

Bischof, Linda Marie; Hochweber, Jan; Hartig, Johannes; Klieme, Eckhard  
**Schulentwicklung im Verlauf eines Jahrzehnts. Erste Ergebnisse des PISA-Schulpanels**

Jude, Nina [Hrsg.]; Klieme, Eckhard [Hrsg.]: PISA 2009 - Impulse für die Schul- und Unterrichtsforschung. Weinheim u.a. : Beltz 2013, S. 172-199. - (Zeitschrift für Pädagogik, Beiheft; 59)

urn:nbn:de:0111-opus-78264



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](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://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](http://www.beltz.de) · [www.juventa.de](http://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

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

Das „Programme for International Student Assessment“ (PISA) ist ein internationales Projekt, in dem alle drei Jahre die Kompetenzen von fünfzehnjährigen Schülerinnen und Schülern erfasst werden. Die Erfassung der Kompetenzen erfolgt in jedem Durchgang mit anderem Schwerpunkt; je drei Erhebungen mit Schwerpunkten in Lesen, Mathematik und Naturwissenschaften bilden einen Zyklus. Im Jahr 2009 stand mit Beginn des zweiten Zyklus erstmals seit PISA 2000 wieder die Messung der Lesekompetenz (*Reading Literacy*) der Schülerinnen und Schüler im Mittelpunkt. Während die PISA-Ergebnisse Aufschluss über Veränderungen in den Kompetenzen sowie bestimmten weiteren Merkmalen (z.B. Lernmotivation) fünfzehnjähriger Schülerinnen und Schüler in den beteiligten Staaten geben können, im Sinne von Unterschieden zu vorherigen Kohorten, informieren sie weder über die Entwicklung einzelner Schülerinnen und Schüler noch über die Entwicklung von Schulen. Dies liegt im Design der PISA-Studie begründet. Schulen werden zu jedem Messzeitpunkt unabhängig voneinander zufällig gezogen und gelangen entsprechend nur zufällig mehrmals in die PISA-Stichprobe (Klieme & Steiner, 2008). PISA erlaubt also keine echten längsschnittlichen Aussagen über die Entwicklung von Schülerinnen und Schülern bzw. Schulen, sondern nur Trendergebnisse auf der Ebene der Staaten. So können die Frage nach den Bedingungsfaktoren für individuelle Leistungsentwicklung und die Frage nach Bedingungsfaktoren für die Entwicklung des Leistungsniveaus auf Schulebene nicht befriedigend beantwortet werden. Hierzu sind Erweiterungen des PISA-Designs erforderlich, die bislang nur innerhalb Deutschlands realisiert werden konnten (vgl. Klieme, Jude, Baumert & Prenzel, 2010, S. 295).

Eine Möglichkeit, über den Zeitraum seit der ersten PISA-Erhebung hinweg längsschnittliche Informationen zu Entwicklungen auf Schulebene zu erhalten, ergab sich mit dem PISA-Schulpanel. Ergänzend zur internationalen Stichprobe für PISA 2009 wurde in Deutschland national eine Zusatzstichprobe von 59 Gymnasien gezogen, die auch an PISA 2000 teilgenommen hatten.<sup>1</sup> Die Untersuchung beschränkte sich auf Gymnasien, da das Gymnasium eine Schulart ist, die in jedem Bundesland in Deutschland existiert, und weil Umwandlungen der Schulart bei Gymnasien in diesem Zeitraum so gut wie nie vorkamen. Die Schulpanel-Stichprobe ermöglicht die längsschnittliche Analyse der Entwicklung der teilnehmenden Gymnasien über einen Zeitraum von neun Jahren. Im vorliegenden Beitrag wurde untersucht, in welchem Zusammenhang ausgewählte

1 Da wir uns hier nur mit deutschen Daten beschäftigen, unterscheiden wir nicht zwischen den Schulen, die bei PISA 2000 in den internationalen Vergleich gingen (Baumert et al., 2001), und der erweiterten Stichprobe PISA-E, die Grundlage des Vergleichs zwischen den Bundesländern war (Baumert et al., 2003).

Schulmerkmale und deren Veränderung mit der Veränderung kognitiver und nicht-kognitiver Merkmale von Schülerinnen und Schülern dieser Schulen stehen.

Im Folgenden gehen wir nach der Darstellung des Hintergrunds unseres Beitrags (Abschnitt 1) auf die Zielsetzung und das Design der Studie (Abschnitt 2) sowie die konkreten Fragestellungen dieses Artikels ein (Abschnitt 3) und beschreiben anschließend unser methodisches Vorgehen (Abschnitt 4). In den Abschnitten 5 und 6 werden die Ergebnisse dargestellt und diskutiert.

## 1. Theoretischer Hintergrund: Von der Schuleffektivitäts- zur Schulentwicklungsforschung

Die vorliegende Untersuchung befasst sich mit der Entwicklung der Schule als Organisation mit dem Ziel, Prädiktoren der Entwicklung von Schulen zu identifizieren. Diese Fragestellung ist der herkömmlichen Schuleffektivitäts- bzw. Schulqualitätsforschung verwandt, beinhaltet aber einen fundamentalen Wechsel der Perspektive. Daher unterscheiden wir im Folgenden explizit zwischen Aspekten der Schuleffektivität (Abschnitt 1.1) und der Schulqualität (Abschnitt 1.2) und schließen dann den Bogen zur Schulentwicklungsforschung (Abschnitt 1.3), um anschließend die hier untersuchten Prädiktoren der Entwicklung von Schulen näher zu erläutern (Abschnitt 1.4).

### 1.1 Schuleffektivität

Die Schuleffektivität wird üblicherweise daran gemessen, welche Ergebnisse eine Schule erreicht („Output“), wobei meist in Rechnung gestellt wird, unter welchen Bedingungen eine Schule arbeitet („Kontext“ und „Input“). Die Auswahl der Kriterien zur Messung effektiver Schulen, also der untersuchten Outputmerkmale, wird immer wieder diskutiert (de Maeyer, van den Bergh, Rymenans, van Petegem & Rijlaarsdam, 2010; Sammons, Hillman & Mortimore, 1995; Scheerens & Bosker, 1997; Seidel, 2008; Teddlie & Reynolds, 2000). Die Auswahl eines einzelnen Effektivitätskriteriums ergibt kein generalisierbares Bild davon, welche Merkmale der Schule eine effektivitätssteigernde Wirkung haben. Dies bedarf einer breiteren Auswahl an Kriterien.

Kriterien effektiver Schulen lassen sich unterschiedlich einteilen. Seidel (2008) beschreibt vier Kategorien von Effektivitätskriterien: kognitive, motivational-affektive, metakognitive und verhaltensbezogene Kriterien. Eine einfachere Einteilung ist die in kognitive und non-kognitive Effektivitätskriterien (Mortimore, Sammons & Thomas, 1994; Opdenakker & van Damme, 2000; van Damme et al., 2006). Zumeist wird die Effektivität von Schulen ausschließlich über kognitive, d.h. über leistungs- bzw. kompetenzbezogene Kriterien gemessen (Scheerens, 2008; Scheerens & Bosker, 1997; Seidel, 2008; van Damme et al., 2006), wobei sehr unterschiedliche Leistungsmaße als Kriterien herangezogen werden, z.B. die Leistung in bestimmten Fächern oder Kompetenzen (vgl. Seidel, 2008). Im Kontext des PISA-Schulpanels stützen wir uns als kognitives

Kriterium auf die Lesekompetenz der Schülerinnen und Schüler, da diese in 2000 und 2009 als Schwerpunkt erhoben wurde.

Die neuere Schulforschung berücksichtigt zunehmend auch non-kognitive, affektive Kriterien (vgl. Diedrich, 2008; Gerecht, 2010; Opdenakker & van Damme, 2000). Kognitive und non-kognitive Kriterien können durchaus zu unterschiedlichen Schlussfolgerungen führen. So wird bei van Damme et al. (2006) deutlich gemacht, dass Schulen einen Einfluss auf non-kognitive Kriterien haben; und dass Schulen, die hinsichtlich eines kognitiven Kriteriums als effektiv eingestuft werden, nicht notwendigerweise auch einen positiven Effekt auf non-kognitive Kriterien wie beispielsweise das Wohlbefinden haben müssen. Im PISA-Schulpanel verwenden wir als ein non-kognitives Kriterium die von den Schülerinnen und Schülern wahrgenommene Beziehungsqualität zu ihren Lehrerinnen und Lehrern (kurz: Beziehungsqualität). Diese kann in Anlehnung an van Damme, de Fraine, van Landeghem, Opdenakker und Onghena (2002) als ein Aspekt des Effektivitätskriteriums Schülerwohlbefinden aufgefasst werden. Diese Einordnung der Beziehungsqualität als Indikator für Schülerwohlbefinden kann kritisch betrachtet werden, denn die Beziehungsqualität wurde, beispielsweise in PISA 2009, als Indikator für das Schulklima herangezogen (vgl. Hertel, Hochweber, Steinert & Klieme, 2010; siehe auch Eder, 2002). Das Klima einer Schule ist allerdings streng genommen kein Effektivitätskriterium wie das Leistungsniveau der Schülerinnen und Schüler, sondern ein Prozessmerkmal der Schule und somit ein Merkmal der Schulqualität. Verschiedene Forscher fanden jedoch einen positiven Effekt eines guten Schulklimas auf Schülermerkmale, wie beispielsweise die Lernmotivation der Schülerinnen und Schüler (Brophy, 2000; Fraser, 1994; Götz, Frenzel & Pekrun, 2008). Zudem ist das Schulklima „sensibel für Veränderungen im Schulsystem bzw. Entwicklungsmaßnahmen einzelner Schulen“ (Eder, 2006, S. 626). Dies bestärkt uns darin, neben Veränderungen im Leseleistungsniveau Veränderungen in der Beziehungsqualität als Kriterium für die Entwicklung der Schulen zu nutzen.

## 1.2 Schulqualität

Schulqualität wird in der Forschungsliteratur manchmal gleichgesetzt mit Schuleffektivität. Präziser ist jedoch eine klare Trennung: Als Aspekte der Schulqualität bezeichnen wir jene Merkmale der pädagogischen Arbeit und der Schulorganisation, also „Prozessmerkmale“, die Schuleffektivität (direkt oder indirekt) im Sinne eines hohen Outputs ermöglichen. Die Schulforschung hat vielfältige Faktoren identifiziert, die in diesem Sinne „gute Schulen“ ausmachen. Eine Liste solcher Faktoren wurde auf der Grundlage praktischer Erfahrungen erstmals von Edmonds (1979) vorgeschlagen: Qualität des Curriculums – damals gleichgesetzt mit einer Betonung von „Basic Skills“ –, hohe Leistungserwartungen, Schulklima, Führungsqualität der Schulleitung sowie die Nutzung von Leistungsmessungen und Evaluation. Neuere Übersichten, die sich auf Meta-Analysen zahlreicher empirischer Studien stützen, bestätigen diese Faktoren und ergänzen sie um Konsens und Kooperation der Lehrerinnen und Lehrer sowie die Einbindung



der Eltern (so etwa Creemers & Reezigt, 1996; Scheerens, 2008). Scheerens (2008) führt zudem effektivitätssteigernde Faktoren an, die sich auf den Unterricht beziehen: Klassenklima, Klassenmanagement/effektive Lernzeit, strukturierte Instruktionen, Differenzierung und adaptive Instruktionen sowie Feedback. Creemers und Reezigt (1996) hingegen benennen ergänzende Faktoren der Schulebene, die guten Unterricht ermöglichen: Fokus auf Grundfertigkeiten/Lernzeit sowie die Koordination von Curriculum und Instruktion über Klassen und Jahrgangsstufen hinweg.

