

Jude, Nina; Hartig, Johannes; Schipolowski, Stefan; Böhme, Katrin; Stanat, Petra
Definition und Messung von Lesekompetenz. PISA und die Bildungsstandards

Jude, Nina [Hrsg.]; Klieme, Eckhard [Hrsg.]: PISA 2009 - Impulse für die Schul- und Unterrichtsforschung. Weinheim u.a. : Beltz 2013, S. 200-228. - (Zeitschrift für Pädagogik, Beiheft; 59)

urn:nbn:de:0111-opus-78285



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

Grundlagen und konzeptionelle Ansätze

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

Definition und Messung von Lesekompetenz

PISA und die Bildungsstandards

1. Einleitung

Lesekompetenz ist spätestens seit der ersten Durchführung des *Programme for International Student Assessment (PISA)* im Jahr 2000 ein regelmäßig erhobener Indikator im Rahmen des internationalen und nationalen Bildungsmonitoring. Neben PISA haben im vergangenen Jahrzehnt verschiedene *Large Scale Assessments* Aussagen über das Niveau und die Verteilung der Lesekompetenz von Schülerinnen und Schülern in Deutschland getroffen. Dazu gehören beispielsweise Studien wie Deutsch Englisch Schülerleistungen International (DESI), die Internationale Grundschul-Lese-Untersuchung (IGLU bzw. PIRLS) sowie Untersuchungen, die das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) auf Basis der länderübergreifenden Bildungsstandards der Kultusministerkonferenz durchführt (Bos et al., 2007; KMK, 2004a, 2004b; Klieme et al., 2008; Köller, Knigge & Tesch, 2010; Stanat, Pant, Böhme & Richter, 2012).

Auch wenn diese Studien jeweils Aussagen über den gleichen Kompetenzbereich, die Lesekompetenz, anstreben, so beruhen sie doch auf theoretisch und fachdidaktisch teils unterschiedlich fundierten Definitionen dieses Konstrukts. Unterschiede finden sich sowohl im verwendeten Testmaterial als auch in der Definition von Kompetenzstufen zur kriteriumsorientierten Beschreibung der von Schülerinnen und Schülern erreichten Kompetenzstände. Es stellt sich daher die Frage, inwieweit die in den genannten Studien erfassten Konstrukte vergleichbar sind und wie ihre Ergebnisse aufeinander bezogen werden können.

Für einen solchen Vergleich bietet es sich aktuell an, zwei Studien näher zu betrachten, die bezüglich ihrer Genese und Verwendung von Kompetenzstufenmodellen aufeinander aufbauen: PISA und der zeitgleich in der Sekundarstufe I durchgeführte Ländervergleich des IQB, der auf den Bildungsstandards basiert (Klieme et al, 2010; Köller et al., 2010).

Im Rahmen von PISA wird die Lesekompetenz von Fünfzehnjährigen in Deutschland seit dem Jahr 2000 im dreijährigen Rhythmus getestet und im internationalen Vergleich berichtet. Die erweiterten, „PISA-E“ genannten Erhebungen, die in den Jahren

2000, 2003 und 2006 durchgeführt wurden, ermöglichten es darüber hinaus, diese Kompetenzen auf der Ebene der Länder in der Bundesrepublik Deutschland darzustellen (Baumert et al., 2002; Prenzel et al., 2005, 2008).

Mit der Einführung der Ländervergleiche auf Basis der Bildungsstandards der KMK, die im Fach Deutsch für den Mittleren Schulabschluss im Jahr 2004 erfolgte (KMK, 2004a), wurde PISA-E im Jahr 2009 abgelöst. Während die internationale Erhebung von PISA 2009 weiterhin mit einer für Deutschland repräsentativen Stichprobe die Lesekompetenz von Fünfzehnjährigen erfasste, bezog sich die Kompetenzmessung auf Basis der Bildungsstandards im Jahr 2009 auf Schülerinnen und Schüler der neunten Jahrgangsstufe (Klieme et al., 2010; Köller et al., 2010). Wie in Abschnitt 4.2 dargestellt wird, überschneiden sich die Zielpopulationen beider Studien. Die Studien verwendeten für ihre Testung allerdings jeweils einen eigenen Lesetest.

Der vorliegende Artikel geht der Frage nach, welche Beziehung zwischen den in beiden Studien gemessenen Lesekompetenzkonstrukten besteht und inwieweit die auf den Testergebnissen der Studien basierenden Kompetenzverteilungen der Schülerinnen und Schüler in Deutschland empirisch vergleichbar sind.

Dargestellt werden zunächst die Zielsetzungen und die theoretischen Definitionen der Lesekompetenz in den beiden Studien sowie ihre jeweilige Operationalisierung in Testaufgaben und in kriterial beschreibbaren Kompetenzstufenmodellen. Anhand einer gemeinsamen Stichprobe von Schülerinnen und Schülern, die sowohl die Lesetestaufgaben von PISA 2009 als auch die der Bildungsstandards bearbeiteten, werden die empirischen Verteilungen auf den Kompetenzstufen verglichen und der latente Zusammenhang zwischen den Lesetestwerten analysiert.

2. Das Konzept der Lesekompetenz

Lesekompetenz wird als ein zentraler Aspekt von Grundbildung angesehen. Lesen zu können stellt nicht nur eine wesentliche Voraussetzung für den weiteren Kompetenzerwerb in der Schule dar, sondern ist grundlegend für die Weiterentwicklung eigenen Wissens und eigener Fähigkeiten sowie für die Teilhabe am kulturellen und gesellschaftlichen Leben auch außerhalb institutioneller Kontexte (Naumann, Artelt, Schneider & Stanat, 2010). Dabei ist Lesen kein passiver Prozess, sondern eine aktive Konstruktionsleistung des Individuums, das die im Text enthaltenen Inhalte mit seinem jeweils individuellen Vor- und Weltwissen in Verbindung bringen muss (Artelt, Stanat, Schneider & Schiefele, 2001).

Die Fähigkeit, verstehend zu lesen, umfasst neben kognitiven Prozessen auch motivationale und emotionale Aspekte (Groeben & Hurrelmann, 2004). Theoretische Konzeptionen beschreiben den Zusammenhang zwischen Lesekompetenz und Lesemotivation, und es kann argumentiert werden, dass eine aktive Lesepraxis eine Motivation zum Lesen voraussetzt (Artelt, Naumann & Schneider, 2010). Auch empirisch lassen sich Zusammenhänge zwischen Lesekompetenz und Leseinteresse sowie Lesemotivation aufzeigen (Guthrie & Wigfield, 2000; Artelt et al., 2010).

Entsprechend berücksichtigen die theoretischen Rahmenkonzeptionen von Large Scale Assessments motivationale Aspekte der Lesekompetenz (bspw. OECD, 2009a). Diese werden jedoch konzeptionell und methodisch nicht als immanenter Aspekt der Lesekompetenz aufgefasst, sondern als mit ihr zusammenhängende, separat erfassbare Aspekte. Sowohl die Testaufgaben als auch die Kompetenzstufenmodelle fokussieren daher auf das Textverstehen. Dieser Definition folgt auch der vorliegende Text.

Die Bedeutung, die der Lesekompetenz als universelles Kulturwerkzeug zukommt, ist auch in einer sich verändernden Medienlandschaft nicht geringer geworden. Im Umkehrschluss ist bei einer geringen Lesekompetenz von einem Chancennachteil nicht nur für das Individuum, sondern auch für die Gesellschaft insgesamt auszugehen.

Nicht zuletzt aus diesen Gründen ist die Erfassung von Lesekompetenz als ein wichtiger Aspekt grundlegender Bildung fester Bestandteil von Schulleistungsstudien sowohl im Grundschulalter als auch am Ende der Pflichtschulzeit. Zu letzteren zählen die Kompetenzerfassung bei Fünfzehnjährigen im internationalen Vergleich in PISA und die Kompetenzmessung auf Basis der Bildungsstandards in der neunten Jahrgangsstufe im nationalen Ländervergleich. Lesekompetenz wird dabei je nach Fokus der Studie und der zu testenden Altersgruppe unterschiedlich definiert und erfasst.

PISA und auch die Studien auf der Grundlage der Bildungsstandards arbeiten zur differenzierteren Beschreibung der Lesekompetenz mit dem Verfahren des *Standard-setting*, aus dem Kompetenzstufenmodelle resultieren. Kompetenzstufen dienen einer kriteriumsorientierten Beschreibung von Kompetenz, die über eine rein numerische Quantifizierung hinausgehen soll. Sie werden gebildet, indem die aus der Messung resultierende, kontinuierliche quantitative Kompetenzskala in qualitativ unterscheidbare Abschnitte unterteilt wird. Die so entstandenen Kompetenzstufen lassen sich inhaltlich beschreiben und es wird definiert, über welche Kompetenzen Personen mit einer bestimmten Testpunktzahl verfügen, d.h. welche Aufgabenanforderungen sie mit einer bestimmten Wahrscheinlichkeit bewältigen können. Die Bildung der Kompetenzstufen, also die Bestimmung von Schwellen zwischen verschiedenen Abschnitten der kontinuierlichen Skala, wird dabei von inhaltlichen und psychometrischen Vorgaben bestimmt (Hartig & Klieme, 2006; Hartig & Jude, 2007; Pant, Tiffin-Richards & Köller, 2010).

Im Folgenden werden die Konzepte der Lesekompetenz in PISA und den Bildungsstandards, die jeweiligen Verfahren der Testentwicklung sowie die Kompetenzstufenbildung in beiden Studien dargestellt und zusammenfassend verglichen.

2.1 Das Programme for International Student Assessment (PISA)

Hintergrund und Zielsetzung

PISA als international vergleichende Schulleistungsstudie wird von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) initiiert und erfasst die Fähigkeiten von 15-jährigen Schülerinnen und Schülern in den Bereichen der Lesekompetenz, der mathematischen Kompetenz und der naturwissenschaftlichen Kompetenz in den OECD-Staaten sowie in OECD-Partnerstaaten. PISA wird seit dem Jahr 2000 im

Abstand von drei Jahren regelmäßig international durchgeführt. In Deutschland erfolgt die Beteiligung im Auftrag der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK).

PISA ermöglicht Aussagen darüber, inwieweit Schülerinnen und Schüler gegen Ende ihrer Pflichtschulzeit über grundlegende Kompetenzen verfügen. Die Ergebnisse geben außerdem „Auskunft über demografische Kontextindikatoren sowie Merkmale der häuslichen und schulischen Umwelt, die mit den Kompetenzen in Zusammenhang stehen können, und erlauben die Analyse von Veränderungen über die Zeit“ auf Ebene der verschiedenen Bildungssysteme (Jude & Klieme, 2010, S. 13).

Die regelmäßige Teilnahme Deutschlands an PISA ermöglicht die systematische Sammlung von Informationen über das deutsche Bildungssystem und die Analyse von Veränderungen (Trends) in den Leistungen von Schülerinnen und Schülern sowie den Kontextbedingungen schulischer Lehr-Lernprozesse. PISA ist damit Teil der im Jahr 2006 verabschiedeten Gesamtstrategie der KMK zum Bildungsmonitoring (KMK, 2006), in der die Teilnahme an international vergleichenden Schulleistungsstudien eine zentrale Rolle spielt. Die Ergebnisse dieser internationalen Vergleiche ergänzen die Informationen, die aus der zentralen Überprüfung der Bildungsstandards gewonnen werden, und gehen in die gemeinsame Bildungsberichterstattung von Bund und Ländern ein (Autorengruppe Bildungsberichterstattung, 2012).

Studienkonzeption und Konstruktdefinition

PISA ist in regelmäßigen Erhebungszyklen aufgebaut: Im Abstand von jeweils drei Jahren werden drei Kompetenzbereiche getestet, wobei die Schwerpunkte zwischen den Erhebungen wechseln. Im Jahr 2000 war Lesekompetenz der Schwerpunkt, gefolgt von Mathematik im Jahr 2003 und den Naturwissenschaften im Jahr 2006. In PISA 2009 bildete die Lesekompetenz erneut den inhaltlichen Schwerpunkt.

