs. Der positive Trend in der Mathematik korrespondiert mit einer Verringerung der Anteile von Achtklässlerinnen und Achtklässlern sowie Neuntklässlerinnen und Neuntklässlern und einer Zunahme des Anteils der Zehntklässlerinnen und Zehntklässler zwischen PISA 2003 und PISA 2009. Berücksichtigt man Studien, die bereichs- sowie jahrgangsstufenspezifische Lernzuwächse von Schülerinnen und Schülern ermitteln konnten (vgl. Abschnitt 3), erscheint es plausibel, dass sich dies gerade im Bereich der Mathematik leistungsförderlich ausgewirkt hat.

Um einen noch genaueren Eindruck zu erhalten, wurde eine vertiefende deskriptive Analyse basierend auf neu erzeugten Gewichten der Schülerinnen und Schüler durchgeführt (vgl. Abschnitt 5.2). Dabei wurde als wichtige Kontrollvariable die Schulartzugehörigkeit berücksichtigt. Differenziert wurde hier zwischen dem Gymnasium einerseits und den nicht-gymnasialen Schularten andererseits, da das Gymnasium als einzige

Schulart im interessierenden Zeitraum und in allen Bundesländern von bildungsstrukturellen Reformmaßnahmen relativ unberührt geblieben ist. Die neuen Gewichte der Schülerinnen und Schüler wurden gebildet, indem die originalen Gewichte der PISA-Erhebungen 2006 und 2009 (a) je Schulart (Gymnasium versus nicht-gymnasiale Schularten) und (b) je Schulart und Jahrgangsstufe mit dem Quotienten der Gewichtsumme in PISA 2003 und der Gewichtsumme in PISA 2006 beziehungsweise PISA 2009 multipliziert wurden. Basierend auf den neu generierten Gewichten wurden dann die Kompetenzmittelwerte in den drei Domänen für die PISA-Erhebungen 2006 und 2009 neu berechnet. Die Resultate lassen sich als (fiktive) Ergebnisse bei Konstanthalten (a) der Schulart beziehungsweise (b) der Schulart und der Jahrgangsstufenverteilung für die Fünfzehnjährigen seit PISA 2003 interpretieren. Die Ergebnisse mit den originalen und den neu erzeugten Gewichten der Schülerinnen und Schüler sind in Tabelle 3 gegenübergestellt.

Domäne/Analyse	PISA 2003	PISA 2006	PISA 2009
Lesen			
Mit Originalgewichten	491	495	497
Neu gewichtet: Schulart	–	493	492
Neu gewichtet: Schulart + Jahrgangsstufe	–	489	489
Mathematik			
Mit Originalgewichten	503	504	513
Neu gewichtet: Schulart	–	502	507
Neu gewichtet: Schulart + Jahrgangsstufe	–	498	502
Naturwissenschaften			
Mit Originalgewichten	–	516	520
Neu gewichtet: Schulart	–	513	515
Neu gewichtet: Schulart + Jahrgangsstufe	–	511	511

Tab. 3: Mittlere Kompetenzen der Fünfzehnjährigen in den PISA-Erhebungen 2003 bis 2009 bei Verwendung der originalen Stichprobengewichte und bei Neugewichtung der PISA-Erhebungen 2006 und 2009 basierend auf (a) der Schulart sowie (b) der Schulart und der Jahrgangsstufenverteilung der Fünfzehnjährigen in PISA 2003