Ausgehend vor allem von den Arbeiten Fends (z.B. 1998) liegen inzwischen auch im deutschen Sprachraum verschiedene Modelle vor, die diesen Ansatz weiter ausdifferenzieren. Ditton (2000a, 2000b, 2007) unterscheidet Individual-, Unterrichts-, Schul- und Kontextebene und bezeichnet sein Modell als „4-plus-Modell“, da die Kontextebene noch genauer unterteilt werden kann. Den auf den verschiedenen Ebenen verorteten Faktoren wurde in der Schulqualitätsforschung eine unterschiedliche Wirksamkeit insbesondere bezüglich der Leistungsergebnisse der Schülerinnen und Schüler zugesprochen. Ditton (2000a, 2000b) verweist hier auf die „These der primären Bedeutung proximaler Faktoren“ (d.h. den Schülerinnen und Schülern näher stehender Faktoren), nach welcher die direkte Lernumgebung der Schülerinnen und Schüler den größten Effekt auf die Schülerleistung hat. Diese These wird beispielsweise von Klieme, Steinert und Hochweber (2010) gestützt, die nachweisen, dass Kooperation und geteilte Ziele in Deutsch-Fachkollegien zwar den Unterricht vielfältiger machen und Disziplinprobleme verringern, aber nicht auf die individuelle Leistungs- und Motivationsentwicklung der Schülerinnen und Schüler durchschlagen.

Gemeinhin werden Faktoren, die in Schulqualitätsstudien als wirksam identifiziert wurden, auch als bedeutsam für die Schulentwicklung angesehen. Im vorliegenden Beitrag wollen wir unser Augenmerk auf proximale und distale Faktoren richten, die in den letzten Jahren auch bildungspolitisch eine hohe Aufmerksamkeit bekommen haben und für die im Schulpanel Veränderungsmaße vorliegen, um zu prüfen, welche Rolle sie für die Entwicklung von Schulen spielen. Konkret werden in unserer Untersuchung individuelle Förderung (als proximaler Faktor für Lernprozesse) sowie ganztägige Schulorganisation und Evaluationsaktivitäten (als distale Faktoren) als Prädiktoren einbezogen.

### 1.3 Schulentwicklung

Anders als die Schuleffektivitäts- und Schulqualitätsforschung, für welche die Wirkung von Schule (für den einzelnen Schüler/die einzelne Schülerin) in einer bestimmten Kohorte, d.h. zu einem gegebenen Zeitpunkt, im Mittelpunkt steht, zielt die Schulentwicklungsforschung auf die Beschreibung und Prädiktion der Veränderungen von Schulen ab. Die Bedeutung dieses Perspektivwechsels ist in der Schulforschung allerdings bislang kaum systematisch aufgearbeitet (vgl. dazu Klieme & Steinert, 2008). Selbst neuere Ansätze zu einem „Dynamic model of school effectiveness“ (Creemers & Kyriakides, 2008) oder einer Theorie der *Schulgestaltung* (Fend, 2008) stützen sich

nach wie vor stark auf die querschnittliche Schulforschung (wozu wir hier auch Studien zählen, die *Schülerinnen und Schüler* wiederholt testen, sofern sie die Schul- und Unterrichtsmerkmale nur einmal messen) und auf Analysen effektiver Schulen, während längsschnittliche Erhebungen nur in wenigen Fällen vorgenommen wurden. In der Methodologie sozialwissenschaftlicher Forschung ist hinlänglich bekannt, dass Entwicklungsfragen nicht mit Querschnittsdaten untersucht werden können (Steinhage & Blossfeld, 1999). Längsschnittliche Betrachtungen von Schulen sind beispielsweise unerlässlich, um rekursive Effekte, d.h. wechselseitige Abhängigkeiten zwischen Bedingungsfaktoren und Kriterien, identifizieren zu können. Beispiele erster deutscher Untersuchungen mit längsschnittlichem Design auf Schulebene sind die Studie zur Entwicklung von Ganztagschulen (StEG; vgl. Klieme, Fischer, Holtappels, Rauschenbach & Stecher, 2010), die Begleitforschung zum Projekt „Selbstständige Schule“ (Holtappels, Klemm & Rolff, 2008) sowie die Untersuchung von Trautwein, Neumann, Nagy, Lüdtke und Maaz (2010) zur gymnasialen Oberstufe. International stellt die Arbeit von Bryk, Sebring, Allensworth, Luppescu und Easton (2010) zur Schulreform der Jahre 1988-1996 in Chicago ein wegweisendes Beispiel dar.

Schulentwicklung im engeren Sinn kann definiert werden als die „bewusste und systematische Entwicklung von Einzelschulen“ (Rolff, 1998, S. 326). Diese Art der Schulentwicklung wird von Rolff als *intentionale* Schulentwicklung oder Schulentwicklung 1. Ordnung bezeichnet. In der Literatur finden sich verschiedene Ansätze zur Schulentwicklung, die aus unterschiedlichen Perspektiven und Bezugstheorien hervorgehen (Bohl, 2009; Rahm, 2005). Schulentwicklung findet auf verschiedenen Ebenen statt, so wird Schulentwicklung von Rolff als „Trias von personaler Entwicklung, Unterrichtsentwicklung und Organisationsentwicklung“ beschrieben (1998). Für substantielle Veränderungen müssen Veränderungen auf einer Ebene zu Veränderungen auf den anderen Ebenen führen, so dass der Ansatzpunkt der Schulentwicklung nicht festgelegt ist (Rolff, 2010). In diesem Beitrag interessieren uns v.a. die Aktivitäten auf Ebene der Organisation und die Bedeutung der Organisationsentwicklung für die Entwicklung von Schulen.

#### 1.4 Prädiktoren der Entwicklung von Schulen

In der Schulentwicklungsforschung muss, wie bereits angesprochen, die Relevanz einzelner Prädiktoren nicht mit jener in der Schulqualitätsforschung übereinstimmen. Während im Bereich der Schulqualitätsforschung nach der „These der primären Bedeutung proximaler Faktoren“ (siehe Abschnitt 1.2) die schülernahen Faktoren den größten Effekt auf die individuelle Schülerleistung haben, könnten für die Entwicklung von Schulen gerade die sogenannten distalen Faktoren, die Faktoren auf Schulebene, eine besondere Bedeutung besitzen, denn Veränderungen der Schule als Organisation müssen sich auf die Organisation selbst beziehen und von Kollegien und Schulleitungen getragen werden. Systemtheoretisch umformuliert: Was für das Lehr-Lern-System proximal ist, ist für das System Schule als Organisation distal, und umgekehrt.

Im Folgenden stellen wir die schon angesprochenen Faktoren, die individuelle Förderung, die ganztägige Schulorganisation und Evaluationsaktivitäten, dar.

### **Individuelle Förderung**

Eine wirksame individuelle Förderung von Schülerinnen und Schülern wurde in den letzten Jahren verstärkt eingefordert. Dies findet unter anderem in einer stärkeren Verankerung in den rechtlichen Richtlinien Ausdruck. So ist beispielsweise in Nordrhein-Westfalen „mit Verabschiedung des neuen Schulgesetzes im Jahr 2006 die individuelle Förderung jeder Schülerin und jedes Schülers als Leitidee nordrhein-westfälischer Bildungspolitik erstmals verankert worden“ (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2010, S. 3). Auch in den Richtlinien anderer Bundesländer wird die hohe Relevanz individueller Förderung mehr oder weniger deutlich hervorgehoben.

Allgemein bedeutet individuelle Förderung nach Fischer, Mönks und Westphal (2008) „die Anpassung des Forder-Förder-Angebots der vorschulischen und schulischen Umwelt an die kognitiven, sozial-emotionalen und psycho-motorischen Forder-Förder-Bedürfnisse des Kindes und Jugendlichen mit dem Ziel seiner optimalen Begabungsentfaltung und Persönlichkeitsentwicklung“ (S. 1). Individuelle Förderung der Schülerinnen und Schüler kann auf vielfältige Weise erreicht werden, und die Art der Umsetzung wird in den rechtlichen Richtlinien nicht festgelegt. Klieme und Warwas (2011) unterscheiden drei Varianten, die in der Schulpraxis in Deutschland verbreitet sind: (a) Trainings- und Zusatzangebote, z.B. für den Spracherwerb, (b) vielfältige, offene Lerngelegenheiten und (c) adaptiven Unterricht. Empirische Ergebnisse deuten darauf hin, dass Training und adaptiver Unterricht dann lernwirksam sind, wenn sie lernpsychologisch und didaktisch fundiert sowie gut strukturiert eingesetzt werden, während vielfältige Angebote und offene Lerngelegenheiten, wenn überhaupt, eher im affektiven Bereich wirken.

Adaptiver Unterricht wurde in PISA 2009 für Deutschland anhand einer Skala zu Differenzierungsmaßnahmen im Deutschunterricht untersucht (Hertel et al., 2010, S. 134-135), aber nicht in PISA 2000. In der vorliegenden Studie beschränken wir uns daher auf extracurriculare Angebote sowie die Wahrnehmung von Förderkursen in Deutsch durch die Schülerinnen und Schüler und prüfen, ob der Umfang extracurricularer Angebote einen Nutzen für die Entwicklung des Leseleistungsniveaus und der Beziehungsqualität hat. Angenommen wird, dass das Ausmaß solcher Angebote über den Zeitraum von neun Jahren keinen oder nur einen geringen Einfluss auf die Kriterien für die Entwicklung der Schulen hat, weil es in der Regel wenig mit den oben genannten Kernprozessen der schulischen Organisation (z.B. Entwicklung von Schulcurriculum und Leistungserwartungen; Zusammenarbeit zwischen Schulleitung, Kollegium und Eltern; Evaluationspraxis) verbunden ist.

### **Ganztägige Schulorganisation**

Seit Anfang des neuen Jahrhunderts bildet der Ausbau von Ganztagschulen einen Schwerpunkt der Bildungspolitik in Deutschland. Die ersten PISA-Ergebnisse lieferten dafür Argumente, aber die Ziele sind vielfältig. So sieht das Forum Bildung (Arbeitsstab Forum Bildung, 2001) in Ganztagschulen eine Möglichkeit der individuellen Förderung, da man den Schülerinnen und Schülern hier sowohl zeitlich als auch erzieherisch besser gerecht werden kann als in Halbtagschulen (Klemm, 2011; vgl. auch Rabenstein, 2009). Verschiedene Funktionen von Ganztagschule werden auch bei Holtappels (2006, S. 9-11) deutlich; u.a. werden „ganztägige Schulformen als Reaktion auf gewandelte Bildungsanforderungen“ verstanden und die „erweiterte Schulzeit zur Entwicklung der Lernkultur und der Förderungsintensität“ genutzt, häufig aber steht der Betreuungsbedarf berufstätiger oder alleinerziehender Eltern im Vordergrund.