Das Verständnis von Lesekompetenz in PISA basiert auf dem Literacy-Konzept und umfasst nach Definition der OECD „die Fähigkeit einer Person, geschriebene Texte zu verstehen, zu nutzen, über sie zu reflektieren und sich mit ihnen auseinanderzusetzen, um eigene Ziele zu erreichen, das eigene Wissen und Potenzial weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ (OECD, 2010, S. 24). PISA fokussiert auf die Kompetenzen der Jugendlichen zum Ende der Pflichtschulzeit, daher werden grundlegende Lesefertigkeiten, die zu einem früheren Zeitpunkt im Kompetenzerwerb stattfinden, nicht separat getestet.

Die Rahmenkonzeption von PISA differenziert in ihrer Definition von Lesekompetenz empirisch unterscheidbare Teildimensionen (OECD, 2009a). Diese beziehen sich zum einen auf die Anforderungen, die eine Leseaufgabe an eine Leserin oder einen Leser stellen kann: a) Informationen suchen und extrahieren (Informationsentnahme), b) textbezogenen Informationen kombinieren und Interpretationen ableiten (Integration und Interpretation), und c) über einen Text reflektieren und diesen bewerten (Reflexion). Zusätzlich zur Berichtslegung über die Gesamtskala der Lesekompetenz werden für diese Teildimensionen Kompetenzskalen gebildet und die Kompetenzstände darauf

separat berichtet. Darüber hinaus unterscheidet PISA in seinem Aufgabenmaterial zwischen kontinuierlichen und nichtkontinuierlichen Texten und berichtet auch hierfür die von den Schülerinnen und Schülern erreichten Kompetenzverteilungen auf separaten Skalen. Neben der Gesamtskala der Lesekompetenz, die in dem vorliegenden Artikel verwendet wird, werden also auf Basis des Aufgabenmaterials insgesamt fünf Teildimensionen der Lesekompetenz unterschieden, die jeweils auf spezifische Aspekte der Lesekompetenz fokussieren.

Unterschiedliche Inhalte der Lesetexte, die sich aus verschiedenen Funktionen, Situationen und Kontexten des Lesens ergeben, werden nicht als empirisch trennbare Teilkompetenzen angesehen. Sie gehen jedoch inhaltlich in die Entwicklung des Testmaterials ein, um eine breite und wirklichkeitsnahe Gestaltung des Aufgabenmaterials sicherzustellen. Dazu gehören entsprechend der Lebenswelt der Fünfzehnjährigen persönliche Texte, öffentliche und bildungsbezogene Texte sowie Texte mit berufsbezogenen Inhalten.

Aufgabenentwicklung zur Messung der Lesekompetenz in PISA

Die Entwicklung von Testaufgaben für PISA basiert auf der Rahmenkonzeption der Lesekompetenz und erfolgt in Zusammenarbeit von mehreren internationalen Expertengremien. Hierzu gehören das internationale PISA-Konsortium, die internationalen Expertenausschüsse unter Federführung der OECD sowie nationale Sachverständige in allen Teilnehmerstaaten. Aus diesem Vorgehen resultieren Testaufgaben, die eine breite kulturelle Diversität aufweisen und in den Lesetexten die unterschiedlichen Lebenskontexte von Fünfzehnjährigen berücksichtigen, die an dieser international vergleichenden Studie teilnehmen (OECD, 2012). Diese Lesetexte und Testaufgaben werden von den Teilnehmerstaaten hinsichtlich ihrer kulturellen Angemessenheit eingeschätzt.

Die Testentwicklung erfolgt in der Regel englischsprachig, so dass die Aufgaben vor der Testung in die jeweilige Landessprache übersetzt werden müssen. In einem mehrstufigen Übersetzungsprozess wird sichergestellt, dass die Lesetexte in allen sprachlichen Fassungen authentisch wirken und die Aufgaben sowohl inhaltlich als auch in Bezug auf das sprachliche Anspruchsniveau der Vorlage entsprechen. Alle neu entwickelten Testaufgaben werden vor ihrem Einsatz in der Hauptuntersuchung in einer internationalen Pilotierungsstudie erprobt. Die Ergebnisse dieser Pilotierung liefern Hinweise auf die internationale Vergleichbarkeit der Testaufgaben, auf ihre Messeigenschaften in allen Teilnehmerstaaten sowie auf die empirische Abdeckung der Kompetenzstufen. Auf Basis dieser Ergebnisse werden die Testaufgaben für die Hauptstudie ausgewählt. In der 2009 durchgeführten Hauptstudie wurden neben neu entwickelten Testaufgaben auch bereits existierende Aufgaben aus PISA 2000 eingesetzt (OECD, 2010).

Im Erhebungsdesign von PISA werden innerhalb von zwei Zeitstunden Testaufgaben aus drei verschiedenen Kompetenzbereichen bearbeitet. Um mehr Testmaterial einsetzen zu können, als in der vorgesehenen Zeit bearbeitet werden kann, wird ein sogenanntes „Multi-Matrix-Design“ verwendet. Schülerinnen und Schüler bearbeiten

unterschiedliche Testhefte, die jeweils nur eine Teilmenge der insgesamt verwendeten Testaufgaben enthalten (Naumann et al., 2010).

Um trotz der eingeschränkten Testzeit eine reliable Messung zu ermöglichen, sind die verwendeten Lesetexte eher kurz, ihr Umfang beträgt selten mehr als eine Seite. In PISA 2009 wurden international 29 Lesetexte (Stimuli) mit insgesamt 131 Testaufgaben in der Hauptstudie eingesetzt. Pro Lesetext waren also durchschnittlich fünf Fragen zu beantworten (OECD, 2009b).

Etwa zwei Drittel der Testaufgaben in PISA beziehen sich auf kontinuierliche Lesetexte (60%), die restlichen Texte sind nicht kontinuierlich oder weisen ein gemischtes Format auf. Die Hälfte der Testaufgaben ist in Multiple-Choice-Antwortformaten zu bearbeiten (48%), bei den restlichen Aufgaben sind offene Antworten zu geben. Diese können entweder aus eindeutigen Richtiglösungen („closed constructed response“, 10%) oder aus kurzen („short response“, 8%) beziehungsweise längeren offenen Antworten bestehen („open constructed response“, 34%). Die Antworten der Schülerinnen und Schüler werden bei offenen Formaten durch geschulte Rater beurteilt.

Die oben dargestellten Teildimensionen der Lesekompetenz spiegeln sich wie folgt in den Testaufgaben wider: Die Hälfte der Aufgaben bezieht sich auf die Teildimension der Integration und Interpretation, jeweils ein Viertel der Aufgaben auf die Teildimensionen der Informationsentnahme beziehungsweise der Reflexion (vgl. Tabelle 1).

Teildimensionen Antwortformate	Informations- entnahme	Integration und Interpretation	Reflexion	Gesamt
Multiple choice	6	38	8	52 (40%)
Complex multiple choice	3	6	1	10 (8%)
Closed constructed response	9	4	0	13 (10%)
Short response	10	1	0	11 (8%)
Open constructed response	3	18	24	45 (34%)
	31 (24%)	67 (51%)	33 (25%)	131 (100%)

Tab. 1: Lesetestaufgaben in PISA 2009, Verteilung auf Teildimensionen und Antwortformate (vgl. OECD, 2012, S. 41)

Die PISA-Lesetexte beziehen sich zu etwa gleichen Teilen auf persönliche, öffentliche und bildungsbezogene Inhalte; berufsbezogene Inhalte sind demgegenüber etwas weniger stark repräsentiert (OECD, 2009b).

Das Kompetenzstufenmodell für den Bereich Lesen in PISA

PISA nutzt zur Beschreibung der Lesekompetenz eine kontinuierliche Skala. In der ersten PISA-Erhebung im Jahr 2000 wurde diese Skala auf einen Mittelwert von 500 und eine Standardabweichung von 100 für die damals teilnehmenden OECD-Staaten normiert. Im Jahr 2009 wurden die verwendeten Testaufgaben zunächst unabhängig kalibriert und anschließend durch ein *Equating* auf der Skala von PISA 2000 verankert (OECD, 2012). Durch die Veränderung der Zusammensetzung der OECD-Staaten sowie Veränderungen im Niveau der Lesekompetenz der 15-Jährigen seit dem Jahr 2000 liegt für PISA 2009 der Durchschnitt aller OECD-Staaten bei 493 Punkten, die Standardabweichung bei 93 Punkten (OECD, 2010).

Das Kompetenzstufenmodell für den Bereich Lesen von PISA 2000 umfasste fünf Kompetenzstufen (Adams & Wu, 2002). Für PISA 2009 wurde das Kompetenzstufenmodell analog zur entsprechend überarbeiteten Rahmenkonzeption erweitert, und zwar speziell im unteren und oberen Kompetenzbereich. Die unterste Kompetenzstufe I wurde aufgeteilt in die Stufen Ib und Ia, und die Kompetenzstufe VI wurde im oberen Kompetenzbereich neu hinzugefügt. Dadurch ist eine stärkere Differenzierung möglich als noch im Jahr 2000. Das Kompetenzstufenmodell für die Lesekompetenz in PISA 2009 umfasst jetzt die in Tabelle 2 dargestellten sieben kriterial beschreibbaren Kompetenzstufen (OECD, 2010).

Das Kompetenzstufenmodell von PISA 2009 definiert mit der Stufe Ib explizit eine untere Grenze, unterhalb derer sich die mit den PISA-Testaufgaben erfasste Lesekompetenz nicht mehr beschreiben lässt. Über die Lesekompetenz von Personen, die sich unterhalb dieser Kompetenzstufe befinden, kann anhand von PISA lediglich die Aussage gemacht werden, welche Leseanforderungen sie mit großer Wahrscheinlichkeit *nicht* lösen. Ihre Lesekompetenz kann anhand der PISA-Testaufgaben also nicht inhaltlich kriterial beschrieben werden.

Die Definition der Kompetenzstufen basiert auf inhaltlichen und psychometrischen Kriterien und erfolgte in einem zweistufigen Verfahren. Im ersten Schritt wurden die Testaufgaben in Relation zu den Kompetenzaspekten der Rahmenkonzeption analysiert, d.h. die einzelnen Testaufgaben wurden im Hinblick darauf beurteilt, inwieweit sie die Kompetenzaspekte der Rahmenkonzeption repräsentieren. Daraus resultierte die Beschreibung der kognitiven Anforderungen jeder Testaufgabe – von leichten Aufgaben bis hin zu schwierigen Aufgaben. Diese Beschreibungen wurden zusammen mit den empirischen Schwierigkeiten jeder Testaufgabe verwendet, um in einem zweiten Schritt die Grenzen zwischen den Kompetenzstufen zu definieren. Die Setzung von Schwellenwerten („cut-offs“) bestimmte letztlich die Zuordnung von Testaufgaben zu Kompetenzstufen (vgl. zum genauen Vorgehen OECD, 2012). Die resultierenden Kompetenzstufenmodelle bilden die Grundlage für die Ergebnisberichte von PISA.