Der Vergleich der Befunde mit den originalen und den neuen Gewichten weist darauf hin, dass der signifikante positive Trend bei den Fünfzehnjährigen in der Mathematik zwischen PISA 2003 und PISA 2009 bereits in erheblichem Umfang durch Veränderungen in der Schulartverteilung erklärt werden kann – die Differenz zwischen PISA 2003 und PISA 2009 beträgt bei Berücksichtigung der Schulart nur noch vier Punkte. Da-

rüber hinaus erweisen sich aber auch die Veränderungen in der Jahrgangsstufenverteilung als bedeutsam, es ergibt sich bei Einbezug beider Merkmale eine annähernde Nivellierung der beobachteten Punktedifferenz. Auch für die mittleren Lesekompetenzen sowie die mittleren naturwissenschaftlichen Kompetenzen, bei denen allerdings keine statistische Absicherung des Trends möglich war (vgl. Abschnitt 6.1), resultiert bei Berücksichtigung der Schulart eine deutliche Verringerung der Punktedifferenz zwischen den PISA-Erhebungen. In den Naturwissenschaften, nicht aber im Bereich des Lesens, führt die zusätzliche Berücksichtigung der Jahrgangsstufenverteilung bei der Bildung neuer Stichprobengewichte zu einer weitgehenden Angleichung der ermittelten Kompetenzstände für die Erhebungszeitpunkte.

7. Fazit und Ausblick

In dem vorliegenden Beitrag wurden Trends der mittleren Kompetenzen in den Domänen Lesen, Mathematik und Naturwissenschaften in den PISA-Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler vergleichend analysiert. Herangezogen wurden dazu die Ergebnisse der PISA-Erhebungen der Jahre 2003, 2006 und 2009 (Lesen und Mathematik) beziehungsweise der Jahre 2006 und 2009 (Naturwissenschaften). Als ein schulstrukturelles Merkmal, das zur Erklärung der Trends bei den Fünfzehnjährigen und von möglichen differentiellen Trends bei Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern geeignet erscheint, wurde exemplarisch die Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen untersucht.

Für die Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler lässt sich im Lesen und in den Naturwissenschaften keine der Differenzen zwischen den Erhebungszeitpunkten zufallskritisch absichern. Dagegen zeigt sich für die Fünfzehnjährigen, anders als für die Neuntklässlerinnen und Neuntklässler, in der Mathematik ein signifikanter positiver Trend zwischen PISA 2003 und PISA 2009. Diese Disparität weist darauf hin, dass die mit den deutschen PISA-Daten möglichen alters- und jahrgangsstufenbasierten Trendanalysen durchaus zu unterschiedlichen Befunden führen können. Entsprechend könnte es interessant sein, die Trendberichterstattung bei PISA in Deutschland um Ergebnisse für die Neuntklässlerinnen und Neuntklässler zu ergänzen. Die dadurch verfügbaren zusätzlichen Informationen besitzen praktische Relevanz. So liegt beispielsweise in der Unterrichtsforschung – für welche die Organisation von Schule in Jahrgangsstufen eine zentrale Rolle spielt – die Bezugnahme auf die Kompetenzen der Neuntklässlerinnen und Neuntklässler näher als auf jene der Fünfzehnjährigen.

In Übereinstimmung mit den von uns formulierten Erwartungen zeigte sich, dass die domänenspezifischen Trends bei den Fünfzehnjährigen zumindest deskriptiv mit Veränderungen in der Verteilung auf die Jahrgangsstufen in Verbindung zu bringen sind. Die Tendenz hin zum Besuch höherer Jahrgangsstufen korrespondierte erwartungsgemäß insbesondere mit einem bedeutsamen positiven Trend im Bereich Mathematik, während im Lesen kein vergleichbarer Kompetenzzuwachs zu verzeichnen war. Für die Natur-