Die bundesweite Studie zur Entwicklung von Ganztagschulen (Klieme, Fischer et al., 2010) zeigt, dass fach- und lernbezogene Angebote deutlich seltener vorkommen als sportliche, musische oder Freizeit-Angebote, und dass die individuelle Nutzung oft sporadisch ist. Individuelle Wirkungen sind eher schwach, aber vor allem im affektiven Bereich und im Sozialverhalten statistisch nachweisbar. Hinsichtlich ihrer Schulleistungen profitieren teilnehmende Schülerinnen und Schüler, sofern die Qualität der Angebote gut ist. Die Einführung des Ganztagsbetriebs ist zunächst eine schulorganisatorische Maßnahme: Es müssen Konzepte entwickelt und in Gremien verabschiedet, Räume und Zeitabläufe neu geplant, externe Partner gefunden, Finanzen verwaltet und Personal eingestellt werden. Wir untersuchen, inwieweit diese Veränderungen der Schulorganisation einen Beitrag zur Entwicklung von Schulen leisten.

### **Evaluationsaktivitäten**

Evaluationsaktivitäten können als ein distaler Faktor aufgefasst werden, da sie meistens von den Lernaktivitäten der einzelnen Schülerin/des einzelnen Schülers weiter entfernt auf der Ebene der Schule angesiedelt sind. Dennoch ist die „Implementierung systematischer interner und externer Evaluationsverfahren“ nach van Ackeren und Brauckmann (2010) eine der „wahrscheinliche[n] Ursachen für den Erfolg von Bildungssystemen“ (S. 42). Klieme (2005) sieht Evaluation als eine „Abfolge von teilweise einander bedingenden Entscheidungen“. Evaluation ist „der Prozess, innerhalb dessen eine zweckgerichtete Auswahl von Bewertungskriterien erfolgt, eine Institution oder Maßnahme auf Basis dieser Kriterien systematisch untersucht und bewertet wird und eine Kommunikation über die Bewertung stattfindet mit dem Ziel, Konsequenzen abzuleiten“ (S. 41).

Evaluation lässt sich auf unterschiedliche Art klassifizieren (vgl. Holtappels, 2003). Unterteilt man Evaluation nach den durchführenden Personen, unterscheidet man *interne* Evaluation bzw. Selbstevaluation, bei der die Beteiligten/die Schulen die Evaluation selber durchführen, von der *externen* Evaluation, die durch externe Personen/Institutionen durchgeführt wird. Evaluation von Schule hat gerade in den letzten Jahren stark an Bedeutung gewonnen (Boller, 2009) und wird immer stärker gesetzlich verankert (Altrichter & Maag Merki, 2010; Berkemeyer & Müller, 2010; Maag Merki, 2009).

Da die Evaluationsmaßnahmen meist unmittelbar auf Veränderungen auf der Ebene von Schulen abzielen, lässt sich hier ein stärkerer Einfluss auf die Entwicklung von Schulen vermuten.

## 2. Zielsetzung und Design des Schulpanels

Ziel des Schulpanels ist die Erfassung der Entwicklung der beteiligten 59 Gymnasien über einen Zeitraum von knapp einem Jahrzehnt, einschließlich der Möglichkeit, Stabilität und Veränderung in ausgewählten schulischen Merkmalen zu untersuchen. Daneben können aufgrund des längsschnittlichen Designs Hinweise auf mögliche Wirkungszusammenhänge zwischen den untersuchten Schulmerkmalen gewonnen werden. Die Abbildung von Veränderungsprozessen auf Schulebene ist in querschnittlichen Erhebungen, wie sie in der Schuleffektivitätsforschung vorwiegend genutzt werden, nicht möglich, weswegen auf Basis dieser Studien keine kausalen Rückschlüsse gezogen werden können. Klieme und Steinert (2008, S. 229-230) beschreiben zwar erste Studien, in denen „eine Verbindung des Schulentwicklungsdiskurses und quantitativ-empirischer Schuleffektivitätsforschung erkennbar“ ist, machen jedoch den Mangel an längsschnittlichen Untersuchungen auf Schulebene deutlich, die nicht nur die Ausgangsbedingungen der Schule (Input) und Ergebnisse der Schülerinnen und Schüler (Output), sondern auch die Prozesse an der Schule berücksichtigen. Die vorliegende Studie soll einen Beitrag zu dieser Form der Schulentwicklungsforschung darstellen.

Die Datenerhebung für das Schulpanel erfolgte zu zwei Messzeitpunkten in einem Abstand von neun Jahren: Die Daten aus der Teilnahme der 59 Gymnasien an PISA 2000 wurden ergänzt durch eine wiederholte Untersuchung dieser Schulen mittels der Fragebögen und Leistungstests aus PISA 2009 (den Schülerinnen und Schülern wurde eine leicht modifizierte Version des Fragebogens vorgegeben, siehe Abschnitt 4.2). Neben quantitativen Daten, die in beiden Jahren erhoben wurden und den Schwerpunkt des Designs ausmachten, wurden im Frühjahr 2010, also zehn Jahre nach der ersten Untersuchung, mittels retrospektiver Interviews qualitative Daten zu Veränderungen an den Schulen gewonnen.

## 3. Fragestellungen

Mit Hilfe der gewonnenen Daten sollten folgende Fragen untersucht werden: Welche Veränderungen zeigen die Schulen hinsichtlich ihrer Entwicklungsaktivitäten? Wie hängen diese Veränderungen mit Veränderungen im Leistungsniveau der Schulen bzw. Veränderungen in der wahrgenommenen Beziehungsqualität an den Schulen von 2000 zu 2009 zusammen? Daraus ergaben sich drei Hypothesen zur Veränderung von Schulmerkmalen, welche sich aus der veränderten gesetzlichen Grundlage und veränderten bildungspolitischen Prioritäten ergeben (siehe Abschnitt 1), und drei Hypothesen zum Zusammenhang mit dem Leistungsniveau bzw. der Beziehungsqualität an den Schulen,

die auf theoretischen Überlegungen beruhen (ebenfalls Abschnitt 1). Diese werden im Folgenden noch einmal gebündelt dargestellt:

1. Schülerinnen und Schüler nehmen vermehrt an Kursen zur individuellen Förderung teil.
2. Es findet eine verstärkte Ausrichtung der Schulen zu ganztägigen Lerngelegenheiten statt.
3. An den Schulen zeigt sich ein deutlicher Zuwachs von interner und externer Evaluation.
4. Die Schulorganisation als Ganztagschule beeinflusst die Veränderung der wahrgenommenen Beziehungsqualität positiv.
5. Das Niveau der individuellen Förderung einzelner Schülerinnen und Schüler (Wahrnehmung von Kursen zur individuellen Förderung; Breite des extracurricularen Angebots) steht in keiner bzw. einer geringfügigen Beziehung mit Veränderungen im Leseleistungsniveau und der wahrgenommenen Beziehungsqualität der Schulen.
6. Veränderungen hin zu mehr Evaluationsaktivitäten haben einen praktisch relevanten positiven Effekt auf die Entwicklung der Schule, insbesondere auf das Leistungsniveau im Lesen an den Schulen.

## 4. Methoden<sup>2</sup>

### 4.1 Stichprobe

Die Zielpopulation des Schulpanels bestand aus Gymnasien aus ganz Deutschland. Aus den Gymnasien, die bereits an PISA 2000 teilgenommen hatten, wurde nach Bundesland stratifiziert eine zufällige Stichprobe von 60 Schulen gezogen; im Falle fehlender Teilnahmebereitschaft am Follow-up wurde, ebenfalls nach Zufall, eine Ersatzschule bestimmt. Zu einer Teilnahme an der weiterführenden Erhebung der Fragebogen- und Leistungsdaten erklärten sich schließlich 59 der insgesamt 90 angeschriebenen Gymnasien (inklusive Ersatzschulen) bereit. In der realisierten Stichprobe sind alle Bundesländer außer Bremen vertreten.

Die Stichproben der Schülerinnen und Schüler innerhalb der Schulen bestand in 2000 aus max. 28 ( $M = 25.3$ ) Fünfzehnjährigen (Geburtsdatum zwischen dem 1. Februar 1984 und dem 31. Januar 1985) und in 2009 aus max. 25 ( $M = 21.6$ ) Fünfzehnjährigen (Geburtsjahr 1993). Insgesamt ergab sich in 2000 eine Fallzahl von  $N = 1493$  und in 2009 von  $N = 1273$  Schülerinnen und Schülern. Die um einen Monat abweichenden Zeiträume beim Geburtsdatum der Schülerinnen und Schüler ergaben sich durch ein verändertes Er-

---

2 Wir danken dem IEA Data Processing and Research Center (IEA DPC) für die Datenerhebung, -erfassung und -kodierung sowie für die Herstellung der Schulkontakte. Claus Carstensen (Universität Bamberg) danken wir für die Skalierung der Leistungsdaten.

hebungszeitfenster. Im Jahr 2000 fand die Erhebung der PISA-Daten im Mai und Juni statt, in 2009 einen Monat früher im April und Mai. Durch die entsprechend angepassten Zeiträume für das Geburtsdatum der Schülerinnen und Schüler ergaben sich für beide Messzeitpunkte vergleichbare Altersverteilungen der Stichproben. Den Schulleiterfragebogen füllten in 2000  $N = 55$  Schulleiterinnen und Schulleiter und in 2009  $N = 56$  Schulleiterinnen und Schulleiter aus. Interviewdaten liegen für  $N = 47$  Schulen vor.

## 4.2 Instrumente

Die quantitative Datenbasis des Schulpanels beruht auf den Fragebögen und Leistungstests aus PISA 2000 und PISA 2009. Eine Ausnahme stellt der Schülerfragebogen dar, der für das Schulpanel in einer vom PISA-Schülerfragebogen 2009 abweichenden Version vorgegeben wurde, um eine höhere Übereinstimmung mit den Fragen in PISA 2000 zu erreichen. Die Vergleichbarkeit der Leistungsdaten wurde durch die Verwendung von Link-Items zur Reading Literacy in beiden Jahren möglich. Während die Fragebögen und Leistungstests eine Momentaufnahme der jeweiligen Messzeitpunkte darstellten, wurden in den retrospektiven leitfadengestützten Interviews der qualitativen Erhebung Veränderungen an der Schule von 2000 bis 2009 aus der Sicht einer Vertreterin bzw. eines Vertreters der (erweiterten) Schulleitung (bzw. für vier Schulen aus der Sicht einer von der Schulleitung benannten Lehrkraft) erfragt.

### **Lesekompetenz**

Lesekompetenz wird hier entsprechend der Definition der Hauptstudie PISA 2009 aufgefasst: „Lesekompetenz wird im Sinne einer Basiskompetenz verstanden, von der angenommen wird, dass sie in modernen Gesellschaften für eine befriedigende Lebensführung in persönlicher und wirtschaftlicher Hinsicht sowie für eine aktive Teilnahme am gesellschaftlichen Leben notwendig ist. Dies beinhaltet eine funktionale Sicht auf Lesekompetenz als basales Kulturwerkzeug. Die Beherrschung der Muttersprache in Wort und Schrift gehört in allen modernen Gesellschaften zum Kernbestand kultureller und sprachlicher Literalität“ (Naumann, Artelt, Schneider & Stanat, 2010, S. 24). Lesekompetenz umfasst „understanding, using, reflecting on and engaging with written texts, in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society“ (Artelt, Naumann & Schneider, 2010, S. 74; OECD, 2009, S. 23).