Stufe	Wertebereich	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
VI	> 698	Jugendliche auf dieser Stufe können Schlussfolgerungen, Vergleiche und Gegenüberstellungen detailgenau und präzise anstellen. Dabei entwickeln sie ein volles und detailliertes Verständnis eines oder mehrerer Texte und verbinden dabei unter Umständen gedanklich Informationen aus mehreren Texten miteinander. Hierbei kann auch die Auseinandersetzung mit ungewohnten Ideen gefordert sein, genauso wie der kompetente Umgang mit konkurrierenden Informationen und abstrakten Interpretationskategorien sowie hohe Präzision im Umgang mit zum Teil unauffälligen Textdetails.
V	626-697	Jugendliche auf dieser Stufe können sowohl mehrere tief eingebettete Informationen finden, ordnen und herausfinden, welche davon jeweils relevant sind, als auch ausgehend von Fachwissen eine kritische Beurteilung oder Hypothese anstellen. Die Aufgaben dieser Stufe setzen in der Regel ein volles und detailliertes Verständnis von Texten voraus, deren Inhalt oder Form ungewohnt ist. Zudem muss mit Konzepten umgegangen werden können, die im Gegensatz zum Erwarteten stehen.
IV	553-625	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, linguistischen oder thematischen Verknüpfungen in einem Text über mehrere Abschnitte zu folgen, oftmals ohne Verfügbarkeit eindeutiger Kennzeichen im Text, um eingebettete Informationen zu finden, zu interpretieren und zu bewerten oder um psychologische oder philosophische Bedeutungen zu erschließen. Insgesamt muss ein genaues Verständnis langer oder komplexer Texte, deren Inhalt oder Form ungewohnt sein kann, unter Beweis gestellt werden.
III	480-552	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, vorhandenes Wissen über die Organisation und den Aufbau von Texten zu nutzen, implizite oder explizite logische Relationen (z.B. Ursache-Wirkungs-Beziehungen) über mehrere Sätze oder Textabschnitte zu erkennen, mit dem Ziel, Informationen im Text zu lokalisieren, zu interpretieren und zu bewerten. Einige Aufgaben verlangen vom Leser/von der Leserin, einen Zusammenhang zu begreifen oder die Bedeutung eines Wortes oder Satzes zu analysieren. Häufig sind die benötigten Informationen dabei nicht leicht sichtbar oder Passagen des Textes laufen eigenen Erwartungen zuwider.
II	408-479	Jugendliche auf dieser Stufe können innerhalb eines Textabschnitts logischen und linguistischen Verknüpfungen folgen, mit dem Ziel, Informationen im Text zu lokalisieren oder zu interpretieren; im Text oder über Textabschnitte verteilte Informationen aufeinander beziehen, um die Absicht des Autors zu erschließen. Bei Aufgaben dieser Stufe müssen unter Umständen auf der Grundlage eines einzigen Textbestandteils Vergleiche und Gegenüberstellungen vorgenommen werden oder es müssen, ausgehend von eigenen Erfahrungen oder Standpunkten, Vergleiche angestellt oder Zusammenhänge zwischen dem Text und nicht im Text enthaltenen Informationen erkannt werden.
Ia	335-407	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, in einem Text zu einem vertrauten Thema eine oder mehrere unabhängige, explizit ausgedrückte Informationen zu lokalisieren, das Hauptthema oder die Absicht des Autors zu erkennen oder einen einfachen Zusammenhang zwischen den im Text enthaltenen Informationen und allgemeinem Alltagswissen herzustellen. Die erforderlichen Informationen sind in der Regel leicht sichtbar, und es sind nur wenige beziehungsweise keine konkurrierenden Informationen vorhanden. Der Leser wird explizit auf die entscheidenden Elemente in der Aufgabe und im Text hingewiesen.
Ib	262-334	Jugendliche auf dieser Stufe können in einem kurzen, syntaktisch einfachen Text aus einem gewohnten Kontext, dessen Form vertraut ist (z.B. in einer einfachen Liste oder Erzählung), eine einzige, explizit ausgedrückte Information lokalisieren, die leicht sichtbar ist. Der Text enthält in der Regel Hilfestellungen für den Leser, wie Wiederholungen, Bilder oder bekannte Symbole. Es gibt kaum konkurrierende Informationen. Bei anderen Aufgaben müssen einfache Zusammenhänge zwischen benachbarten Informationsteilen hergestellt werden.
unter Ib	< 262	Keine kriteriale Beschreibung der Kompetenz möglich.

Tab. 2: Wertebereich auf der PISA-Skala und inhaltliche Beschreibung der Kompetenzstufen von PISA 2009 (Naumann et al., 2010)

2.2 Die Studien auf Basis der Bildungsstandards

Hintergrund und Zielsetzung

Vor dem Hintergrund des unbefriedigenden Abschneidens der Schülerinnen und Schüler in deutschen Schulen in internationalen Vergleichsstudien seit TIMSS (Baumert, Bos & Lehmann, 2000) wurde eine Reihe von Maßnahmen zur Qualitätssicherung und -entwicklung im deutschen Schulsystem beschlossen (KMK, 1997, 2006). Eine wichtige Grundlage dieser Maßnahmen bilden die abschlussbezogenen länderübergreifenden Bildungsstandards, die seit 2003/2004 unter anderem für den Mittleren Schulabschluss im Fach Deutsch vorliegen. Bildungsstandards stellen verbindlich und fachspezifisch formulierte Kernziele schulischer Bildung dar und benennen jene Kompetenzen, die Schülerinnen und Schüler im jeweiligen Fach bis zu einem bestimmten Zeitpunkt in ihrer Bildungsbiografie im Regelfall erworben haben sollten. Bildungsstandards sollen zudem helfen, den Fokus von einer Input- hin zu einer Output-Orientierung zu verlagern, bei der nicht Lerninhalte, sondern Bildungsergebnisse die zentrale Größe in der Steuerung des deutschen Bildungssystems darstellen.

Im Zuge der Einführung der Bildungsstandards wurde im Jahr 2004 das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) als An-Institut der Humboldt-Universität zu Berlin gegründet, dessen Auftrag in der Weiterentwicklung, Operationalisierung, Normierung und Überprüfung des Erreichens der Bildungsstandards besteht. Mit den sog. Plöner Beschlüssen aus dem Jahr 2006 (KMK, 2006) wurde zudem entschieden, auf Basis der Bildungsstandards regelmäßig nationale Ländervergleiche durchzuführen. Diese nationalen Studien ergänzen seit dem Jahr 2009 die internationalen Schulleistungsvergleiche PISA (Sekundarstufe I) und IGLU/PIRLS sowie TIMSS (Primarstufe), indem sie detaillierte Informationen darüber bereitstellen, inwieweit die Schülerinnen und Schüler in den 16 deutschen Ländern die in den Bildungsstandards definierten Leistungserwartungen erfüllen (Köller et al., 2010; Stanat et al., 2012).

Studienkonzeption und Konstruktdefinition

Die Durchführung der Ländervergleiche auf Basis der Bildungsstandards im Bereich der Sekundarstufe I ist an die internationalen PISA-Studien angebunden und erfolgt daher ebenfalls im dreijährigen Zyklus. Dabei wird jeweils eine Auswahl der Kompetenzbereiche getestet, für die in der Sekundarstufe I Bildungsstandards vorliegen. Im Ländervergleich 2009 wurden Kompetenzen in der Verkehrssprache Deutsch, darunter die Lesekompetenz, sowie in der ersten Fremdsprache Englisch bzw. Französisch erhoben. In diesem Durchgang lag auch bei PISA der Schwerpunkt auf der Lesekompetenz. Während jedoch bei PISA die internationale Rahmenkonzeption die Grundlage des Lesetests bildete (vgl. Abschnitt 2.1, Studienkonzeption und Konstruktdefinition), basierte die Messung der Lesekompetenz im Ländervergleich auf den Bildungsstandards.

In den Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss (MSA) umfasst der Kompetenzbereich „Lesen – mit Texten und Medien umgehen“ zum einen das Verstehen von literarischen Texten und Sachtexten, zum anderen jedoch auch die

ästhetische Dimension des Lesens sowie die Förderung von Leseinteresse und Lese Freude (KMK, 2004a). Da die beiden letztgenannten Aspekte im Kontext eines Large Scale Assessments kaum mit Testaufgaben operationalisierbar sind, liegt der Fokus der Konstruktdefinition auf den Standards „literarische Texte verstehen und nutzen“ sowie „Sach- und Gebrauchstexte verstehen und nutzen“ (vgl. KMK, 2004a, S. 14-15). Gemäß dieser Definition wird Lesekompetenz in den Studien des IQB auf Basis der Bildungsstandards als die Fähigkeit verstanden, auch längere Texte sowie nichtlineare Texte, wie z.B. Schaubilder, zu verstehen und im Detail zu erfassen. Hierzu gehören die zielgerichtete Entnahme und das Ordnen von Informationen aus einem Text sowie die Verknüpfung von verschiedenen im Text enthaltenen Informationen miteinander und mit dem eigenen Vorwissen. Schülerinnen und Schüler, die die Regelstandards erreichen, sind ferner in der Lage, diese Informationen zu interpretieren und hinsichtlich ihrer Aussagekraft und Wertung zu prüfen, um die Intention eines Textes zu erkennen und begründete Schlussfolgerungen zu ziehen (vgl. Böhme, Neumann & Schipolowski, 2010). Lesekompetenz umfasst somit Rezeptionsleistungen vom Detailverstehen über Globalverstehen bis hin zur kritischen Reflexion von Texten. Des Weiteren beinhaltet die Definition der Lesekompetenz im Rahmen der Bildungsstandards auch die Kenntnis sprachlicher Gestaltungsmittel sowie das Erkennen und Unterscheiden verschiedener literarischer Gattungen, Textsorten und -funktionen, die in den Bildungsstandards als weitere wesentliche Aspekte des Verstehens und der Nutzung von Texten genannt werden.

Auf Basis dieser Konstruktdefinition erfolgt die Entwicklung von Testaufgaben zur Erfassung der Lesekompetenz in den nationalen Ländervergleichsstudien.

Entwicklung von Testaufgaben zur Messung der Lesekompetenz

Bei der Entwicklung von Testaufgaben steht der Innovationsanspruch der Bildungsstandards im Vordergrund, der – in Anlehnung an PISA – unter anderem in der Kompetenzorientierung sowie in der stärkeren Berücksichtigung kontinuierlicher und diskontinuierlicher Sachtexte zum Ausdruck kommt. Gleichzeitig wird ein gewisses Maß an curricularer Validität angestrebt, die durch die Genese der Bildungsstandards gewährleistet wird. Diese beruhen auf fachspezifischen Kompetenzmodellen, die sowohl international anerkannte Modelle wie die theoretischen Grundlagen der PISA-Studien berücksichtigen als auch „aus der Erfahrung der Schulpraxis heraus entwickelt wurden“ (KMK, 2004a, S. 3). Zudem soll der im Folgenden beschriebene Aufgabenentwicklungsprozess durch die Einbeziehung unterrichtender Lehrkräfte sicherstellen, dass Elemente der aktuellen unterrichtlichen Gestaltung in die Testentwicklung einfließen.

Die Testaufgaben für die bildungsstandardbasierten Kompetenzmessungen werden von erfahrenen Lehrkräften aus ganz Deutschland entwickelt. Diese werden durch Expertinnen und Experten aus der jeweiligen Fachdidaktik, der empirischen Bildungsforschung und Psychometrie geschult und beraten. Jede Aufgabe wird in einem mehrstufigen Review-Prozess wiederholt beurteilt, in kleinen Schülergruppen präpilotiert, überarbeitet und schließlich im Rahmen einer größeren Pilotierungsstudie erprobt. Basierend auf den empirischen Ergebnissen dieser Pilotierung werden besonders geeig-

nete Aufgaben ausgewählt, gegebenenfalls erneut überarbeitet und in einer für ganz Deutschland repräsentativen Erhebung normiert (Normierungsstudie). Die aus diesem Entwicklungsprozess resultierenden Aufgaben werden schließlich im Rahmen verschiedener Studien zur Kompetenzmessung auf Basis der Bildungsstandards eingesetzt (Pöhlmann, Neumann, Tesch & Köller, 2010; Granzer, 2009).

Im 2009 durchgeführten Ländervergleich Sprachen kamen für das Fach Deutsch im Bereich Lesen acht authentische Lesetexte (Stimuli) zum Einsatz, darunter drei literarische Texte und fünf Sachtexte. Die Länge der Textstimuli betrug zwischen einer halben und drei Seiten, wobei etwa anderthalb Seiten die Regel waren. Es handelte sich in jedem Fall um Fließtexte, die jedoch teilweise durch Abbildungen ergänzt oder aus mehreren Quellen zusammengeführt wurden. Pro Stimulus waren zwischen fünf und zwölf Testaufgaben zu bearbeiten, die teilweise aus mehreren Teilaufgaben (Items) bestanden. Insgesamt kamen 75 Testaufgaben zum Einsatz, zu denen insgesamt 99 Items bearbeitet werden sollten. Aufgrund der begrenzten Testzeit wurde – wie bei der PISA-Erhebung – ein komplexes Testdesign verwendet (vgl. Böhme, Leucht et al., 2010), bei dem jedem Schüler bzw. jeder Schülerin nur ein Teil der gesamten Aufgabenmenge vorgelegt wurde.