wissenschaften konnten keine eindeutigen Erwartungen abgeleitet werden. Die Analyse beschränkte sich hier – in Übereinstimmung mit dem Vorgehen der OECD bei der internationalen Trendberichterstattung (OECD, 2010) – auf die PISA-Erhebungen 2006 und 2009. Für diesen Zeitraum resultierte ein deskriptiv schwacher und nicht signifikanter positiver Trend, der mit eher geringfügigen Veränderungen in der Jahrgangsstufenverteilung korrespondierte. Die potenzielle Relevanz der Jahrgangsstufenverteilung für die Erklärung der Trends bei den Fünfzehnjährigen konnte in den Analysen unter Verwendung neuer Stichprobengewichte auch unter Hinzunahme der Schulart (Gymnasium versus nicht-gymnasiale Schularten) gestützt werden. Für alle Domänen resultierte, dass die beobachteten Punktedifferenzen weitgehend verschwinden, wenn Veränderungen in der Schulart- sowie Jahrgangsstufenverteilung der Fünfzehnjährigen berücksichtigt werden. Dies könnte darauf hindeuten, dass sich die bildungsstrukturellen Maßnahmen, die mehr Fünfzehnjährigen den Besuch höherer Jahrgangsstufen ermöglichten, tatsächlich in positiven Veränderungen in den Kompetenzmittelwerten (zumindest in der Mathematik) niedergeschlagen haben. Die Kompetenzen der Neuntklässlerinnen und Neuntklässler erwiesen sich demgegenüber als nahezu unverändert, die Differenzen zwischen den Erhebungszeitpunkten waren zudem statistisch nicht abzusichern. Allerdings ist bei solchen Schlussfolgerungen große Vorsicht geboten, da die beobachteten Beziehungen weder statistisch abgesichert sind, noch weitere möglicherweise einflussreiche Faktoren berücksichtigt wurden.

Zusätzliche potenzielle Einschränkungen bei der Interpretation der Ergebnisse dieser Studie ergeben sich aus den bereits angesprochenen Befunden von Carstensen et al. (2008), die zeigen konnten, dass alternative Skalierungsverfahren zu voneinander abweichenden Trendergebnissen für die Lesekompetenz der Fünfzehnjährigen in Deutschland führen (vgl. Abschnitt 5.2). Demnach wurden in den von der OECD berichteten Trends für die PISA-Erhebungen 2000 bis 2006 die Kompetenzen der Schülerinnen und Schüler für PISA 2003 und PISA 2006 überschätzt. Welche Implikationen diese Befunde für den vorliegenden Beitrag besitzen, lässt sich ohne die Durchführung vergleichbarer Analysen unter Einschluss von PISA 2009 nicht abschließend klären. Allerdings könnten auch im vorliegenden Fall abweichende Befunde resultieren, die sich – angesichts der Verwendung der internationalen Aufgabenschwierigkeitsparameter in den entsprechenden Skalierungen – auch auf die Stichprobe der Neuntklässlerinnen und Neuntklässler erstrecken könnten.

Die von uns durchgeführte Studie könnte in mindestens zwei Richtungen erweitert werden. Erstens haben wir uns in unseren Analysen auf die Betrachtung von Trends in den Kompetenzmittelwerten der Schülerinnen und Schüler beschränkt. Neben den Kompetenzmittelwerten gibt es noch andere statistische Kennwerte, die für die Beurteilung von Kompetenzverläufen bedeutsam sind und in zukünftigen Untersuchungen einbezogen werden könnten, insbesondere die Streuung der Kompetenzen. Zweitens haben wir als einen möglichen Faktor zur Erklärung der Veränderungen in den Kompetenzen bei den Fünfzehnjährigen und von differentiellen Trends bei Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern die Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen betrachtet. Unsere Analysen konnten dabei erste Anhaltspunkte

für mögliche Erklärungszusammenhänge liefern, blieben aber letztlich rein deskriptiv und erlaubten es nicht, die Beziehung zwischen Trendverläufen und potenziellen Bedingungsmerkmalen inferenzstatistisch abzusichern. Eine noch aussagekräftigere Auswertung wäre beispielsweise auf Grundlage einer gemeinsamen (marginalen) Skalierung (siehe Carstensen et al., 2008) der verschiedenen PISA-Erhebungen möglich. Dabei könnten die latenten Differenzen zwischen den Erhebungszeitpunkten direkt mit möglichen Erklärungsvariablen in Beziehung gesetzt werden.