Für die Ermittlung des Leistungsniveaus im Lesen wurden für das Schulpanel 36 Link-Items verwendet, die sowohl in PISA 2000 als auch 2009 vorgegeben wurden. Um Unterschiede im Testheftdesign zwischen beiden Zyklen zu kontrollieren, wurden im Schulpanel nicht nur die Testhefte aus der internationalen PISA-2009-Studie verwendet, sondern zusätzlich auch Testhefte in der Zusammenstellung aus PISA 2000. Auswirkungen der Unterschiede im Testheftdesign auf die Itemschwierigkeiten wurden im Skalierungsmodell berücksichtigt und statistisch kontrolliert. Für die Ermittlung des Lesekompetenzniveaus in den Schulen lagen im Schnitt für 2000 pro Schule

$N = 19.56$  Leistungswerte vor ( $Min = 9$ ;  $Max = 22$ ). Aus der Erhebung 2009 liegen im Mittel  $N = 14.39$  Leistungswerte pro Schule vor ( $Min = 6$ ;  $Max = 19$ ).

Da die Lesekompetenz der Schülerinnen und Schüler über die Link-Items der Testhefte aus 2000 und 2009 ermittelt wurde und bedingt durch das Testheftdesign nicht jede Schülerin/jeder Schüler Link-Items zur Erfassung ihrer/seiner Lesekompetenz bearbeitet hat, ist die Gesamtzahl der Leistungswerte pro Schule niedriger als die Gesamtzahl der in der Schule getesteten und befragten Schülerinnen und Schüler.

Zur Schätzung der Lesekompetenz wurden die Daten der Schülerinnen und Schüler aus den Jahren 2000 und 2009 gemeinsam skaliert, die Itemschwierigkeiten wurden frei geschätzt. Damit sind die resultierenden Werte nicht auf der internationalen PISA-Skala verankert, aber optimal für den Vergleich des Leistungsniveaus zwischen 2000 und 2009. Für die Analysen wurden Punktschätzer (Weighted Likelihood Estimates, WLEs) der individuellen Lesekompetenz verwendet.

### **Wahrnehmung der Beziehungsqualität durch die Schülerschaft**

Die Skala zur Beziehungsqualität gibt an, inwiefern sich Schülerinnen und Schüler durch ihre Lehrpersonen gut behandelt fühlen. Die wahrgenommene Beziehungsqualität wurde als ein Aspekt von Schülerwohlbefinden über drei Items aus dem Schülerfragebogen erfasst. Die Items erforderten eine Antwort im Antwortrange von „stimme überhaupt nicht zu“ (1) bis „stimme ganz zu“ (4) auf Aussagen wie „Die meisten meiner Lehrer/innen interessieren sich für das, was ich zu sagen habe“. Im Erhebungsjahr 2000 hat diese Skala eine Reliabilität von  $\alpha = .68$  und in 2009 von  $\alpha = .71$ . Ein höherer Wert auf der Skala steht für eine positivere wahrgenommene Beziehungsqualität.

### **Individuelle Förderung**

Als Indikatoren der individuellen Förderung durch die Schulen wurden die Breite des extracurricularen Angebots und die Teilnahme der Schülerschaft an Förderangeboten berücksichtigt.

#### *Breite des extracurricularen Angebots*

Sowohl im Jahr 2000 als auch im Jahr 2009 wurden im PISA-Schulleiterfragebogen außerunterrichtliche Angebote erfasst. Die verwendeten Items unterscheiden sich für beide Messzeitpunkte erheblich, so dass die Angaben nur eingeschränkt zueinander in Beziehung gesetzt werden können. Als Indikator des extracurricularen Angebots einer Schule wird zu beiden Erhebungszeitpunkten der prozentuale Anteil der von der Schule angebotenen extracurricularen Optionen an den im Schulleiterfragebogen erfragten extracurricularen Optionen verwendet. Im Jahr 2000 gaben Schulleiterinnen und Schulleiter für 35 extracurriculare Angebote an, ob diese an ihrer Schule stattfinden. Für 29 der Items lagen Ja/Nein-Antworten vor; für sechs Items wurden die Häufigkeitsangaben der Schulleiter von 1 (nie) bis 4 (mehr als viermal im Jahr) umkodiert zu Ja/Nein-Antworten. Im Jahr 2009 bejahten oder verneinten die Schulleiter das Bestehen von 22 ex-



tracurricularen Angeboten. Zu den Angeboten, die zu beiden Erhebungszeitpunkten erfasst wurden, gehören beispielsweise Leseclubs, musisch-künstlerische Angebote, mathematisch-naturwissenschaftliche Angebote und eine Schülerzeitung. Die Frage nach dem Angebot eines Debattierklubs hingegen wurde nur im Erhebungsjahr 2009 gestellt, während die Frage nach Präsentationen von Verlagen oder Buchhandlungen spezifisch für das Jahr 2000 war. Der Index zur Breite des extracurricularen Angebots (Prozentanteil der vom Schulleiter bejahten Angebote) hat in 2000 eine Reliabilität von  $\alpha = .70$  und in 2009 von  $\alpha = .74$ .

#### *Teilnahme der Schülerinnen und Schüler an Deutsch-Förderangeboten*

Neben der Bereitstellung von extracurricularen Angeboten interessiert die Nutzung von Förderangeboten durch die Schülerinnen und Schüler. Da im vorliegenden Beitrag die PISA-Lesekompetenz als Leistungsmaß verwendet wurde, wird die folgende Frage nach der Teilnahme an Förder-/Stützkursen in Deutsch mit den Antwortkategorien „Nein, nie“ (0), „Ja, gelegentlich“ (1) und „Ja, regelmäßig“ (2) berücksichtigt: „Hast du in den letzten 3 Jahren in der Schule zusätzlichen Unterricht gehabt, um deine Leistungen zu verbessern?“.

### **Ganztägige Schulorganisation**

Fragen zur ganztägigen Schulorganisation wurden aufgrund des Fehlens entsprechender Items im PISA-2000-Schulleiterfragebogen im Rahmen des retrospektiven Interviews gestellt. Eine Schule wurde in Anlehnung an die Definition der Kultusministerkonferenz (KMK, 2011, S. 9) als Ganztagschule definiert, wenn an mindestens drei Tagen in der Woche ein ganztägiges Angebot besteht, dieses mindestens sieben Zeitstunden umfasst und Mittagsverpflegung für die Schülerschaft angeboten wird. Eine Schule wird als Ganztagschule kodiert (1), wenn sie diese drei Kriterien erfüllt. Treffen weniger als die drei Kriterien zu, ist die Schule nach dieser Kodierung keine Ganztagschule (0). Das Kriterium der konzeptuellen Verbindung von Unterricht und Ganztagsangebot, das die KMK zusätzlich bei der Kategorisierung einer Schule als Ganztagschule voraussetzt, konnte hier nicht berücksichtigt werden; diese Forderung bestand im Jahr 2000 noch nicht.

### **Evaluation**

Evaluationsaktivitäten an den Schulen wurden im Rahmen der retrospektiven Interviews erhoben. Hier wurden die Interviewpartner nach Veränderungen sowohl in der internen als auch in der externen Evaluation befragt. Aus diesen Angaben wurde jeweils eine Variable zum Vorhandensein von Evaluation zwischen 2000 und 2009 erzeugt. Für Schulen, für die im Interview ein konstantes oder ein vermehrtes Ausmaß an interner bzw. externer Evaluation im Zeitraum 2000 bis 2009 angegeben wurde, wurde Vorhandensein von interner bzw. externer Evaluation (1) kodiert. Bei Schulen, für die eine Abnahme der Evaluationsaktivitäten bzw. kein Stattfinden von Evaluationsaktivitäten berichtet wurde, wurde Fehlen interner bzw. externer Evaluation (0) kodiert. Da eine

Abnahme von Evaluationsaktivitäten weder für interne noch für externe Evaluation berichtet wurde, enthält die Fehlen-Kategorie ausschließlich Schulen, die keine Evaluationsaktivitäten berichteten. Für beide Prädiktoren wurde demnach – anders als bei den anderen Schulmerkmalen, für die messzeitpunktspezifische Angaben vorlagen – nur jeweils eine Variable erzeugt, die das Vorhandensein von Evaluationsaktivitäten im Zeitraum 2000 bis 2009 erfasste.

### **Zusammensetzung der Schülerschaft**

Die Zusammensetzung der Schülerschaft wurde hinsichtlich zweier Komponenten untersucht: soziale Komposition und Migrantanteil. Die soziale Komposition der jeweiligen Schule ergab sich durch den mittleren sozioökonomischen Status der Familien ihrer Schülerinnen und Schüler. Die Angaben der Eltern und Schülerinnen und Schüler zum Elternberuf wurden mit Hilfe der International Standard Classification of Occupation (ISCO-88) klassifiziert und in den internationalen sozioökonomischen Index (International Socio-Economic Index of Occupational Status, ISEI) umgerechnet (vgl. Ganzeboom & Treiman, 1996). Der sozioökonomische Index wurde anhand der Elternangaben bestimmt, bei fehlenden Werten wurden die Angaben aus dem Schülerfragebogen herangezogen. Auf dieser Basis wurde für die Schülerinnen und Schüler der HISEI (highest ISEI), der höchste sozioökonomische Index der Familie, gebildet, indem der jeweils höchste Wert aus Vater- und Mutter-ISEI ausgewählt wurde.

Eine Schätzung für den Anteil an Migranten wurde je Schule aus den Schülerangaben zum Geburtsort der Schülerin bzw. des Schülers, ihrer/seiner Mutter und ihres/seines Vaters ermittelt. Der Migrationsstatus wurde hier gemäß dem Vorgehen bei PISA 2009 definiert. Eine Schülerin/ein Schüler gilt dabei als Schüler/in mit Migrationshintergrund, wenn mindestens ein Elternteil im Ausland geboren wurde. Eine genaue Darstellung der Definition kann Stanat, Rauch und Segeritz (2010, S. 203-204) entnommen werden.

### *4.3 Analysemethoden*

Angesichts der hierarchischen Struktur der Daten (Schüler geschachtelt in Schulen) erscheint für die Untersuchung der Stabilität der Schulmerkmale bzw. für die Vorhersage von Veränderungen in den Schulmerkmalen die Analyse mithilfe von Mehrebenenmodellen als die angemessene Vorgehensweise. Ein Einsatz gängiger längsschnittlicher Mehrebenenmodelle – als Mehrebenen-Regressionsmodell (z.B. Singer & Willett, 2003) oder in Form von Mehrebenen-Pfad- bzw. Strukturgleichungsmodellen (z.B. Kaplan, 2008) – ist im vorliegenden Fall allerdings nicht möglich, da solche Modelle wiederholte Erhebungen mit denselben Schülerinnen und Schülern voraussetzen. Eine Messwiederholung liegt jedoch nur auf der Ebene der Schulen vor. Eine naheliegende Lösung bestünde darin, die Daten messzeitpunktspezifisch zu aggregieren und auf Mehrebenenmodelle zugunsten einer Auswertung mit den Schulen als Analyseeinheiten zu verzichten. Insbesondere im Falle kleiner Gruppen (wenige Schüler je Schule)

und geringer Intraklassenkorrelationen (geringer Varianzanteil auf Schulebene) könnte es mit dieser Methode zu fehlerhaften Resultaten kommen, und es sollte auf alternative mehrebenenanalytische Auswertungsstrategien zurückgegriffen werden.