Die Antwortformate umfassten neben Multiple-Choice-Aufgaben mit vier Antwortalternativen (ca. 44% der Teilaufgaben) auch sog. Richtig-Falsch-Aufgaben, bei denen für vorgegebene Aussagen entschieden werden musste, ob sie zutreffen oder nicht (12%). Ein weiteres geschlossenes Antwortformat stellten Zuordnungs- und Umordnungsaufgaben dar (8%). Bei offenen Antwortformaten wurde unterschieden zwischen Kurzantworten, für die eine intersubjektiv eindeutige Richtiglösung definiert werden kann (15%), und Aufgabenstellungen, deren Beantwortung überwiegend Interpretationsleistungen erfordert und bei deren Auswertung daher ein gewisser Beurteilungsspielraum besteht (21%).

Das Kompetenzstufenmodell für den Bereich Lesen

Seit April 2009 liegt für den Bereich „Lesen – mit Texten und Medien umgehen“ ein fachdidaktisch fundiertes Kompetenzstufenmodell zu den Bildungsstandards für den Mittleren Schulabschluss (MSA) vor (Bremerich-Vos et al., 2010). Dieses Modell wurde für Schülerinnen und Schüler am Ende der Pflichtschulzeit entwickelt, die den MSA anstreben. Die Datenbasis des Modells bildet eine Normierungsstudie des IQB aus dem Jahr 2008, in die Schülerinnen und Schüler aus den Jahrgangsstufen acht bis zehn aller Schularten des allgemeinbildenden deutschen Schulsystems einbezogen wurden. Die Festlegung der zugrunde liegenden Metrik orientierte sich am internationalen Kompetenzstufenmodell von PISA aus dem Jahr 2000 (Bremerich-Vos et al., 2010); d.h. die Kompetenzskala wurde so definiert, dass sie in der Population der Neuntklässler in Deutschland einen Mittelwert von 496 Punkten und eine Standardabweichung von 92 Punkten aufweist.

Wie in anderen Modellen zu den Bildungsstandards wurden fünf Kompetenzstufen definiert, deren Schwellen im gleichen Abstand zueinander liegen. Für den Bereich der

Lesekompetenz in Deutsch beträgt die Stufenbreite jeweils 80 Punkte. Die unterste sowie die oberste Stufe sind nach unten bzw. nach oben offen. Die Fähigkeit jeder Person mit einem gültigen Testwert kann also einer kriterial beschreibbaren Kompetenzstufe zugeordnet werden.

Die zur Differenzierung der Kompetenzstufen verwendeten Cut-off-Werte und die inhaltlichen Beschreibungen der Kompetenzstufen sind in Tabelle 3 dargestellt.

Da die auf den einzelnen Stufen beschreibbaren Aspekte von Lesekompetenz mit den in den Bildungsstandards formulierten Kompetenzerwartungen verknüpft wurden, geben diese Hinweise darauf, inwieweit Regelanforderungen erfüllt oder verfehlt werden. Leseleistungen auf Kompetenzstufe III werden als das Erreichen der Regelstandards für den MSA interpretiert. Weiterhin wurde festgelegt, dass die Kompetenzstufe II als Mindeststandards zu interpretieren ist, der von allen Schülerinnen und Schülern, die den MSA anstreben, erreicht werden sollte.

Die Stufenbeschreibungen beziehen sich auf die in der Konstruktdefinition genannten Verständnisleistungen wie das Lokalisieren von Informationen, deren Verknüpfung zur Ableitung von Schlussfolgerungen sowie Interpretation und Reflexion (vgl. Abschnitt 2.2, Studienkonzeption und Konstruktdefinition). Ferner wird explizit auf unterschiedliche Textsorten, Textaufbau, Erzählperspektive sowie die Rolle des Antwortformats eingegangen. Die teilweise gegebene Aufgabennähe der Deskriptoren hängt damit zusammen, dass das Modell nicht allein für die Zwecke des Bildungsmonitoring auf Systemebene erstellt wurde, sondern auch im Rahmen von Unterrichtsentwicklung Anwendung finden soll. In diesem Zusammenhang kommt den exemplarischen Aufgabenstellungen, die eine Konkretisierung von Anforderungen ermöglichen, eine größere Bedeutung zu.

2.3 Zusammenfassender Vergleich

In den vorangegangenen Abschnitten wurden die Definitionen der Lesekompetenz in PISA und der Bildungsstandards dargestellt und beschrieben, wie diese Konstrukte mit Testaufgaben und Kompetenzstufenmodellen operationalisiert werden. Diese Informationen werden nun im Folgenden vergleichend gegenübergestellt.

Hintergrund und Zielsetzung der Studien

Beide hier dargestellten Studien erfassen die Lesekompetenz von Schülerinnen und Schülern am Ende der Sekundarstufe I. PISA wählt dabei eine altersbasierte Populationsdefinition der Fünfzehnjährigen, um einen vom jeweiligen Schulsystem unabhängigen, internationalen Vergleich der Kompetenzen innerhalb einer Alterskohorte zu ermöglichen. Den Ländervergleichsstudien auf Basis der Bildungsstandards liegt dagegen eine jahgangsbasierte Populationsdefinition zugrunde, d.h. es werden Schülerinnen und Schüler in der neunten Jahrgangsstufe getestet. Diese beiden Zielgruppen überschneiden sich in der Schülerpopulation in Deutschland zu ca. 50% (Jude & Klieme,

Stufe	Wertebereich	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
V	> 640	<p>Schülerinnen und Schülern lösen hier nicht nur mit hoher Sicherheit Aufgaben der Stufen I bis IV. Es gelingt ihnen auch, Interpretationshypothesen plausibel zu beurteilen und in einem argumentativen Text zentrale Thesen zu identifizieren. Dabei geht es auch um das Erkennen der argumentativen Funktion von Beispielen. Ebenso können in einem umfangreichen Text einzelne Stellen als Fälle von positiven oder negativen Bewertungen identifiziert werden. Wörter und Wortgruppen in fiktionalen Texten, die für die Interpretation zentral sind, können verständlich erläutert werden. Bei einem weiteren Typ von (Teil-) Aufgaben, die auf dieser Stufe gelöst werden können, ist eine nur partiell ausgeführte schematische Darstellung relevanter Elemente eines langen Sachtexts vorgegeben. Fehlende Elemente sind einzufügen, was nur gelingt, wenn im Text verstreute Informationen identifiziert und verknüpft werden können. Darüber hinaus müssen für die korrekte Platzierung der Einträge in dem Schema Analogien berücksichtigt werden. Anders als auf den Stufen I bis IV ist bei einem großen Teil der Aufgaben auf Kompetenzstufe V verlangt, selbstständig einen mehr oder weniger umfangreichen Text zu produzieren. Insgesamt werden auf der Stufe V Leistungen gezeigt, die vermutlich nur unter sehr günstigen schulischen und außerschulischen Lerngelegenheiten erreichbar sind. Wir interpretieren diese Stufe dementsprechend im Sinne der Erreichung eines Maximalstandards.</p>
IV	560-639	<p>Auf dieser Stufe können Aufgaben der bisher beschriebenen Arten gelöst werden, das heißt unter anderem Aufgaben zur Strukturierung von Texten in Abschnitte und zur Angabe der Textsorte. Neben Aufgaben, die sich auf die Verknüpfung von Informationen über Abschnitte hinweg oder auch auf lokale Relationen von Informationen beziehen, sind auch noch Fälle von Lokalisierungsaufgaben im Spiel. Hier gibt es allerdings anders als auf den unteren Stufen kaum noch eine Identität oder offensichtliche Ähnlichkeit von Aufgabenformulierung und Darstellung im Text. Unter Berücksichtigung des gesamten (narrativen) Textes können darüber hinaus verschiedenen Figuren Merkmale zugeordnet werden. Schülerinnen und Schüler können auch über einen Text verstreute Varianten der Bezeichnung eines Gegenstandes nennen und sich begründet für eine von mehreren Interpretationen eines Textabschnitts entscheiden. Im Text nicht explizit genannte, sondern zu erschließende Wissensbestände von Figuren werden ebenso erkannt wie das Motiv eines Erzählers, einen Sachverhalt auf eine bestimmte Weise darzustellen. Auf dieser Stufe können auch Aufgaben gelöst werden, die sich auf Kombinationen von Grafik und kontinuierlichem Text beziehen. Dabei müssen unter anderem Informationen in beiden Formaten miteinander abgeglichen werden und es geht auch darum, den partiell hierarchischen Charakter der Grafik, das heißt die Einbettung einer Liste in eine andere Liste, zu verstehen. Auf Stufe IV werden somit insgesamt Leistungen gezeigt, die bereits über den Vorgaben der KMK liegen, so dass wir diese Stufe im Sinne der Erreichung eines Regelstandards plus interpretieren.</p>
III	480-559	<p>Über die bereits auf den Stufen I und II lösbaren Aufgaben hinaus werden hier Aufgaben gemeistert, in deren Rahmen unter anderem bei längeren Texten vorgegebene Absatzüberschriften in die richtige Reihenfolge zu bringen sind. Bei einem anderen auf die Textgliederung zielenden Aufgabentyp sind Absatzüberschriften Zeilenintervalle zuzuordnen. Hinzu kommen Aufgaben, bei denen bei komplexeren Texten aus mehreren vorgegebenen Alternativen das zutreffende Textthema zu wählen ist. Es gelingt auch – ebenfalls im Blick auf den ganzen Text – im Kontext einer Mehrfachwahlaufgabe die Erzählperspektive anzugeben. Weitere Aufgaben zielen auf die Differenzierung von Textsorten, zum Beispiel auf die Differenz von historisch und phantastisch, und darauf anzugeben, welches von mehreren Merkmalen am ehesten geeignet ist, einen Text der Phantastik zuzuweisen. Es werden auch Aufgaben gelöst, bei denen die Bedeutung von Wörtern kontextuell erschlossen werden muss. Bereits auf ein zentrales Element einer Interpretation als Resultat eines komplexen Schlussprozesses zielt die auf diesem Niveau beantwortbare Frage, auf welche von vier genannten Gruppen sich ein Personalpronomen bezieht. Schülerinnen und Schüler, die diese oder eine höhere Stufe erreichen, zeigen Leistungen, wie sie in den Bildungs-</p>