Literatur

Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15-68). Opladen: Leske + Budrich.

Baumert, J., & Stanat, P. (2006). Internationale Schulleistungsvergleiche. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 291-302). Weinheim/Basel: Beltz.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Heidelberg: Springer.

Carstensen, C. H., Frey, A., Walter, O., & Knoll, S. (2007). Technische Grundlagen des dritten internationalen Vergleichs. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 367-390). Münster: Waxmann.

Carstensen, C. H., Prenzel, M., & Baumert, J. (2008). Trendanalysen in PISA: Wie haben sich die Kompetenzen in Deutschland zwischen PISA 2000 und PISA 2006 entwickelt? *Zeitschrift für Erziehungswissenschaft, Sonderheft 10*, 11-34.

Ehmke, T., Blum, W., Neubrand, M., Jordan, A., & Ulfig, F. (2006). Wie verändert sich die mathematische Kompetenz von der neunten zur zehnten Klassenstufe? In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 63-86). Münster: Waxmann.

Frey, A., Heinze, A., Mildner, D., Hochweber, J., & Asseburg, R. (2010). Mathematische Kompetenz von PISA 2003 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 153-176). Münster: Waxmann.

Gailberger, S., & Willenberg, H. (2008). Leseverstehen Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 60-71). Weinheim/Basel: Beltz.

Gebhardt, E., & Adams, R. J. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement, 8*(3), 305-322.

Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyer.

Herzberg, P. Y., & Frey, A. (2011). Kriteriumsorientierte Diagnostik. In L. Hornke, M. Amelang & M. Kersting (Hrsg.), *Enzyklopädie der Psychologie*, Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik, Band 2 Methoden der psychologischen Diagnostik (S. 281-324). Göttingen: Hogrefe.

Klieme, E., Jude, N., Baumert, J., & Prenzel, M. (2010). PISA 2000-2009: Bilanz der Veränderungen im Schulsystem. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-296). Münster: Waxmann.

KMK (Hrsg.) (2004a). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Luchterhand.

KMK (Hrsg.) (2004b). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Luchterhand.

KMK (Hrsg.) (2004c). *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Luchterhand.

Köller, O., Knigge, M., & Tesch, B. (Hrsg.) (2010). *Sprachliche Kompetenzen im Ländervergleich: Befunde des ersten Ländervergleichs zur Überprüfung der Bildungsstandards für den Mittleren Schulabschluss in den Fächern Deutsch, Englisch und Französisch*. Münster: Waxmann.

Lehmann, R., Peek, R., Gänsfuß, R., & Husfeldt, V. (2002). *LAU 9. Aspekte der Lernausgangslage und der Lernentwicklung − Klassenstufe 9. Ergebnisse einer Längsschnittuntersuchung in Hamburg*. Hamburg: Behörde für Bildung und Sport.

Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-72). Münster: Waxmann.

OECD (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.

OECD (2010). *PISA 2009 results: Learning trends: Changes in student performance since 2000 (Volume V)*. Paris: OECD.

OECD (2011). *PISA 2009 technical report*. Paris: OECD.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (Hrsg.) (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.

Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D., & Hochweber, J. (2010). Naturwissenschaftliche Kompetenz von PISA 2006 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 177-198). Münster: Waxmann.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the last decade: Role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454-499.

Sibberns, H., & Baumert, J. (2001). Anhang A: Stichprobenziehung und Stichprobengewichtung. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 511-524). Opladen: Leske + Budrich.

von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? *IERI Monograph Series, 4*, 9-36.

Anschrift der Autorin/der Autoren

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

Dr. Jan Hochweber, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hochweber@dipf.de

Prof. Dr. Andreas Frey, Friedrich-Schiller-Universität Jena, Institut für Erziehungswissenschaft, Am Planetarium 4, 07737 Jena, Deutschland
E-Mail: andreas.frey@uni-jena.de