Um die Stabilität von Schulmerkmalen sowie die wechselseitige längsschnittliche Beziehung zwischen Schulmerkmalen in PISA 2000 und PISA 2009 zu untersuchen, wurden Mehrebenen-Pfadmodelle verwendet, die für eine Messwiederholung auf der Schulebene angepasst wurden. Dazu wurden Modelle mit einer Schüler- und einer Schulebene spezifiziert; die Ebene der Schulklasse konnte nicht berücksichtigt werden, da die PISA-Stichproben der Fünfzehnjährigen nur auf Basis ihrer Schul- und nicht auf Basis ihrer Klassenzugehörigkeit rekrutiert wurden. Der Typus der verwendeten Modelle wird exemplarisch in Abbildung 1 veranschaulicht. Der untere Teil der Abbildung stellt die Schülerebene (Ebene 1), der obere Teil die Schulebene dar (Ebene 2). Die Pfeile stehen für die angenommenen gerichteten oder ungerichteten Zusammenhänge zwischen den Variablen. Zur Untersuchung der wechselseitigen Beziehung zwischen Schulmerkmalen wurden auf Schulebene Cross-lagged-panel-Modelle spezifiziert (vgl. Abbildung 1). In diesen Modellen wurden jeweils ein Kriterium (Leseleistung oder Beziehungsqualität) und ein Prädiktor berücksichtigt. Zu diesen Prädiktoren zählten das Ganztagsangebot, der Index zur Breite des extracurricularen Angebots, die Teilnahme der Schülerinnen und Schüler an Deutsch-Förderkursen, der HISEI, der Migrantenanteil und das Vorhandensein von interner bzw. externer Evaluation. Die Stabilität der Schulmerkmale wurde in Modellen analysiert, in die jeweils nur das interessie-

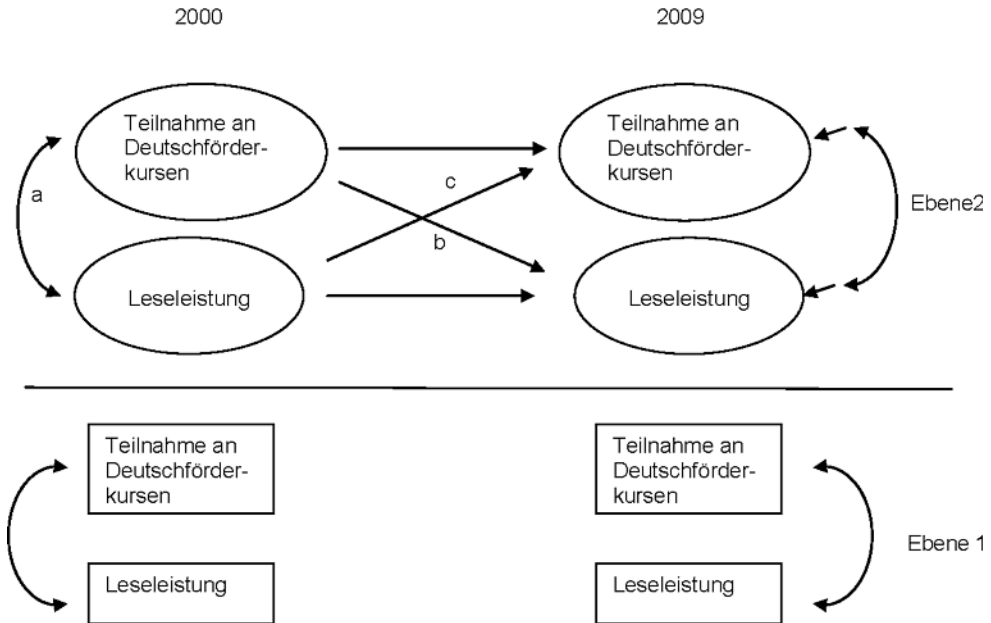


Abb. 1: Schematische Darstellung der Cross-lagged-panel-Modelle am Beispiel der Leseleistung und der Teilnahme an Deutsch-Förderkursen durch die Schülerinnen und Schüler

rende Merkmal einging. Die Stabilität wurde dann über die Korrelation des Merkmals auf Schulebene zwischen PISA 2000 und 2009 ermittelt.

Einen Sonderfall unter den auf Schulebene einbezogenen Variablen stellt die Evaluation dar. Hierfür liegen keine unabhängig in 2000 und 2009 erhobenen Daten vor (siehe Abschnitt 4.2, Evaluation). In die Analysen ging folglich nur eine Variable „Vorhandensein interner Evaluation“ bzw. „Vorhandensein externer Evaluation“ als Prädiktor für die Kriterien in 2009 ein.

Im Falle von Merkmalen, die auf Schülerebene erhoben wurden, wurden auf Schülerebene lediglich die Varianzen und – bei zwei Merkmalen dieses Typs in einem Modell (vgl. Abbildung 1) – die Kovarianzen innerhalb desselben Messzeitpunkts frei geschätzt. Alle Kovarianzen zwischen den Zeitpunkten wurden dagegen auf null gesetzt, da auf Schülerebene keine Messwiederholung stattgefunden hatte.

Die Ermittlung der Stabilität der Schulmerkmale und die Analysen zur Vorhersage von Veränderung erfolgten mithilfe von Mplus 5. Die Veränderungen im Niveau der Schulmerkmale zwischen 2000 und 2009 wurden mit SPSS 18 analysiert.

## 5. Ergebnisse

Im Folgenden werden die Ergebnisse zur Veränderung im Niveau der Schulqualitätsmerkmale und unserer Kriterien für die Entwicklung von Schulen zwischen 2000 und 2009 dargestellt (Abschnitt 5.1 und 5.2). In Abschnitt 5.3 wird die Stabilität der Schulmerkmale berichtet, in 5.4 wird dann auf die Vorhersage der Veränderung der Kriterien eingegangen.

### 5.1 Veränderungen im Niveau der Schulqualitätsmerkmale

#### **Veränderung in der Förderpraxis der Schulen**

Ein t-Test für abhängige Stichproben zeigt, dass Schülerinnen und Schüler vermehrt an Förderkursen in Deutsch teilnehmen ( $t [58] = 3.142, p = .003$ ). Der Anteil der Schülerinnen und Schüler, die mindestens gelegentlich an Förderkursen in Deutsch teilnehmen, ist in 2009 mit 5% der Schülerinnen und Schüler, für die Angaben zur Teilnahme vorliegen, weiterhin sehr gering, jedoch signifikant höher als in 2000 (2%).

#### **Veränderung der Schulorganisation durch Ganztagsangebote**

Es zeigt sich für unsere gymnasiale Stichprobe ein signifikanter Trend hin zur Ganztagschule (Wilcoxon-Test:  $Z = -4.123, p < .001$ ). 17 Schulen waren 2000 nach unserer Definition keine Ganztagschulen, erfüllen im Jahr 2009 jedoch die Kriterien. Für 14 Schulen ergibt sich auf Grundlage unserer Definition keine Veränderung der Schulorganisation, hiervon erfüllten 2 bereits im Erhebungsjahr 2000 die Kriterien einer Ganztagschule, während die restlichen 12 in beiden Erhebungen die Kriterien nicht erfüllen.

### **Veränderung in der Evaluationspraxis der Schulen**

Die Befragung der Schulleiter zu Evaluationsmaßnahmen an ihren Schulen und deren Veränderung weist deutlich auf eine Zunahme der Evaluation hin. Für Maßnahmen der internen Evaluation gaben 23 der 33 Schulleiterinnen und Schulleiter, die hierzu Aussagen machten, eine Zunahme der Evaluationsaktivitäten an (70%), während 4 Personen berichteten, dass die Evaluationsmaßnahmen an ihrer Schule auf einem konstanten Niveau geblieben sind. An 6 der befragten Schulen (18%) fand bis 2009 keine interne Evaluation statt. Ein Nachlassen der Maßnahmen interner Evaluation wurde für keine Schule berichtet.

Eine Zunahme von Maßnahmen der externen Evaluation berichteten 19 der 30 Schulleiterinnen und Schulleiter (63%), die hierzu Angaben machten. Ein konstantes Niveau externer Evaluation wurde von 3 Personen angegeben. Ein weiterhin bestehendes Fehlen von externer Evaluation wurde aus 8 Schulen berichtet (27%). Ein Nachlassen im Umfang externer Evaluation wurde nicht berichtet.

## *5.2 Veränderungen im Niveau der Kriterien*

### **Veränderung im Leistungsniveau**

Ein t-Test für abhängige Stichproben zeigt keinen signifikanten Unterschied im mittleren Niveau der Leseleistung der Schulen zwischen 2000 und 2009 ( $t [58] = .559$ ,  $p = .578$ ).

### **Veränderung in der wahrgenommenen Beziehungsqualität**

Die mittlere wahrgenommene Beziehungsqualität in den Schulen verbessert sich von 2000 zu 2009 signifikant ( $t [58] = -3.121$ ,  $p = .003$ ).

## *5.3 Stabilität der Schulmerkmale über die Zeit*

Die Stabilität der Merkmale über die Zeit wurde basierend auf aggregierten Daten (Schulmittelwerte), aber auch mit Hilfe von Mehrebenenmodellen bestimmt. Tabelle 1 gibt einen Überblick über die ermittelten Stabilitäten der Merkmale über den Zeitraum von neun Jahren. Neben den Stabilitäten werden für beide Erhebungszeitpunkte die Intraklassenkorrelationen (ICC) des jeweiligen Merkmals aufgeführt. Die ICC gibt den Anteil der Varianz in der jeweiligen Variablen an, der auf Varianz zwischen den Schulen zurückzuführen ist.

Die tabellarische Darstellung der Ergebnisse zur Stabilität von Leistung, Beziehungsqualität, sozio-ökonomischem Status, Migration und Teilnahme an Förderkursen verdeutlicht teilweise große Unterschiede in der festgestellten Stabilität zwischen den Analysemethoden (Mehrebenen vs. aggregiert). In den Analysen mit den aggregierten Daten zeigen sich generell schwächere Zusammenhänge als in den Mehrebenenanalysen.

Merkmal	Korrelation 2000 – 2009 (S.E.) (basierend auf aggregierten Daten)	Korrelation 2000 – 2009 (S.E.) (basierend auf Mehrebenen- analysen)	ICC 2000	ICC 2009
Leseleistung	0.32** (0.12)	0.84** (0.32)	0.03	0.04
Wahrgenommene Beziehungsqualität	-0.03 (0.13)	-0.08 (0.25)	0.06	0.04
Sozio-ökonomischer Status (HISEI)	0.67*** (0.07)	0.97*** (0.12)	0.07	0.09
Migration	0.66*** (0.07)	0.99*** (0.08)	0.08	0.09
Teilnahme an Deutsch-Förderkursen	0.11 (0.13)	0.41 (0.51)	0.01	0.05

Anmerkungen: ICC 2000 ist die Intraklassenkorrelation im Erhebungsjahr 2000; ICC 2009 ist die Intraklassenkorrelation im Erhebungsjahr 2009; \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$  (zweiseitige Testung).

Tab. 1: Stabilitäten der Merkmale von Schulen, die sowohl mehrebenenanalytisch als auch aggregiert auf Schulebene betrachtet werden können

Für die auf Schulebene erfassten Merkmale „ganztägige Schulorganisation“ und „Breite des extracurricularen Angebots“ können ebenfalls Stabilitäten berichtet werden. Im Falle beider Merkmale resultiert kein signifikanter Zusammenhang zwischen den beiden Erhebungszeitpunkten (extracurriculares Angebot:  $r = -0.01$ , S.E. = 0.13, n.s.; ganztägige Schulorganisation:  $r = 0.20$ , S.E. = 0.17, n.s.).