Stufe	Wertebereich	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
II	400-479	<p>standards der KMK für den Bereich Lesen verfasst sind. Die Stufe III kann dementsprechend im Sinne der Erreichung von Regelstandards interpretiert werden.</p> <p>Auch auf dieser Stufe werden vor allem Aufgaben gelöst, bei denen es um Lokalisieren geht. Die gesuchten Informationen sind häufig weniger auffällig platziert. Über Aufgaben hinaus, bei denen Relationen von im Text benachbarten Informationen zu explizieren sind, werden auch einige Aufgaben bewältigt, bei denen die Verbindung von Informationen verlangt ist, die über mehrere Abschnitte verteilt sind. Vereinzelt werden hier auch Aufgaben gelöst, bei denen es darum geht, auf der Basis mehrerer Informationen eine angemessene Bezeichnung für das Handlungsmotiv eines Protagonisten auszuwählen. Häufig sind wie auf Stufe I Teile der Aufgabe mit Teilen der gesuchten Information bezeichnungs- oder bedeutungsgleich oder sie sind ihnen ähnlich. Das für die Lösung einiger Aufgaben nötige Weltwissen kann als weithin geteilt angesehen werden. So muss zum Beispiel im Kontext einer Aufgabe zu einem ICE-Prospekt geschlossen werden, dass man ein Problem mit einer Fahrkarte im Wagen mit einem Service Point zur Sprache bringen kann. Bei einer Reihe von Aufgaben, bei denen Lokalisieren gefragt ist, gibt es einen, manchmal auch mehrere attraktive Distraktoren im Text. Gelöst werden auch Aufgaben, bei denen ein mehrfaches Durchsuchen von Texten nach einzelnen Informationen gefordert ist. Neben Mehrfachwahlaufgaben kommen Aufgaben vor, bei denen ein Kurztext zu schreiben ist. Darüber hinaus gibt es einige Zuordnungs- und mehrteilige Wahrfalsch-Aufgaben. Schülerinnen und Schüler, die diese Kompetenzstufe erreichen, zeigen Leistungen, mit denen die in den Bildungsstandards der KMK implizierten Kompetenzstände noch nicht erreicht werden. Da einfache Verknüpfungen aber bereits gelingen und auch über den Text verstreute Informationen integriert werden können, interpretieren wir diese Stufe im Sinne der Erreichung von Mindeststandards.</p>
I	< 400	<p>Es können einzelne Informationen im Text lokalisiert werden, vor allem dann, wenn sie auffällig platziert sind (z. B. grafisch hervorgehoben oder zu Beginn oder am Ende von Absätzen). Gesucht werden häufig Bezeichnungen für Konkretes wie Eigennamen. Dabei kommt es gelegentlich vor, dass der Text plausible Optionen enthält, also Informationen, die der gesuchten mehr oder weniger ähneln. Es gelingt ebenfalls, auch längere und komplexere Texte auf das Vorhandensein einer einzelnen Information hin zu durchsuchen. Darüber hinaus werden Aufgaben gelöst, bei denen es um die Herstellung lokaler Kohärenz geht. Im Text benachbarte Informationen, deren Relationen nicht explizit sind, werden auf der Basis von verbreitetem Weltwissen beziehungsweise vor allem aufgrund von Sprachwissen miteinander verknüpft. Dabei ist das, was in der Aufgabe als Information bereits gegeben ist, oft mit dem Wortlaut beziehungsweise der Bedeutung eines Teils der Information, die im Text identifiziert werden muss, identisch oder verwandt. So ist zum Beispiel gefragt „Wer verfolgt die Prinzessin?“, und im Text ist der (Teil-) Satz „... dass sich die Prinzessin auf der Flucht vor den ... Trollen weigert weiterzulaufen“, zu ergänzen. Gelöst werden auch Aufgaben, bei denen angesichts eines strukturell einfachen und kurzen Textes das Textthema identifiziert werden muss beziehungsweise bei denen thematisch Relevantes von weniger Relevantem unterschieden werden muss. Es handelt sich meistens um geschlossene Aufgabenformate, vor allem um Mehrfachwahlaufgaben. Schülerinnen und Schüler, die auf dieser Stufe liegen, verfehlen deutlich die Vorgaben, wie sie in den Bildungsstandards der KMK für den Bereich Lesen verfasst sind. Die Stufe I beschreibt dementsprechend einen Leistungsabschnitt, auf dem auch ein Bildungsminimum im Sinne von Mindeststandards nicht erreicht wird.</p>

Tab. 3: Wertebereich auf der Skala der Bildungsstandards im Bereich Lesen in der Sekundarstufe I und inhaltliche Beschreibung der Kompetenzstufen des Lesemodells auf Basis der Bildungsstandards für den Mittleren Schulabschluss (Bremerich-Vos et al., 2010, S. 37)

2010). Es ist daher davon auszugehen, dass sich die im Durchschnitt erreichten Kompetenzniveaus dieser Populationen unterscheiden, weil die Fünfzehnjährigen in Deutschland in Abhängigkeit vom Einschulungsalter und der Bildungslaufbahn die Klassen 7 bis 10 besuchen (vgl. auch Mildner, Hochweber & Frey in diesem Heft).

Konstruktdefinition

Beide Studien greifen auf umfassende Modelle der Lesekompetenz zurück, die sowohl kognitive als auch motivationale Aspekte beinhalten. Letztere werden in beiden Studien bewusst nicht als Teil der über Testaufgaben zu messenden Kompetenz operationalisiert, sondern über begleitende Kontextfragebögen erhoben. Unterschiede finden sich in der konzeptionellen Ausrichtung der Konstruktdefinitionen jedoch dahingehend, dass PISA explizit von einem allgemeingültigen Literacy-Konzept ausgeht, das für alle teilnehmenden Staaten relevant ist, während das Konzept der Lesekompetenz bei den Bildungsstandards auf das deutsche Bildungssystem ausgerichtet ist. So flossen bereits in die Entwicklung der Bildungsstandards Erfahrungen aus der Schulpraxis in Deutschland ein. Darüber hinaus wird der Unterrichtsbezug jedoch in erster Linie durch den Aufgabenentwicklungsprozess realisiert, d.h. auf der Ebene der Operationalisierung.

PISA und die Studie auf Basis der Bildungsstandards differenzieren das Konstrukt der Lesekompetenz konzeptuell weiter aus, etwa durch die Unterscheidung von Textformaten. Beide Studien gehen also von einer Mehrdimensionalität der Lesekompetenz aus. In der Skalierung und Berichtlegung weist jedoch nur PISA empirisch fünf trennbare Teildimensionen der Lesekompetenz aus. Gemeinsam ist beiden Studien wiederum, dass sie die Lesekompetenz von Schülerinnen und Schülern (auch) auf einer Gesamtskala berichten.

Aufgabenentwicklung

Basierend auf der jeweiligen Konstruktdefinition erfolgt die Aufgabenentwicklung in beiden Studien durch geschulte Expertengruppen. Für die Studie auf Basis der Bildungsstandards sind dies Lehrerinnen und Lehrer aus Deutschland, die durch Fachdidaktiker und Psychometriker aus dem deutschsprachigen Raum angeleitet werden. Die Testaufgaben sollen einerseits kompetenzorientiert sein, in ihrer Gestaltung jedoch auch die Unterrichtspraxis in deutschen Schulen berücksichtigen und einen gewissen curricularen Bezug aufweisen. Bei PISA werden die Testaufgaben hingegen gemäß der internationalen Ausrichtung des Programms unter Beteiligung der Teilnehmerstaaten von internationalen Expertengruppen entwickelt. Sie beinhalten kulturell vielfältiges Lesematerial aus verschiedenen Staaten, das der Alltagswelt von Fünfzehnjährigen entsprechen soll. Die Testaufgaben von PISA sind nicht explizit curricular ausgerichtet.

Weitere Unterscheide zwischen beiden Studien bestehen dahingehend, dass für PISA die konzeptionell angenommenen Teildimensionen der Lesekompetenz nicht nur bei der Aufgabenentwicklung, sondern auch bei der Auswertung und Berichtlegung berücksichtigt werden. Darüber hinaus werden insgesamt mehr offene Antwortformate ver-

wendet als bei der Studie auf Basis der Bildungsstandards. Die in PISA verwendeten Lesetexte sind demgegenüber im Mittel kürzer als die Lesetexte im bildungsstandardbasierten Ländervergleich.

Kompetenzstufenmodell

PISA und die Studie auf Basis der Bildungsstandards veranschaulichen die kontinuierliche Kompetenzskala mit Hilfe von Kompetenzstufenmodellen. Das Kompetenzstufenmodell der Bildungsstandards für den Mittleren Schulabschluss im Bereich Lesen orientiert sich mit der Unterscheidung von fünf Kompetenzstufen am Modell für den Bereich Lesen von PISA 2000. Auch die Skalierung des Lesetests entspricht in wesentlichen Aspekten dem methodischen Vorgehen von PISA 2000.

Trotz ihrer Ähnlichkeit unterscheiden sich die Kompetenzstufenmodelle jedoch in zwei wesentlichen Punkten: Während das Kompetenzstufenmodell von PISA eine untere Grenze ausweist („unter Kompetenzstufe Ib“), unterhalb derer der Test definitionsgemäß nicht mehr in der Lage ist, valide zu messen, weicht das Modell der Lesekompetenz der Bildungsstandards hiervon ab: Die Kompetenzstufe I ist nach unten offen. Jeder Schülerin und jedem Schüler, die bzw. der an der Testung auf Basis der Bildungsstandards teilnimmt, wird also mindestens eine Lesekompetenz der Stufe I attestiert, sofern die Testung nicht vollständig verweigert wurde. In PISA hingegen wird die Lesekompetenz der Schülerinnen und Schüler unter Kompetenzstufe Ib nicht mehr kriterial beschrieben.

Ein weiterer wesentlicher Unterschied besteht in der Zielpopulation der Modelle. Während im PISA-Modell die Kompetenz der Fünfzehnjährigen aller Schulformate abgebildet werden soll, wurde das Modell der Bildungsstandards für die leistungsstärkere Population derjenigen Schülerinnen und Schüler entwickelt, die in Deutschland den Mittleren Schulabschluss anstreben. Um in der leistungsheterogenen Population der Fünfzehnjährigen im oberen und unteren Bereich besser differenzieren zu können, weist das für PISA 2009 überarbeitete Modell der Lesekompetenz nun sieben Kompetenzstufen auf, d.h. es werden mehr Stufen ausgewiesen als im Modell der Bildungsstandards.

Die hier vergleichend gegenübergestellten konzeptuellen Gemeinsamkeiten und Unterschiede der beiden Studien im vorliegenden Artikel bilden den Hintergrund, vor dem die empirischen Ergebnisse im Folgenden diskutiert und interpretiert werden sollen. Ein inhaltlicher Vergleich der Kompetenzstufenbeschreibungen in den beiden Studien ist explizit nicht vorgesehen. Hierzu wäre eine umfassende Einschätzung durch eine Gruppe unabhängiger Experten nötig. Der Vergleich solcher Einschätzungen wäre zwar durchaus interessant, für die Fragestellung des vorliegenden Artikels, nämlich der empirischen Vergleichbarkeit der Ergebnisse beider Studien, jedoch nicht erforderlich.

Um die empirischen Befundmuster der Analysen dieses Artikels einordnen zu können, werden im Folgenden zunächst empirische Befunde vergleichbarer Studien dargestellt.

3. Empirische Befunde zur Vergleichbarkeit von internationalen und nationalen Kompetenztests

Die verschiedenen bisher in Deutschland durchgeführten nationalen und internationalen Schulleistungsstudien untersuchen teilweise den gleichen Kompetenzbereich, es werden jedoch unterschiedliche, speziell für den Zweck der jeweiligen Studie entwickelte Testaufgaben eingesetzt, die jeweils auf separaten Kompetenzdefinitionen basieren. Dabei stellt sich die Frage, inwieweit die Ergebnisse dieser unterschiedlichen Studien zum gleichen Kompetenzbereich vergleichbar sind. Im Folgenden wird dargestellt, wie sich diese Frage untersuchen lässt und welche Ergebnisse Analysen von Daten früherer Studien in Deutschland erbracht haben.

Verschiedene konzeptionelle und methodische Ansätze sind denkbar, um dieser Frage nachzugehen. So lässt sich analysieren, a) ob verschiedenen Studien die gleiche Kompetenzdefinition zugrunde liegt, b) ob die gleichen oder zumindest inhaltlich vergleichbare Testinstrumente verwendet werden und c) wie eng die aus unterschiedlichen Tests resultierenden Testwerte zusammenhängen. Darüber hinausgehend kann d) im Sinne eines *Equating* untersucht werden, ob Schülerinnen und Schüler auf Basis ihrer Testwerte in verschiedenen Tests der gleichen Kompetenz(stufe) zugewiesen werden.

Inzwischen haben verschiedene Publikationen die empirischen Zusammenhänge zwischen Kompetenztestwerten aus in Deutschland durchgeführten Schulleistungsstudien analysiert. Dabei sind zum einen die Zusammenhänge zwischen inhaltlich sehr ähnlichen Kompetenzkonstrukten, etwa der Lesekompetenz in verschiedenen Erhebungen, interessant. Zum anderen sind die Zusammenhänge zwischen unterschiedlichen Kompetenzbereichen bedeutsam, wie beispielsweise zwischen der Lesekompetenz und der mathematischen Kompetenz. Die Stärke dieser Zusammenhänge kann dann vergleichend interpretiert werden. Aus theoretischer Sicht wäre zu erwarten, dass die Zusammenhänge zwischen inhaltlich ähnlichen Kompetenzkonstrukten (Lesekompetenz in PISA und in Studien zu den Bildungsstandards) stärker ausfallen sollten als zwischen inhaltlich unterschiedlichen Kompetenzkonstrukten. Die Stärke der korrelativen Zusammenhänge zwischen verschiedenen Kompetenzbereichen kann also als Grundlage für die Interpretation von Korrelationen zwischen Messungen zum gleichen Kompetenzbereich in verschiedenen Studien dienen. Im Folgenden werden aktuelle empirische Ergebnisse zum Zusammenhang zwischen Tests zu unterschiedlichen Kompetenzbereichen sowie zwischen Tests zu denselben Kompetenzbereichen aus nationalen und internationalen Studien dargestellt.