#### 5.4. Vorhersage von Leseleistung und Beziehungsqualität 2009

Im Folgenden werden die Ergebnisse der Cross-lagged-panel-Modelle zur Vorhersage der Veränderung von Leseleistung und Beziehungsqualität dargestellt. Tabelle 2 gibt einen Überblick über die Prädiktion der Leseleistung. Die Spalte „Korrelation 2000 mit Leseleistung“ gibt die Korrelation zwischen der Leseleistung in 2000 und dem entsprechenden Prädiktor wieder (vgl. Abbildung 1, Pfeil a). Die Spalte „Effekte auf Leseleistung 2009“ enthält die standardisierten Regressionskoeffizienten für die Vorhersage der Leseleistung 2009 durch den jeweiligen Prädiktor 2000 (vgl. Abbildung 1, Pfeil b), die Spalte „Effekte von Leseleistung 2000“ enthält die standardisierten Regressionskoeffi-

zienten für die Vorhersage des Prädiktors 2009 durch die Leseleistung 2000 (vgl. Abbildung 1, Pfeil c). In Tabelle 3 werden die Ergebnisse zur Prädiktion der Beziehungsqualität gezeigt. Die Darstellung erfolgt nach dem gleichen Prinzip wie in Tabelle 2 für die Leseleistung.

Die Art der Signifikanztestung wurde orientiert an den Befunden bisheriger Forschung und unseren formulierten Hypothesen gewählt. So wurden Zusammenhänge, für die keine Richtungsannahmen getroffen wurden, zweiseitig getestet. Zusammenhänge, für die eine Richtungsannahme plausibel schien, wurden einseitig getestet. So zeigt die bisherige Forschung einen positiven Zusammenhang von HISEI und Leseleistung (Ehmke & Baumert, 2007; Ehmke & Jude, 2010) und einen negativen Zusammenhang von Migrantensstatus und Leseleistung (Stanat et al., 2010; Schwippert, Hornberg, Freiberg & Stubbe, 2007; Walter & Taskinen, 2007), weswegen die Zusammenhänge im Jahr 2000 einseitig geprüft wurden. Für die Breite des extracurricularen Angebots und die Teilnahme an Förderkursen nahmen wir für das Erhebungsjahr 2000 einen negativen Zusammenhang mit der Leseleistung an, da diese Fördermaßnahmen im Regelfall

Prädiktor	Korrelation 2000 mit Leseleistung (S. E.)	Effekt auf Leseleistung 2009 (S. E.)	Effekt von Leseleistung 2000 (S. E.)
<b>Individuelle Förderung</b>			
Breite des extracurricularen Angebots	-0.36 (0.23)	-0.08 (0.26)	0.30 (0.20)
Teilnahme an Deutsch-Förderkursen	-0.70* (0.26)	0.29 (0.85)	0.26 (0.72)
<b>Ganztägige Schulorganisation</b>			
<b>Evaluationsaktivitäten</b>			
Interne Evaluation	-----	0.49* (0.27)	-----
Externe Evaluation	-----	0.27 (0.42)	-----
<b>Zusammensetzung der Schülerschaft</b>			
HISEI	0.53** (0.16)	0.26 (0.28)	-0.33 (0.20)
Migration	-0.08 (0.32)	-0.41 (0.23)	0.10 (0.16)

Anmerkungen: Einseitig getestete Zusammenhänge sind kursiv, zweiseitig getestete Zusammenhänge normal dargestellt; \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

Tab. 2: Ergebnisdarstellung für die Prädiktion von Leseleistung basierend auf den Cross-lagged-panel-Modellen

remedial eingesetzt werden. Der Effekt von Evaluationsaktivitäten auf die Leseleistung in 2009 wurde aufgrund der gerichteten Annahme ebenfalls einseitig getestet. Für die Beziehungsqualität wurde – in Anlehnung an Ergebnisse aus StEG – ein positiver Effekt einer Veränderung hin zur ganztägigen Schulorganisation angenommen, entsprechend erfolgte die Testung hier ebenfalls einseitig.

Aus Tabelle 2 wird unter anderem ersichtlich, dass eine signifikante positive Korrelation zwischen dem mittleren sozio-ökonomischen Status in 2000 und der mittleren Leseleistung des gleichen Erhebungszeitpunkts besteht. Ein Effekt des mittleren sozio-ökonomischen Status im Jahr 2000 auf das mittlere Niveau der Leseleistung in 2009 unter Kontrolle der mittleren Leseleistung des ersten Messzeitpunktes lässt sich dagegen nicht nachweisen. Für die Teilnahme an Förderkursen in Deutsch ergibt sich zum ersten Erhebungszeitpunkt eine negative Korrelation mit der mittleren Leseleistung.

<b>Prädiktor</b>	<b>Korrelation 2000 mit Beziehungsqualität (S. E.)</b>	<b>Effekt auf Beziehungsqualität 2009 (S. E.)</b>	<b>Effekt von Beziehungsqualität 2000 (S. E.)</b>
<b>Individuelle Förderung</b>			
Breite des extracurricularen Angebots	-0.06 (0.16)	-0.05 (0.19)	0.22 (0.16)
Teilnahme an Deutsch-Förderkursen	-0.04 (0.31)	0.47 (0.37)	0.12 (0.22)
<b>Ganztägige Schulorganisation</b>	<i>0.30* (0.16)</i>	<i>0.45* (0.27)</i>	0.05 (0.19)
<b>Evaluationsaktivitäten</b>			
Interne Evaluation	-----	0.40* (0.19)	-----
Externe Evaluation	-----	-0.28 (0.19)	-----
<b>Zusammensetzung der Schülerschaft</b>			
HISEI	-0.08 (0.23)	-0.08 (0.19)	0.03 (0.13)
Migration	-0.29 (0.22)	-0.32 (0.26)	-0.16 (0.17)

*Anmerkungen:* Einseitig getestete Zusammenhänge sind kursiv, zweiseitig getestete Zusammenhänge normal dargestellt; \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

Tab. 3: Ergebnisdarstellung für die Prädiktion von Beziehungsqualität basierend auf den Cross-lagged-panel-Modellen



Auch in den Cross-lagged-panel-Modellen mit der Beziehungsqualität als Kriterium resultieren einige signifikante Zusammenhänge (vgl. Tabelle 3). Zum ersten Erhebungszeitpunkt im Jahr 2000 hängt eine ganztägige Schulorganisation positiv mit der mittleren Beziehungsqualität zusammen. Dieser Effekt zeigt sich auch längsschnittlich unter Kontrolle der Beziehungsqualität in 2000. Ein Effekt der mittleren Beziehungsqualität 2000 auf die ganztägige Schulorganisation 2009 (unter Kontrolle der Schulorganisation 2000) findet sich, wie zu erwarten war, dagegen nicht.

Wurden im Zeitraum von 2000 bis 2009 interne Evaluationsmaßnahmen an den Schulen durchgeführt, wirkt sich dies positiv sowohl auf die mittlere Leseleistung 2009 als auch auf die mittlere Beziehungsqualität 2009 aus. Maßnahmen der individuellen Förderung 2000 (Breite des extracurricularen Angebots und Teilnahme an Deutsch-Förderkursen) dagegen stehen erwartungsgemäß in keiner Beziehung zur Leseleistung und Beziehungsqualität im Jahr 2009.<sup>3</sup>

## 6. Zusammenfassung und Diskussion

Im vorliegenden Beitrag wurde die Entwicklung der Schulen im PISA-Schulpanel hinsichtlich verschiedener Aspekte untersucht. Erstes Ziel war es, Veränderungen von bzw. Niveauunterschiede in ausgewählten Schulmerkmalen zwischen 2000 und 2009 zu bestimmen.

Bei den untersuchten Gymnasien zeigt sich die erwartete Entwicklung der Organisationsform der Schulen hin zu mehr Ganztagschulbetrieb. Eine wichtige Rolle spielten hier sicherlich die starke Förderung des Ausbaus von Ganztagschulen durch Gelder des Investitionsprogramms „Zukunft Bildung und Betreuung (IZBB)“ (2003 bis 2009) sowie der Bedarf der Schulen nach einer Ausweitung des Unterrichts auf den Nachmittag als Folge der Einführung des achtjährigen Gymnasiums.

Die leicht gewachsene Teilnahme der Schülerinnen und Schüler an Deutsch-Förder- und Stützkursen könnte als Indikator für ein verstärktes Förderangebot und eine erhöhte Sensibilisierung an den Schulen für individuelle Förderbedarfe nach dem schlechten Abschneiden in PISA 2000 interpretiert werden. Ein deutlicher Zuwachs zwischen 2000 und 2009 zeigte sich bei den Maßnahmen der internen Evaluation. Nur knapp 18% der befragten Interviewpartner berichteten, dass im Jahr 2009 wie bereits im Jahr 2000 keine interne Evaluation stattfand. Die restlichen Schulen berichteten ein mindestens konstantes Ausmaß an interner Evaluation oder sogar vermehrte Aktivitäten der internen Evaluation. Die untersuchten Schulen wurden demnach (freiwillig oder aufgrund äußeren Drucks) vermehrt selbst aktiv im Bemühen, Genaueres über ihre aktuelle Situation zu erfahren. Auch bei den Maßnahmen der externen Evaluation konnte

3 Da im Falle der Merkmale der individuellen Förderung von keinem Effekt auf Leseleistung und Beziehungsqualität in 2009 ausgegangen wurde, wurde bei diesen Tests statt eines Signifikanzniveaus von 5% ein Niveau von 20% angesetzt. Die Effekte sind nach diesem Kriterium nicht signifikant ( $p > .20$ ).

eine deutliche Zunahme im zurückliegenden Jahrzehnt festgestellt werden. Allerdings gaben im Vergleich zur internen Evaluation mehr – knapp ein Viertel – der Schulen an, dass Maßnahmen der externen Evaluation weiterhin fehlten. Dieser Befund überrascht, da man erwarten kann, dass an allen Schulen externe Evaluation stattgefunden hat, z.B. in Form von Vergleichsarbeiten und Schulinspektionen. Ein Grund hierfür könnte darin liegen, dass die Schulinspektion an diesen Schulen tatsächlich noch nicht stattgefunden hat; eine weitere Erklärung für diese Diskrepanz könnte die Wahrnehmung des Konzepts der Evaluation durch die Schulleiter sein: Es wurden keine objektiven Fragen nach Vorkommen und Häufigkeit einzelner Evaluationsmaßnahmen gestellt, stattdessen wurde über eine offene Frage erfasst, welche Veränderungen in der externen Evaluation die Interviewpartner in den vergangenen zehn Jahren sehen. So wurden die Interviewpartner gefragt, ob sie Veränderungen wahrnehmen („Hat sich die Evaluationspraxis an Ihrer Schule in den letzten zehn Jahren verändert?“) und wie diese aussahen („Welche Veränderungen gab es?“). Möglicherweise gehören für viele Interviewpartner z.B. Vergleichsarbeiten nicht zum Konzept „externe Evaluation“. Dies könnte zum einen daran liegen, dass Schulen externe Evaluation oftmals mit Schulinspektion gleichsetzen, und zum anderen daran, dass von administrativer Seite gegenüber den Schulen der Unterrichtsentwicklungsaspekt von Vergleichsarbeiten stärker betont wird als der Evaluationsaspekt (Isaac, 2010; Orth, 2002; Peek, 2009; Petilliot-Becker & Müller-Rosigkeit, 2010; Ramm, 2007), so dass die Wahrnehmung der Schulleiterinnen und Schulleiter von Vergleichsarbeiten als Instrument der externen Evaluation ggfs. nicht gegeben ist.