Die latenten, also messfehlerbereinigten Korrelationen der Testwerte für die verschiedenen bei PISA erfassten Kompetenzbereiche liegen international im Durchschnitt zwischen $r = .77$ und $r = .89$ (PISA 2003, Leutner, Klieme, Meyer & Wirth, 2004, S. 167) beziehungsweise zwischen $r = .79$ und $r = .88$ (PISA 2006, vgl. OECD, 2009c, S. 215). Beispielsweise betragen die Korrelationen zwischen den Testwerten in Naturwissenschaften und Lesen in der deutschen Stichprobe von PISA 2000 $r = .87$, zwischen Naturwissenschaften und Mathematik $r = .83$, und zwischen Mathematik und Lesekompetenz $r = .84$ (vgl. Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001, S. 221).

Bereits in PISA 2000 wurden die in Deutschland durchgeführten internationalen Schulleistungsstudien durch nationale Testverfahren ergänzt, so beispielsweise in den Bereichen Mathematik und Naturwissenschaften. Die nationalen Tests waren stärker an den Lehrplänen für Schulen in Deutschland orientiert als die internationalen Tests. Zwischen den internationalen und den national ergänzten Mathematiktests ergaben sich dabei Korrelationen von $r = .91$ (PISA 2000, vgl. Klieme, Neubrand & Lüdtke, 2001, S. 157) beziehungsweise von $r = .92$ (PISA 2003, vgl. Blum et al., 2004, S. 63). Die Korrelation zwischen den internationalen und den national ergänzten Tests im Bereich Naturwissenschaften betrug in PISA 2000 $r = .84$ (vgl. Prenzel et al., 2001, S. 224). Es kann festgestellt werden, dass die Zusammenhänge sich zwischen den Kompetenzbereichen (Mathematik und Naturwissenschaften) durchaus unterscheiden.

Seit Einführung der länderübergreifenden Bildungsstandards in Deutschland stellt sich zudem die Frage, wie die auf dieser Basis entwickelten Testaufgaben mit den internationalen Tests zusammenhängen. Für den Bereich der Mathematik untersuchten Hartig und Frey (2012) diese Fragestellung anhand von Daten aus der gemeinsamen Erhebung von PISA und Tests zu den Bildungsstandards in Mathematik im Jahr 2006. Der latente Zusammenhang zwischen den Testwerten im Bildungsstandards-Mathematiktest und im PISA-Mathematiktest fiel mit einer Korrelation von $r = .95$ signifikant stärker aus als die Korrelation zwischen Bildungsstandards-Mathematik und PISA-Lesen ($r = .77$) sowie zwischen Bildungsstandards-Mathematik und PISA-Naturwissenschaften ($r = .87$). Trotzdem kommen die Autoren auf Basis ihrer Analysen zu dem Schluss, dass ein differenziertes Modell, welches den Bildungsstandards-Mathematiktest und den PISA-Mathematiktest als zwei korrelierte, jedoch distinkte Dimensionen ausweist, die Varianz in den vorliegenden Daten besser erklären kann als ein eindimensionales Modell. Sie begründen dieses Ergebnis mit einem unterschiedlichen Validitätsbereich der beiden Tests, letztendlich also der Zielsetzung der Studien, die sich in der Entwicklung von Testaufgaben widerspiegelt.

Auch für den Bereich der Lesekompetenz liegen erste Analysen zum Zusammenhang zwischen nationalen und internationalen Kompetenztests vor. Für den Vergleich zwischen der Lesekompetenzmessung auf Basis der Bildungsstandards in der Grundschule und im Rahmen von IGLU berichten Pietsch, Böhme, Robitzsch und Stubbe (2009) inhaltliche Übereinstimmungen zwischen den verwendeten Kompetenzstufenmodellen. Die empirischen Zusammenhänge zwischen den Testverfahren liegen auf latenter Ebene bei $r = .84$. Die Stärke dieser latenten Zusammenhänge ist vergleichbar mit jener zwischen den PISA-Tests zu den drei Kompetenzbereichen (vgl. etwa die oben genannten Ergebnisse aus PISA 2003). Die Autorinnen und Autoren kommen zu dem Schluss, „dass ein unmittelbarer, linearer Vergleich der Ergebnisse aus IGLU 2006 und den länderübergreifenden Bildungsstandards nicht sinnvoll ist“ (Pietsch et al., 2009, S. 426). Sie vermuten, dass die geringe Korrelation auf deutliche Unterschiede in der Aufgabenentwicklung und in den verwendeten Aufgabentypen zwischen den beiden Studien, insbesondere aber auf die Verwendung verschiedener Arten von Stimulustexten zurückzuführen ist.

Anhand der hier dargestellten Ergebnisse wird deutlich, dass Tests aus unterschiedlichen Studien, auch wenn sie auf das gleiche Kompetenzkonstrukt abzielen, in der Regel nicht vollständig das Gleiche messen. Insbesondere bei der gemeinsamen Modellierung von Daten aus internationalen Studien mit denen aus nationalen Erhebungen erscheint der sprachliche beziehungsweise nationale Ursprung der Testaufgaben selbst als eine der Ursachen für die jeweils spezifischen Varianzanteile (Artelt & Baumert, 2004; Jurecka, 2010).

Aufbauend auf den hier dargestellten Befunden beschäftigt sich auch der vorliegende Artikel mit der Frage des empirischen Zusammenhangs zwischen Testwerten aus einem nationalen und einem internationalen Kompetenztest. Im Folgenden werden die Fragestellung, die Datengrundlage und das methodische Vorgehen dargestellt.

4. Methode

4.1 Fragestellungen

Der vorliegende Artikel geht der Frage nach, inwieweit die Kompetenzverteilungen der Schülerinnen und Schüler in Deutschland, die auf der Grundlage des Lesetests in PISA 2009 und des Lesetests im Rahmen des bildungsstandardbasierten Ländervergleichs 2009 ermittelt wurden, empirisch vergleichbar sind.

Diese Fragestellung lässt sich wie folgt weiter differenzieren:

- a) Wie unterscheiden sich die Verteilungen von Schülerinnen und Schülern, die beide Tests bearbeitet haben, auf den jeweiligen Kompetenzstufen der Lesekompetenz?
- b) Welcher korrelative Zusammenhang besteht zwischen den mit beiden Testverfahren gemessenen Kompetenzwerten?

4.2 Datengrundlage

Die Datengrundlage für die folgenden Analysen bildet eine Stichprobe von 8061 Schülerinnen (49,3%) und Schülern der 9. Jahrgangsstufe aus 201 Schulen in Deutschland. Dabei wurden alle Schularten des allgemeinbildenden Schulsystems einbezogen. Aufgrund des Erhebungsdesigns wurden Schülerinnen und Schüler aus beruflichen Schulen sowie aus Förderschulen in dieser Stichprobe allerdings nicht einbezogen. Das Alter der Schülerinnen und Schüler betrug im Durchschnitt 15 Jahre (Minimum 13, Maximum 19). Diese Stichprobe verband den bildungsstandardbasierten Ländervergleich Sprachen 2009 mit der Erhebung von PISA 2009. Für beide Studien wurden jeweils eigenständige Stichproben gezogen, wobei die hier verwendete gemeinsame Stichprobe lediglich einen Teil der Gesamtstichproben zur bildungsstandardbasierten Erhebung und von PISA darstellt (vgl. Abb. 1). Die Schülerinnen und Schüler bearbeiteten am ersten Testtag die internationalen PISA-Aufgaben und an einem zweiten Testtag die Testaufgaben des bildungsstandardbasierten Ländervergleichs.

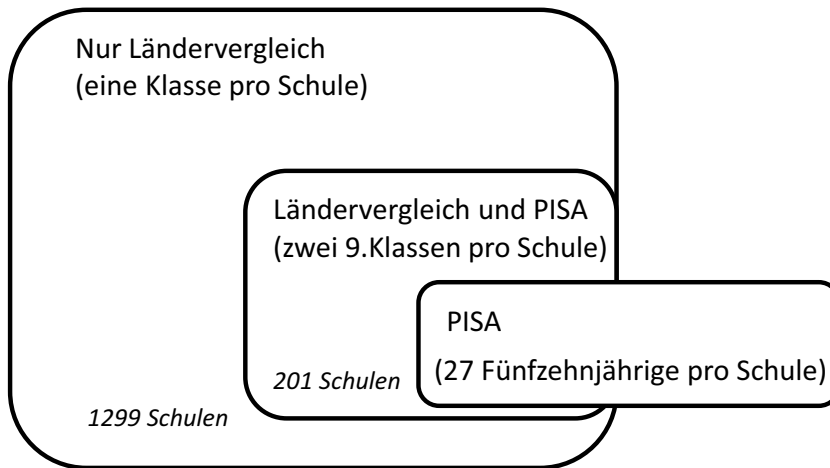


Abb. 1: Stichproben im Ländervergleich und der neunten Klassen in PISA 2009

4.3 Analytisches Vorgehen

Der Vergleich der empirischen Verteilungen der Schülerinnen und Schüler auf die Kompetenzstufen von PISA und den Bildungsstandards basiert auf den für beide Studien separat skalierten Testwerten. Pro Studie wurden für jede Person fünf *Plausible Values* (PVs) geschätzt (Köller et al., 2010; OECD, 2012). Anhand dieser PVs wurden für jede Studie separat die prozentualen Verteilungen der Stichprobe auf den Kompetenzstufen erzeugt und einander gegenübergestellt. Verwendet wurde hierfür die Software SPSS mit dem Zusatzmodul Macro für *Replicate Weights* (Balanced Repeated Replication, vgl. OECD, 2012, 2009d).

In diese Analyse gingen die Rohwerte von 101 Einzelaufgaben aus PISA sowie von 99 Einzelaufgaben zu den Bildungsstandards aus dem Kompetenzbereich Lesen ein. Zur Beurteilung der gemeinsamen dimensional Struktur der beiden Kompetenztests wurde die Passung von zwei Raschmodellen verglichen. Geschätzt wurden ein eindimensionales Raschmodell, bei dem alle Aufgaben auf einem gemeinsamen latenten Faktor luden, sowie ein zweidimensionales Raschmodell. Im zweidimensionalen Modell definierte die Zugehörigkeit der Items zu den Tests die Ladungsstruktur, d.h. alle Aufgaben des PISA-Lesetests luden ausschließlich auf einem Faktor und alle Aufgaben des Bildungsstandards-Lesetests luden ausschließlich auf einem zweiten Faktor (Adams, Wilson & Wang, 1997). Die latente Korrelation zwischen den beiden Faktoren wurde frei geschätzt.

In den Analysen wurden Schüलगewichte zur Berücksichtigung unterschiedlicher Ziehungswahrscheinlichkeiten verwendet. Der Schachtelung der Schülerinnen und

Schüler in Schulen wurde durch die Verwendung der in Mplus implementierten Pseudo-Maximum-Likelihood-Schätzung (PML) für komplexe Stichproben (Asparouhov & Muthén, 2005) Rechnung getragen. Hierzu wurde die Software Mplus 6.1 verwendet (Muthén & Muthén, 2009). Vergleichskriterien für die Beurteilung der Modellpassung waren der AIC und BIC (Akaike, 1974; Schwarz, 1978).

5. Ergebnisse

5.1 Verteilung auf die Kompetenzstufen

Um zu vergleichen, wie sich die Personen der gemeinsamen Stichprobe auf die Kompetenzstufen von PISA und den Bildungsstandards verteilen, wurden die Verteilungen jeweils getrennt für die Kompetenzstufen nach PISA und die Kompetenzstufen der Bildungsstandards gebildet. Jede Person wurde also zweimal zugeordnet: Einmal anhand ihres Testwertes in PISA auf die Kompetenzstufen von PISA 2009 und einmal anhand des Testwertes, den sie in den Aufgaben auf Basis der Bildungsstandards erzielt hat, auf die Kompetenzstufen des Bildungsstandardmodells. Anhand dieser Zuordnung kann keine Aussage darüber getroffen werden, ob sich eine Person jeweils auf den gleichen Stufen in beiden Modellen befindet. Die entsprechenden Verteilungen auf den Kompetenzstufen sind vergleichend in Tabelle 4 gegenübergestellt.¹

Wie anhand der prozentualen Häufigkeiten ersichtlich wird, unterscheiden sich die Verteilungen der Personen auf die Kompetenzstufen nicht bedeutsam zwischen PISA und den Bildungsstandards. Bedingt durch die Unterschiede im Differenzierungsgrad der Kompetenzmodelle ist die Stufe VI als oberste Kompetenzstufe nur im Modell von PISA belegt (0.3 % der Stichprobe). Dem unteren Kompetenzbereich (Stufe I) werden im Modell der Bildungsstandards 16.3% der Stichprobe zugeordnet. Dieser Wert entspricht in etwa dem Anteil von Personen, die im Modell von PISA zusammen den Stufen Ib und Ia sowie dem Bereich unter Kompetenzstufe Ib zugeordnet werden (16.1%). PISA differenziert in diesem Bereich zusätzlich noch zwischen Personen, die Kompetenzstufe Ib nicht erreichen (0.5%), und jenen auf Kompetenzstufe Ib (3.4%).

1 Es ist zu beachten, dass die hier dargestellte Verteilung der Neuntklässlerinnen und Neuntklässler in Deutschland auf die Kompetenzstufen des Bildungsstandards-Modells weder mit den für die Normierungsstudie zu den Bildungsstandards berichteten Verteilungen vergleichbar ist, noch mit den in Köller et al. (2010) berichteten Verteilungen zum Ländervergleich 2009. Die dort wiedergegebenen Verteilungen beziehen sich nur auf jene Neuntklässlerinnen und Neuntklässler in Deutschland, die den Mittleren Schulabschluss anstreben, während die im vorliegenden Artikel berichtete Verteilung auch jene Neuntklässlerinnen und Neuntklässler einschließt, die den Hauptschulabschluss anstreben.

Kompetenzstufen	PISA		Bildungsstandards für den MSA	
	%	SE	%	SE
VI	0.3	0.1	5.4	0.5
V	5.2	0.4		
IV	21.7	1.0	19.3	0.9
III	31.4	0.8	31.9	0.8
II	25.3	1.0	27.0	1.0
la	12.2	0.7	16.4	1.0
lb	3.4	0.4		
unter lb	0.5	0.1		

Tab. 4: Verteilung der Neuntklässlerinnen und Neuntklässler auf die Kompetenzstufen von PISA 2009 und des Kompetenzstufenmodells zu den Bildungsstandards für den MSA im Fach Deutsch, Bereich Lesen (prozentuale Häufigkeiten und Standardfehler [SE])

5.2 Zusammenhang zwischen den Lesetests

Um zu prüfen, welcher empirische Zusammenhang zwischen dem Lesetest von PISA und dem Lesetest zu den Bildungsstandards besteht, wurden auf Basis der Ergebnisse

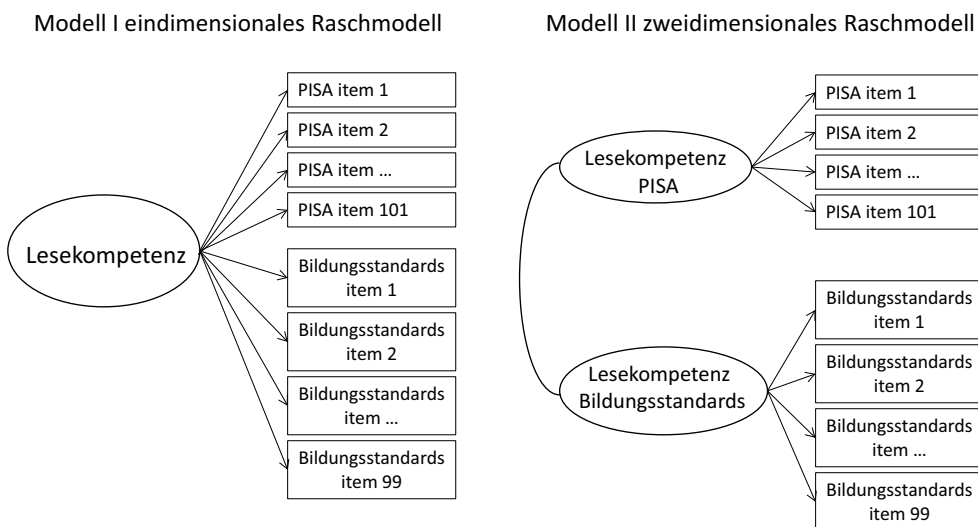


Abb. 2: Eindimensionale und zweidimensionale Raschmodellierung des PISA-Lesetests und des auf den Bildungsstandards basierenden Lesetests (schematische Darstellung)

der Schülerinnen und Schüler in den einzelnen Testaufgaben zwei Modelle spezifiziert und verglichen (vgl. Abb. 2).

Das erste Modell geht von einem gemeinsamen Faktor der Lesekompetenz aus, auf dem alle Testaufgaben aus PISA und dem Ländervergleich gleichermaßen laden. Das zweite Modell spezifiziert zwei korrelierte latente Faktoren der Lesekompetenz – einen für PISA und einen für die Bildungsstandards-Aufgaben –, auf denen jeweils nur die Aufgaben der zugehörigen Studie laden. Beim Vergleich der beiden Modelle auf Basis des AIC und BIC erweist sich die Passung des zweidimensionalen Modells (AIC = 473154; BIC = 474623) für die Beschreibung der in den Daten vorliegenden Zusammenhänge der Passung des eindimensionalen Modells (AIC = 474706; BIC = 476161) als überlegen. Die latente Korrelation von $r = .85$ zwischen den Faktoren im zweidimensionalen Modell kann dahingehend interpretiert werden, dass es sich um zwei korrelierte Konstrukte handelt, die jeweils einen eigenen Beitrag zur Erklärung der in den Schülerantworten auftretenden Varianz liefern.

6. Diskussion

Der vorliegende Artikel untersuchte die Frage, inwieweit die auf den Testergebnissen basierenden Kompetenzverteilungen der Schülerinnen und Schüler in Deutschland zwischen dem PISA-Lesetest und dem Lesetest auf Basis der Bildungsstandards für den Mittleren Schulabschluss empirisch vergleichbar sind. Dazu wurden die theoretischen Definitionen der Lesekompetenz von PISA und der Bildungsstandards, der Prozess der Aufgabenentwicklung sowie die Kompetenzmodelle beider Studien vergleichend dargestellt. Darüber hinaus wurden auf Basis einer gemeinsamen Stichprobe die empirischen Zusammenhänge zwischen den Ergebnissen der Leistungstests bestimmt und die Häufigkeitsverteilungen auf den Kompetenzstufen verglichen.

Die Analyse des empirischen Zusammenhangs zwischen den Kompetenzkonstrukten in PISA und der Studie auf Basis der Bildungsstandards hat gezeigt, dass ein zweidimensionales Modell, das separate korrelierte Kompetenzdimensionen annimmt, die in den Daten vorhandene Struktur besser abbildet als ein eindimensionales Modell. Der empirische Zusammenhang zwischen den Testwerten aus der zweidimensionalen Skalierung beträgt $r = .85$. Die latente Korrelation zwischen beiden Tests liegt damit in einem ähnlichen Bereich wie der Zusammenhang zwischen dem Lesetest auf Basis der Bildungsstandards im Fach Deutsch für den Primarbereich und dem IGLU-Lesetest. Zudem ist diese Korrelation vergleichbar mit Zusammenhängen zwischen unterschiedlichen Kompetenzbereichen in PISA, also beispielsweise zwischen Lesen und Mathematik. Die latente Korrelation ist sogar geringer als der empirische Zusammenhang zwischen dem nationalen Test zu den Bildungsstandards-Mathematik und dem internationalen PISA-Mathematiktest.

Dieses Ergebnis wird im Folgenden vor dem Hintergrund der zugrunde liegenden Kompetenzkonzepte, der Zielsetzungen der Studien, der jeweiligen Prozesse der Aufgabenentwicklung und der Gestaltung der Kompetenzmodelle diskutiert.

Kompetenzkonzepte und Zielsetzungen der Studien

Das Konstrukt der Lesekompetenz kann konzeptuell als mehrdimensional angesehen werden (Artelt & Schlagmüller, 2004). PISA berücksichtigt verschiedene Aspekte der Lesekompetenz dadurch, dass diese als fünf getrennte Teildimensionen gemessen und berichtet werden. Die Ergebnisse zeigen, dass die Mittelwertsunterschiede zwischen Staaten in den verschiedenen Aspekten der Lesekompetenz beträchtlich sind (Naumann et al., 2010). Dies deutet darauf hin, dass die unterschiedlichen Teildimensionen der Lesekompetenz in unterschiedlichem Maße zur Gesamtkompetenz beitragen können.

Auch im theoretischen Konzept der Bildungsstandards wird der Mehrdimensionalität des Lesekompetenzkonstrukts Rechnung getragen, indem Anforderungsniveaus und Textsorten differenziert werden (vgl. KMK, 2004a). Diese Aspekte werden zwar bei der Testkonstruktion berücksichtigt, es erfolgt jedoch keine nach Teildimensionen getrennte Skalierung oder Berichtlegung. Analysen zur Dimensionalität der Lesekompetenz in den Tests zu den Bildungsstandards stehen noch aus. Vor diesem Hintergrund kann der vorliegende Artikel beim Vergleich beider Studien nur den über alle Items hinweg skalierten Gesamtestwert berücksichtigen. Es ist anzunehmen, dass die geringe Korrelation zwischen den beiden Lesetests zumindest teilweise durch die Mehrdimensionalität der Kompetenzkonstrukte erklärt werden kann. In weiterführenden Analysen sollte daher die Dimensionalität des Lesetests der Bildungsstandards empirisch überprüft werden.

Testentwicklung und Testaufgaben

Die Testentwicklung auf Grundlage der Bildungsstandards erfolgte mit nationaler Ausrichtung und dem Ziel, Testaufgaben für die Verwendung im Ländervergleich innerhalb von Deutschland bereitzustellen. Die PISA-Aufgabenentwicklung zielt hingegen darauf ab, eine in allen internationalen Teilnehmerstaaten vergleichbare Messung von Lesekompetenz zu ermöglichen. Das verwendete Aufgabenmaterial ist entsprechend breit, entstammt unterschiedlichen Kulturkreisen und ist nicht explizit an Curricula oder Bildungsstandards orientiert.

Dementsprechend finden sich zwischen den Studien auch Unterschiede in den Testaufgaben. Beispielsweise erfordern die Aufgaben zu den Bildungsstandards Kenntnisse über literarische Textgattungen, die in Deutschland im Unterricht der Sekundarstufe I behandelt werden. Die Aufgaben des PISA-Tests beziehen sich hingegen stärker auf den Umgang mit Alltagstexten, wie sie weltweit in jeder Kultur vorkommen könnten. Ein wichtiger formaler Unterschied besteht auch in der Länge der Stimulustexte: Während die Texte in PISA in der Regel eine Seite nicht überschreiten, wurden im bildungsstandardbasierten Ländervergleich überwiegend Texte mit einer Länge von mehr als einer Seite eingesetzt. Zudem verwendet PISA einen höheren Anteil an Aufgaben mit offenem Antwortformat.

Bei der gemeinsamen Skalierung aller Testaufgaben, wie sie im vorliegenden Artikel vorgenommen wurde, spielen diese Aufgabeneigenschaften eine wesentliche Rolle. Da-

her könnte die bessere Passung des zweidimensionalen Modells auf die Daten durchaus auch auf die unterschiedliche Genese der jeweils verwendeten Testaufgaben zurückzuführen sein.