Das mittlere Leistungsniveau der im Schulpanel untersuchten Gymnasien ist von 2000 zu 2009 konstant geblieben. Dieser Befund ist konform mit den Ergebnissen von PISA 2009 (Naumann et al., 2010), wonach sich die Leseleistung der Schülerinnen und Schüler im oberen Kompetenzbereich zwischen PISA 2000 und PISA 2009 nicht signifikant verändert hat. Die von den Schülerinnen und Schülern wahrgenommene Qualität der Beziehung zu ihren Lehrerinnen und Lehrern verbesserte sich hingegen über den Zeitraum von neun Jahren. Hiermit übereinstimmend berichteten die Schülerinnen und Schüler in PISA 2009 eine insgesamt positivere Lehrer-Schüler-Beziehung als in PISA 2000 (Hertel et al., 2010).

Zweites Ziel der Untersuchung war die Bestimmung der Stabilität der Schulmerkmale. Die Analysen auf der Grundlage von aggregierten Daten und die Mehrebenenanalysen führten zu teilweise deutlich unterschiedlichen Ergebnissen. Diese Disparitäten sind primär auf die geringen Varianzanteile auf Schulebene (ICC) und die daraus resultierende geringe Reliabilität der auf Schulebene aggregierten Schülerangaben zurückzuführen (vgl. z.B. Snijders & Bosker, 1999). Infolge dieser geringeren Reliabilität kommt es in den Analysen mit aggregierten Daten zu einer Unterschätzung der Zusammenhänge auf der Schulebene. Die Mehrebenenanalysen dagegen ermöglichen es, die Zusammenhänge basierend auf den tatsächlichen Unterschieden zwischen den Schulen zu schätzen. Die Gegenüberstellung verdeutlicht den Vorteil der mehrebenenanalytischen Vorgehensweise, wenn schulische Merkmale längsschnittlich untersucht werden sollen.

Die Befunde aus den Mehrebenenanalysen weisen bei signifikantem Ergebnis auf eine hohe Stabilität der betrachteten Merkmale hin. Dies gilt insbesondere für die sozi-

ale Zusammensetzung und den Migrantenanteil. Immerhin knapp 10% der im Jahr 2009 beobachteten Varianz des sozialen Status und des Migrationsstatus liegen zwischen den Schulen, und dieser Anteil wird durch die bereits im Jahr 2000 erhobenen Werte des Merkmals nahezu perfekt erklärt. Gymnasien unterscheiden sich also durchaus hinsichtlich der sozialen Selektion und Selbst-Selektion (durch die Eltern), und diese Unterschiede sind außerordentlich stabil. Die Varianz des „Output“, und zwar sowohl der Leseleistung als auch der wahrgenommenen Beziehungsqualität, lag in 2009 hingegen nur zu 4% auf Schulebene. Hinsichtlich der kognitiven wie der affektiven Kriterien sind sich Gymnasien also recht ähnlich. Die verbleibenden Unterschiede auf Schulebene sind im kognitiven Bereich recht stabil, im affektiven Bereich jedoch in hohem Maße instabil, was auf eine erhebliche Abhängigkeit des Schulklimas von den beteiligten Personen, z.B. von der zeitpunktspezifischen Zusammensetzung der Schülerschaft und des Kollegiums, hindeuten könnte. Die geringe Stabilität im Niveau des extracurricularen Angebots der Schulen dagegen könnte zum Teil auch auf die Unterschiedlichkeit der in 2000 und 2009 verwendeten Indizes zurückzuführen sein (siehe Abschnitt 4.2, Individuelle Förderung). Die Schulorganisation (ohne/mit Ganztagsangebot) und die Teilnahme an Förderkursen in Deutsch sind ebenfalls eher instabil. Die mangelnde Stabilität hängt sicherlich mit dem starken Ausbau von Schulen zu Ganztagschulen innerhalb des letzten Jahrzehnts zusammen. Zudem hängt die Teilnahme der Schülerinnen und Schüler an individuellen Förderkursen neben dem Bedarf der Schülerinnen und Schüler an solchen Kursen mit dem von der Schule gestellten Angebot zusammen, welches hier nicht berücksichtigt werden konnte.

Drittes Ziel unseres Beitrags war die Prädiktion der Veränderung in Leseleistung und Beziehungsqualität zwischen 2000 und 2009. Hier ließen sich einzelne Zusammenhänge mit den herangezogenen Schulqualitätsmerkmalen nachweisen. Zum einen zeigt sich der theoretisch plausible negative Zusammenhang zwischen der Teilnahme an Deutsch-Förderkursen und der Leseleistung innerhalb des Erhebungsjahres 2000. Dieser Befund kann wohl nur im Sinne eines rekursiven Effekts interpretiert werden: Schulen mit vergleichsweise niedrigem Leistungsstand im Lesen reagieren auf diese Problematik mit einem Angebot an Förderkursen. Im Längsschnitt dagegen besteht, wie von uns erwartet, kein Zusammenhang, vermutlich, da es sich bei der Teilnahme an Förderkursen primär um eine Maßnahme der individuellen Förderung der Schülerin/des Schülers und nicht um eine Maßnahme zur Entwicklung der „Organisation Schule“ handelt.

Weitere querschnittliche Zusammenhänge im Erhebungsjahr 2000 finden sich erwartungsgemäß zwischen Leistung und sozio-ökonomischem Status (vgl. Ehmke & Jude, 2010) sowie zwischen Beziehungsqualität und Ganztagsbetrieb (vgl. die StEG-Resultate; Klieme, Fischer et al., 2010). Schülerinnen und Schüler an ganztägig arbeitenden Schulen beschreiben die Beziehung zu ihren Lehrkräften im Mittel positiver als Schülerinnen und Schüler an Schulen ohne Ganztagsangebot. Dieser positive Effekt eines Ganztagsangebots auf die wahrgenommene Beziehungsqualität zeigt sich auch im Längsschnitt: Auch unter Kontrolle der wahrgenommenen Beziehungsqualität in 2000 findet sich ein positiver Effekt des Ganztagsangebots 2000 auf die wahrgenommene Beziehungsqualität in 2009. Bedeutsam für die Entwicklung sowohl der Lese-

leistung als auch der Beziehungsqualität ist nach unseren Befunden schließlich nicht die externe Evaluationspraxis, wohl aber die Nutzung interner Evaluation. Schulen, die bereits 2000 interne Evaluation betrieben und dieses Ausmaß an Evaluation beibehalten oder bis 2009 verstärkt haben, zeigen eine höhere Leseleistung und eine günstigere Beziehungsqualität. Eine mögliche Erklärung für diese Konstellation ist die stärkere Ausrichtung der Maßnahmen interner Evaluation auf die konkrete Situation an der Schule als bei externer Evaluation. Während Bildungsökonominnen, ausgerichtet an Wettbewerbsmodellen, externe Evaluation für wirksam erachten (Hanushek & Wößmann, 2010), sehen wir die Auswirkungen primär bei der internen Evaluation, weil diese infolge der Durchführung durch die Schulen selbst eine Beteiligung in der Schule voraussetzt und möglicherweise Prozesse und Aktivitäten in Gang setzt, die bei externer Evaluation nicht zwingend stattfinden.

Die untersuchten Prädiktoren wurden in Abschnitt 1.2 bzw. 1.4 mit der in der Schulqualitätsforschung üblichen Unterscheidung zwischen proximalen Merkmalen und distalen Merkmalen in Beziehung gebracht. In der vorliegenden Studie zeigte sich, dass die individuelle Förderung als proximales Merkmal des Lehr-Lerngeschehens von nur geringer, die distalen Merkmale (ganztägige Schulorganisation und interne Evaluationspraxis) dagegen von stärkerer Bedeutung für die Entwicklung der Schulen sind, insbesondere für das herangezogene affektive Kriterium. Von diesen Befunden sollte jedoch nicht auf die allgemeine Bedeutung proximaler vs. distaler Merkmale für die Schulentwicklung geschlossen werden, da der Ausschnitt der untersuchten Merkmale sehr klein ist. Dennoch legen die Ergebnisse nahe, dass eine weitere Untersuchung der relativen Bedeutung verschiedener Faktoren auf unterschiedlichen Ebenen für die Schulentwicklung unerlässlich ist.

Einschränkend ist anzumerken, dass in den bisherigen Analysen nur bivariate Zusammenhänge untersucht wurden, während multivariate Modelle in diesem Beitrag keinen Platz gefunden haben. Vertiefende Analysen unter Berücksichtigung z.B. der Schülerkomposition sind für weitere Publikationen vorgesehen.

Abschließend kann man festhalten, dass ein Zeitraum von neun Jahren, der für ein ganzes Schulsystem gerade ausreichend sein mag, um sich merklich zu verbessern (vgl. Klieme, Jude et al., 2010, S. 280), für Einzelschulen einen mehr als ausreichenden Erwartungshorizont bietet. Bryk et al. (2010) beispielsweise fanden messbare Entwicklungsunterschiede zwischen Schulen nach etwa 6 Jahren der Schulreform. Andererseits können in einem Zeitraum von fast einem Jahrzehnt viele Ereignisse die Entwicklung der Schulen beeinflusst haben, und die Annahme einer linearen Beziehung zwischen schulischen Bedingungsfaktoren und Ergebnisvariablen wird der Dynamik der Veränderung von Schulen möglicherweise nicht gerecht. Eine Möglichkeit, das Auftreten von für die Entwicklung von Schulen relevanten Ereignissen zu prüfen, liegt in einer weiteren Analyse der Interviewdaten, für die bewusst auf den Zeitraum zwischen der Erhebung 2000 und 2009 fokussiert wurde. Hier können Analysen von Einzelfällen mit besonders auffälligen Veränderungen weitere Erkenntnisse bringen. Darüber hinaus sollen in weiteren Publikationen neben den wenigen hier einbezogenen Schulmerkmalen auch

deren Qualität sowie weitere, durch Schulleiter-, Eltern- und Schülerbefragungen erhobene Schulmerkmale hinzugenommen werden.

Das Konzept eines an große Schulleistungsstudien wie PISA angehängten Schulpanels hat sich mit der vorliegenden Auswertung als fruchtbar erwiesen. Eine wichtige Chance für die Schulforschung wäre es, ein solches Design auch international vergleichend durchführen zu können, wie es bereits für PISA 2012 vorgeschlagen wurde (vgl. Klieme, Backhoff et al., 2010, S. 40).