Verteilung auf den Kompetenzstufen

Die Verteilung auf den Kompetenzstufen unterscheidet sich zwischen PISA und den Bildungsstandards nicht wesentlich. Die prozentualen Anteile an Personen auf den jeweiligen Stufen sind vergleichbar. Erklärt werden kann dieses Ergebnis durch das methodische Vorgehen bei der Bestimmung der Cut-off-Werte für die Kompetenzstufen im Modell der Bildungsstandards. Das Standard-Setting orientierte sich hier explizit an den Stufenbeschreibungen des PISA-Modells und der Stichprobenverteilung der Neuntklässler in Deutschland, wie sie in PISA 2000 ermittelt wurde. Daher war zu erwarten, dass sich die Verteilungen auf den Kompetenzstufen zu den Bildungsstandards, die anhand von Daten der Normierungsstudie festgelegt wurden, nicht wesentlich von der Verteilung für PISA 2009 unterscheiden.

Unterschiede in der Verteilung entstehen lediglich dadurch, dass das Kompetenzstufenmodell von PISA 2009 stärker differenziert als das Modell von PISA 2000 und den Bildungsstandards. Werden die zusätzlich differenzierten Stufen bei PISA allerdings zusammengefasst (die Stufen Ib und Ia sowie die Stufen V und VI) und den jeweils korrespondierenden Stufen der Bildungsstandards gegenübergestellt, zeigen sich für keine der Stufen signifikante Unterschiede in den prozentualen Häufigkeiten zwischen den beiden Studien. Wie die relativ geringe latente Korrelation im zweidimensionalen Modell zeigt, bedeutet dies jedoch nicht zwangsläufig, dass sich dieselben Personen in beiden Studien auf vergleichbaren Kompetenzstufen befinden (vgl. auch Pietsch et al., 2009).

Fazit und Ausblick

Insgesamt lässt sich auf Basis der hier dargestellten Ergebnisse für den Vergleich der Testergebnisse von PISA und zum bildungsstandardbasierten Test feststellen, dass a) die empirischen Verteilungen der Schülerinnen und Schüler in Deutschland auf den jeweiligen Kompetenzstufen prozentual vergleichbar sind, jedoch b) die in beiden Studien erfassten Lesekompetenzen empirisch als korrelierte, gleichzeitig jedoch auch distinkte Konstrukte zu betrachten sind.

Anhand eines erzielten Kompetenzwerts oder einer erreichten Kompetenzstufe in PISA kann also nicht direkt darauf geschlossen werden, welchen Kompetenzwert oder welche Kompetenzstufe dieselbe Person in der Studie auf Basis der Bildungsstandards erreicht hätte, und umgekehrt.

Die wesentlichen Unterschiede in den Zielsetzungen der Studien, in den Konstruktdefinitionen sowie in der Ausgestaltung des Testmaterials bieten mögliche Erklärungen für die vorliegenden Ergebnisse. Diese Ergebnisse reihen sich ein in die bisheri-

gen Ergebnisse zum Vergleich national entwickelter Testaufgaben mit Kompetenztests aus internationalen Large Scale Assessments.

Ausgehend von diesen Befunden und Überlegungen lassen sich mehrere Fragestellungen für weiterführende Analysen identifizieren. Zum einen sollte – wie bereits erwähnt – untersucht werden, ob die Aufgaben auf Basis der Bildungsstandards eine mehrdimensionale Struktur aufweisen, mit der sich wie bei PISA Teildimensionen der Lesekompetenz aufzeigen lassen. Anhand dieser Informationen ließen sich auch komplexere Modelle der Lesekompetenz mit den Daten beider Studien analysieren. Zur weiteren Erklärung der Konstruktdimensionalität empfiehlt es sich, eine Einschätzung der verwendeten Testaufgaben durch einschlägige Experten vornehmen zu lassen. Wie bereits dargestellt wurde, ist davon auszugehen, dass die Testaufgaben aufgrund ihrer unterschiedlichen Entstehungskontexte zu einem Teil auf unterschiedliche Aspekte der Lesekompetenz abzielen. Darüber hinaus könnten anhand von Experteneinschätzungen die jeweils korrespondierenden Kompetenzstufen beider Kompetenzstufenmodelle inhaltlich verglichen werden. Dies könnte Aufschluss darüber geben, warum zwar die Häufigkeitsverteilung auf den Kompetenzstufen weitgehend vergleichbar ist, die erfassten Kompetenzen jedoch nicht identisch sind.

Literatur

- Adams, R., Wilson, M., & Wang, W. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21(1), 1-23.
- Adams, R., & Wu, M. (Hrsg.) (2002). *PISA 2000 Technical Report*. Paris: OECD.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Artelt, C., & Baumert, J. (2004). Zur Vergleichbarkeit von Schülerleistungen bei Leseaufgaben unterschiedlichen sprachlichen Ursprungs. *Zeitschrift für Pädagogische Psychologie*, 18(4), 171-185.
- Artelt, C., Naumann, J., & Schneider, W. (2010). Lesemotivation und Lernstrategien. 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. 73-112). Münster: Waxmann.
- Artelt, C., & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen – Dimensionsanalysen und Ländervergleiche. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz – Vertiefende Analysen im Rahmen von PISA 2000* (S. 169-196). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Artelt, C., Stanat, P., Schneider, W., & Schiefele, U. (2001). Lesekompetenz. Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000* (S. 69-131). Opladen: Leske + Budrich.
- Asparouhov, T., & Muthén, B. O. (2005). *Multivariate Statistical Modeling with Survey Data* (Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference). http://www.fcsm.gov/05papers/Asparouhov_Muthen_IIA.pdf [26.11.2012].
- Autorengruppe Bildungsberichterstattung (2012). *Bildung in Deutschland 2012. Ein indikatoren-gestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf*. Bielefeld: wbv.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K. J., & Weiß, M. (Hrsg.) (2002). *PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.

- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.) (2000). *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske + Budrich.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F., & Carstensen, C. H. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47-92). Münster: Waxmann.
- Böhme, K., Leucht, M., Schipolowski, S., Porsch, R., Knigge, M., & Köller, O. (2010). Anlage und Durchführung des Ländervergleichs. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 65-85). Münster: Waxmann.
- Böhme, K., Neumann, D., & Schipolowski, S. (2010). Beschreibung der im Ländervergleich im Fach Deutsch untersuchten Kompetenzen. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 19-24). Münster: Waxmann.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K., & Valtin, R. (Hrsg.) (2007). *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bremerich-Vos, A., Behrens, U., Böhme, K., Krelle, M., Neumann, D., Robitzsch, A., Schipolowski, S., & Köller, O. (2010). Kompetenzstufenmodelle für das Fach Deutsch. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 37-49). Münster: Waxmann.
- Granzer, D. (2009). Von Bildungsstandards zu ihrer Überprüfung: Grundlagen der Item- und Testentwicklung. In D. Granzer, O. Köller & A. Bremerich-Vos (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 21-30). Weinheim/Basel: Beltz.
- Groeben, N., & Hurrelmann, B. (Hrsg.) (2004). *Lesesozialisation in der Mediengesellschaft: Ein Forschungsüberblick*. Weinheim: Juventa.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and Motivation in Reading. In M. L. Kamil, P. Mosenthal, P. D. Pearson & R. Barr (Hrsg.), *Handbook of Reading Research*. Mahwah: Lawrence Erlbaum Associates.
- Hartig, J., & Frey, A. (2012). Validität des Tests zur Überprüfung des Erreichens der Bildungsstandards in Mathematik. Zusammenhänge mit den bei PISA gemessenen Kompetenzen und Varianz zwischen Schulen und Schulformen. *Diagnostica*, 58, 3-14.
- Hartig, J., & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17-36). Bonn/Berlin: Bundesministerium für Bildung und Forschung.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127-143). Berlin: Springer.
- Jude, N., & Klieme, E. (2010). Das Programme for International Student Assessment (PISA). 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. 11-20). Münster: Waxmann.
- Jurecka, A. (2010). *Zum Zusammenhang von Differentiellen Item Funktionen und Testkultur*. Frankfurt a.M.: Universitätsbibliothek Johann Christian Senckenberg.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (Hrsg.) (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E., Eichler, W., Helmke, A., Lehmann, R. H., Nold, G., Rolff, H.-G., Schröder, K., Thomé, G., & Willenberg, H. (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. DESI-Ergebnisse Band 2*. Weinheim/Basel: Beltz.
- Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen*

- von Schülerinnen und Schülern im internationalen Vergleich (S. 139-190). Opladen: Leske + Budrich.
- KMK (1997). *Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland. Konstanzer Beschluss*. Bonn/Berlin: KMK.
- KMK (2004a). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss (Jahrgangsstufe 10)*. München: Luchterhand.
- KMK (2004b). *Bildungsstandards der Kultusministerkonferenz: Erläuterungen zur Konzeption und Entwicklung*. München: Luchterhand.
- KMK (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Wolters Kluwer.
- Köller, O., Knigge, M., & Tesch, B. (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147-175). Münster: Waxmann.
- Mildner, D., Hochweber, J., & Frey, A. (2013). Vergleichende Analysen der Kompetenzen von Fünfzehnjährigen und Neuntklässlern in den deutschen PISA-Erhebungen 2003 bis 2009. *Zeitschrift für Pädagogik*, 59. Beiheft, 151-157.
- Muthén, L. K., & Muthén, B. O. (2009). *Mplus User's Guide* (6. Aufl.). Los Angeles: Muthén & Muthén.
- 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, P. Stanat & W. Schneider (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- OECD (2009a). *PISA 2009 Assessment Framework – Key Competencies in Reading, Mathematics and Science*. Paris: OECD.
- OECD (2009b). *PISA Take the Test: Sample Questions from OECD's PISA Assessments*. Paris: OECD.
- OECD (2009c). *PISA 2006 Technical Report*. Paris: OECD.
- OECD (2009d). *PISA Data Analysis Manual: SPSS® Second edition*. Paris: OECD.
- OECD (2010). *PISA 2009 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften* (Band I). Berlin: wbv.
- OECD (2012). *PISA 2009 Technical Report*. Paris: OECD.
- Pant, H. A., Tiffin-Richards, S. P., & Köller, O. (2010). Standard-Setting für Kompetenztests im Large-Scale-Assessment. *Zeitschrift für Pädagogik*, 56. Beiheft, 175-188.
- Pietsch, M., Böhme, K., Robitzsch, A., & Stubbe, T. (2009). Das Stufenmodell zur Lesekompetenz der länderübergreifenden Bildungsstandards im Vergleich zu IGLU 2006. In A. Bremerich-Vos, D. Granzer & O. Köller (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 393-416). Weinheim/Basel: Beltz.
- Pöhlmann, C., Neumann, D., Tesch, B., & Köller, O. (2010). Die Bildungsstandards im allgemeinbildenden Schulsystem. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 13-18). Münster: Waxmann.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (Hrsg.) (2008). *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R. H., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (Hrsg.) (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* Münster: Waxmann.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P., & Klopp, A. (2001). Naturwissenschaftliche Grundbildung. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von*

Schülerinnen und Schülern im internationalen Vergleich (S. 191-248). Opladen: Leske + Budrich.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
Stanat, P., Pant, H. A., Böhme, K., & Richter, D. (Hrsg.) (2012). *Ergebnisse des IQB-Ländervergleichs 2011. Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik*. Münster: Waxmann.

Anschrift der Autor(inn)en

Dr. Nina Jude, Deutsches Institut für Internationale Pädagogische Forschung (DIPF),
Schloßstr. 29, 60486 Frankfurt am Main, Deutschland
E-Mail: jude@dipf.de

Prof. Dr. Johannes Hartig, Deutsches Institut für Internationale Pädagogische Forschung
(DIPF), Schloßstr. 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hartig@dipf.de

Dipl.-Psych. Stefan Schipolowski, Humboldt-Universität zu Berlin, Institut zur Qualitäts-
entwicklung im Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: Stefan.Schipolowski@IQB.hu-berlin.de

Dr. Katrin Böhme, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im
Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: Katrin.Boehme@IQB.hu-berlin.de

Prof. Dr. Petra Stanat, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im
Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: iqboffice@IQB.hu-berlin.de