## Literatur

- Altrichter, H., & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In Dies. (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 15-39). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Arbeitsstab Forum Bildung (Hrsg.) (2001). *Empfehlungen und Einzelergebnisse des Forum Bildung*. Bonn: Forum Bildung (Eigendruck).
- 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*. Münster: Waxmann.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (Hrsg.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., & Weiß, M. (Hrsg.) (2003). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Berkemeyer, N., & Müller, S. (2010). Schulinterne Evaluation – nur ein Instrument zur Selbststeuerung von Schulen? In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 195-218). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bohl, T. (2009). Theorien und Konzepte der Schulentwicklung. In S. Blömeke, T. Bohl, L. Haag, G. Lang-Wojtasik & W. Sacher (Hrsg.), *Handbuch Schule* (S. 553-559). Bad Heilbrunn: Klinkhardt/UTB.
- Boller, S. (2009). *Kooperation in der Schulentwicklung. Interdisziplinäre Zusammenarbeit in Evaluationsprojekten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brophy, J. (2000). *Teaching*. Brüssel/Genf: International Academy of Education/International Bureau of Education.
- Bryk, T., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. C. (2010). *Organizing schools for improvement. Lessons from Chicago*. Chicago: The University of Chicago Press.
- Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement*, 7(3), 197-228.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- de Maeyer, S., van den Bergh, H., Rymenans, R., van Petegem, P., & Rijlaarsdam, G. (2010). Effectiveness criteria in school effectiveness studies: Further research on the choice for a multivariate model. *Educational Research Review*, 5, 81-96.
- Diedrich, M. (2008). *Demokratische Schulkultur*. Münster: Waxmann.
- Ditton, H. (2000a). Qualitätskontrolle und -sicherung in Schule und Unterricht – ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik*, 41. Beiheft, 73-92.
- Ditton, H. (2000b). Elemente eines Systems der Qualitätssicherung im schulischen Bereich. In H. Weishaupt (Hrsg.), *Qualitätssicherung im Bildungswesen: Problemlage und aktuelle For-*

- schungsbefunde. Erfurter Studien zur Entwicklung des Bildungswesens* (S. 13-35). Erfurt: Pädagogische Hochschule.
- Ditton, H. (2007). Schulqualität. Modelle zwischen Konstruktion, empirischen Befunden und Implementierung. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 83-92). Frankfurt a.M.: Peter Lang.
- Eder, F. (2002). Unterrichtsklima und Unterrichtsqualität. *Unterrichtswissenschaft*, 30, 213-228.
- Eder, F. (2006). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 622-631). Weinheim/Basel: Beltz.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-27.
- Ehmke, T., & Baumert, J. (2007). Soziale Herkunft und Kompetenzerwerb: Vergleiche zwischen PISA 2000, 2003 und 2006. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 309-335). Münster: Waxmann.
- Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-300). Münster: Waxmann.
- Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fischer, C., Mönks, F. J., & Westphal, U. (Hrsg.) (2008). *Individuelle Förderung: Begabungen entfalten – Persönlichkeit entwickeln*. Berlin: LIT.
- Fraser, B. J. (1994). Research on classrooms and school climate. In D. Gabel (Hrsg.), *Handbook of research on science teaching and learning* (S. 493-541). New York: MacMillan.
- GANZEBOOM, H. B. G., & TREIMAN, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25, 201-239.
- Gerecht, M. (2010). *Schul- und Unterrichtsqualität und ihre erzieherischen Wirkungen*. Münster: Waxmann.
- Götz, T., Frenzel, A. C., & Pekrun, R. (2008). Sozialklima in der Schule. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Psychologie. Band Pädagogische Psychologie* (S. 503-514). Göttingen: Hogrefe.
- Hanushek, E. A., & Wößmann, L. (2010). *The economics of international differences in educational achievement* (NBER Working Paper No. 15949). <http://www.nber.org/papers/w15949.pdf> [16.11.2012].
- Hertel, S., Hochweber, J., Steinert, B., & Klieme, E. (2010). Schulische Rahmenbedingungen und Lerngelegenheiten im Deutschunterricht. 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. 113-151). Münster: Waxmann.
- Holtappels, H. G. (2003). *Schulqualität durch Schulentwicklung und Evaluation. Konzepte, Forschungsbefunde, Instrumente*. München: Luchterhand.
- Holtappels, H. G. (2006). Stichwort: Ganztagschule. *Zeitschrift für Erziehungswissenschaft*, 9(1), 5-29.
- Holtappels, H. G., Klemm, K., & Rolff, H.-G. (Hrsg.) (2008). *Schulentwicklung durch Gestaltungsautonomie. Ergebnisse der Begleitforschung zum Modellvorhaben ‚Selbstständige Schule‘ in Nordrhein-Westfalen*. Münster: Waxmann.
- Isaac, K. (2010). Chancen und Grenzen von Vergleichsarbeiten – Zur schulinternen Nutzung von Daten. *Schulverwaltung NRW*, 21(7/8), 197-199.
- Kaplan, D. (2008). *Structural equation modeling*. Thousand Oaks: Sage.

- Klemm, K. (2011). Das Bildungssystem Deutschlands. Strukturen und Strukturreformen. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (S. 153-164). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E. (2005). Zur Bedeutung von Evaluation für die Schulentwicklung. In K. Maag Merki, A. Sandmeier, P. Schuler & H. Fend (Hrsg.), *Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert* (S. 40-61). Zürich: Pädagogisches Institut der Universität Zürich.
- Klieme, E., Backhoff, E., Blum, W., Buckley, J., Hong, Y., Kaplan, D., Levin, H., Scheerens, J., Schmidt, W., van de Vijver, F., & Vieluf, S. (2010). *Designing PISA as a sustainable database for educational policy and research: The PISA 2012 Context Questionnaire Framework*. Paris: OECD.
- Klieme, E., Fischer, N., Holtappels, H. G., Rauschenbach, T., & Stecher, L. (2010). *Ganztagschule: Entwicklung und Wirkungen. Ergebnisse der Studie zur Entwicklung von Ganztagschulen 2005-2010*. [http://projekt-steg.de/files/pk101111/Ergebnisbroschuere\\_StEG\\_2010-11-11.pdf](http://projekt-steg.de/files/pk101111/Ergebnisbroschuere_StEG_2010-11-11.pdf) [16.11.2012].
- Klieme, E., Jude, N., Baumert, J., & Prenzel, M. (2010). PISA 2000-2009: Bilanz der Veränderungen im Schulsystem. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-300). Münster: Waxmann.
- Klieme, E., & Steinert, B. (2008). Schulentwicklung im Längsschnitt. In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006* (S. 221-238). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E., Steinert, B., & Hochweber, J. (2010). Zur Bedeutung der Schulqualität für Unterricht und Lernergebnisse. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 231-255). Münster: Waxmann.
- Klieme, E., & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, 57(6), 805-818.
- KMK = Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2011). *Definitionenkatalog zur Schulstatistik 2011 (2)*. <http://www.kmk.org/fileadmin/pdf/Statistik/Defkat2011.pdf> [04.12.2012].
- Maag Merki, K. (2009). Evaluation im Bildungsbereich Schule in Deutschland. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch* (S. 157-162). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2010). *Die Initiative „Gütesiegel Individuelle Förderung“: Ein Leitfaden für Schulen*. Neuss: Meinke Print Media Partner.
- Mortimore, P., Sammons, P., & Thomas, S. (1994). School effectiveness and value added measures. *Assessment in Education: Principles, Policy & Practice*, 1(3), 315-332.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- OECD (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. Paris: OECD.
- Opendakker, M.-C., & van Damme, J. (2000). Effects of schools, teaching and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement*, 11(2), 165-196.
- Orth, G. (2002). Vergleichsarbeiten. In H.-G. Rolff & H.-J. Schmidt (Hrsg.), *Brennpunkt Schulleitung und Schulaufsicht* (S. 203-223). Neuwied: Luchterhand.
- Peek, R. (2009). Dateninduzierte Schulentwicklung. In H. Buchen & H.-G. Rolff (Hrsg.), *Professionswissen Schulleitung* (S. 1343-1367). Weinheim/Basel: Beltz.

- Petilliot-Becker, I., & Müller-Rosigkeit, E. (2010). Diagnose- und Vergleichsarbeiten – Zielsetzungen und Konzeption. *Schulverwaltung Baden-Württemberg*, 19(6), 122-124.
- Rabenstein, K. (2009). Individuelle Förderung in unterrichtsergänzenden Angeboten an Ganztagschulen – ein Fallvergleich. In S. Appel, H. Ludwig & U. Rother (Hrsg.), *Jahrbuch Ganztagschule 2010. Vielseitig fördern* (S. 23-33). Schwalbach am Taunus: Wochenschau-Verlag.
- Rahm, S. (2005). *Einführung in die Theorie der Schulentwicklung*. Weinheim/Basel: Beltz.
- Ramm, G. (2007). VERA kommt in die Sekundarstufe. *Schul-Management*, 4, 20-23.
- Rolff, H.-G. (1998). Entwicklung von Einzelschulen: Viel Praxis, wenig Theorie und kaum Forschung. Ein Versuch, Schulentwicklung zu systematisieren. In H.-G. Rolff, K.-O. Bauer, K. Klemm, H. Pfeiffer (Hrsg.), *Jahrbuch der Schulentwicklung*, Bd. 10 (S. 295-325). Weinheim: Juventa.
- Rolff, H.-G. (2010). Schulentwicklung als Trias von Organisations-, Unterrichts- und Personalentwicklung. In T. Bohl, W. Helsper, H. G. Holtappels & C. Schelle (Hrsg.), *Handbuch Schulentwicklung* (S. 29-36). Bad Heilbrunn: Klinkhardt/UTB.
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: a review of school effectiveness research. A report by the Institute of Education for the Office for Standards in Education*. London: Institute of Education.
- Scheerens, J. (2008). *Review and meta-analyses of school and teaching effectiveness*. Twente: University of Twente, Department of Educational Organization and Management.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon Press.
- Schwippert, K., Hornberg, S., Freiberg, M., & Stubbe, T. C. (2007). Lesekompetenzen von Kindern mit Migrationshintergrund im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 249-270). Münster: Waxmann.
- Seidel, T. (2008). Stichwort: Schuleffektivitätskriterien in der internationalen empirischen Forschung. *Zeitschrift für Erziehungswissenschaft*, 11(3), 348-367.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis*. London: Sage.
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. 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. 200-230). Münster: Waxmann.
- Steinhage, N., & Blossfeld, H.-P. (1999). *Zur Problematik von Querschnittsdaten*. <http://nbn-resolving.de/urn:nbn:de:0168-ssoa-57590> [04.12.2012].
- Teddlie, C., & Reynolds, D. (Hrsg.) (2000). *The international handbook of school effectiveness research*. London: Falmer Press.
- Trautwein, U., Neumann, M., Nagy, G., Lüdtke, O., & Maaz, K. (Hrsg.) (2010). *Schulleistungen von Abiturienten: Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- van Ackeren, I., & Brauckmann, S. (2010). Internationale Diskussions-, Forschungs- und Theorieansätze zur Governance im Schulwesen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 41-61). Wiesbaden: VS Verlag für Sozialwissenschaften.
- van Damme, J., de Fraine, B., van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2002). A new study on educational effectiveness in secondary schools in Flanders: An introduction. *School Effectiveness and School Improvement*, 13(4), 383-397.
- van Damme, J., Opdenakker, M.-C., van Landeghem, G., de Fraine, B., Pustjens, H., & van de Gaer, E. (2006). *Educational effectiveness: an introduction to international and Flemish research on schools, teachers and classes*. Leuven: Acco.



Walter, O., & Taskinen, P. (2007). Kompetenzen und bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland: Ein Vergleich mit ausgewählten OECD-Staaten. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 337-366). Münster: Waxmann.

**Anschrift der Autorin/der Autoren**

Dipl. Psych. Linda Marie Bischof, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland  
E-Mail: bischof@dipf.de

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

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

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