

## **Untersuchung von Schulentwicklung anhand von Schuleffektivitätsdaten – „Yes, we can“?!**

### **1. Einleitung<sup>1</sup>**

Der vorliegende Beitrag befasst sich mit der Frage, ob und wie durch Erweiterungen des Designs von Schuleffektivitätsstudien (wie z.B. PISA) Erkenntnisse für Schulentwicklung generiert werden können. Er zeigt die Möglichkeiten, aber auch die Grenzen einer empirischen Verknüpfung von Schuleffektivität und Schulentwicklung auf. Darüber hinaus beschäftigt der Beitrag sich mit der Frage der theoretischen Verknüpfung beider Forschungsansätze. Er behandelt mögliche Gründe für die Schwierigkeit, Effektivität und Entwicklung auch auf theoretischer Ebene zu verknüpfen.

Effektivität und Entwicklung sind zwei zentrale Aspekte von Schule. In der Forschung werden diese Aspekte bisher eher unabhängig voneinander von Mitgliedern unterschiedlicher Forschungsgemeinschaften untersucht. Beide Forschungsrichtungen verfolgen das Ziel, Schulen zu verbessern. Dies erfolgt jedoch aus unterschiedlicher Perspektive und basiert auf unterschiedlichen Methoden, theoretischen Grundlagen und Modellen. Zudem wird bei der genaueren Betrachtung der Ziele ersichtlich, dass die Indikatoren für die Verbesserung von Schule auf unterschiedlichen Ebenen liegen. Im Fokus der Schuleffektivitätsforschung steht die Verbesserung von Schülerergebnissen (Reynolds, Sammons, De Fraine, Townsend & Van Damme, 2011; Seidel, 2008), während der Fokus in der Schulentwicklungsforschung eher auf der Verbesserung der Selbststeuerungsfähigkeit der Schulen und ihrer Entwicklungsprozesse als auf der Verbesserung der Schülerergebnisse liegt (Hopkins, 1996; Reynolds, Hopkins & Stoll, 1993; Stoll, Creemers & Reezigt, 2006).

Die jeweilige Fokussierung (vgl. Abschnitt 2) versuchen Forscherinnen und Forscher seit geraumer Zeit durch eine Verknüpfung von Schuleffektivität und Schulentwicklung aufzulösen, wobei „Verknüpfung“ teilweise unterschiedlich verstanden wird. Einige verstehen die Verknüpfung als die Schaffung einer gemeinsamen Wissensbasis der beiden Forschungsrichtungen (z.B. Harris, 2001), andere verstehen sie als die Schaffung einer Wissensbasis für die Schulentwicklungspraxis, die auf Erkenntnissen der Schuleffektivitätsforschung beruht (z.B. Creemers & Kyriakides, 2008). Unabhängig von dem zugrundeliegenden Verständnis wird davon ausgegangen, dass eine

---

<sup>1</sup> Der vorliegende Beitrag basiert auf Arbeiten im Rahmen des PISA-Schulpanels und einer Promotion am Deutschen Institut für Internationale Pädagogische Forschung (Bischof, Hochweber, Hartig & Klieme, 2013; Bischof, 2017).

Verknüpfung notwendig ist, um die innerhalb des Feldes der Schuleffektivität bzw. innerhalb des Feldes der Schulentwicklung bestehende Fokussierung/Engführung aufzulösen und zu einem umfassenden Bild von Schule zu gelangen (vgl. Abschnitt 3).

Die Diskussion und Versuche einer Verknüpfung fanden vor allem in den 1990er-Jahren und Anfang des 21. Jahrhunderts statt. Deutlich wird dies beispielsweise an Arbeiten wie den Sammelbänden von Gray, Reynolds, Fitz-Gibbon und Jesson (1996), Harris und Bennett (2001), Hoeben (1998), Reynolds, Bollen, Creemers, Hopkins, Stoll und Lagerweij (1996), Stoll und Fink (1996), aber auch in Zeitschriftenbeiträgen und Büchern wie von Creemers und Kyriakides (2008), Creemers & Kyriakides (2010), Creemers und Reezigt (1997), Hopkins (1990), Reezigt, (2001), Reezigt und Creemers (2005), Reynolds, Hopkins, und Stoll (1993) und Reynolds, Teddlie, Hopkins, und Stringfield (2000). Auch strukturelle Veränderungen wie die Umbenennung des International Congress for School Effectiveness, welcher sich bereits bei seinem Beginn im Jahr 1988 mit der Frage der Verknüpfung von Schuleffektivität und Schulentwicklung befasste, in den International Congress for School Effectiveness and Improvement (vgl. <http://www.icsei.net/index.php?id=1714>) weist auf deren Bedeutung hin.

Der vorliegende Beitrag diskutiert, inwiefern Schulentwicklungsfragen anhand von Daten, die im Kontext der Schuleffektivitätsforschung gewonnen wurden, untersucht werden können, genauer gesagt, inwiefern durch spezifische Ergänzung von Designs Aussagen zur Schulentwicklung möglich sind.

Wir stellen zunächst die zentralen Merkmale der Schuleffektivitäts- und Schulentwicklungsforschung sowie die Diskussion über ihre (theoretische) Verknüpfung dar. Anschließend stellen wir Analysen und Ergebnisse aus dem PISA-Schulpanel vor. Das PISA-Schulpanel ist eine Erweiterung des regulären PISA-Designs in Deutschland anhand dessen die Entwicklung von Schulen zwischen 2000 und 2009 in den Blick genommen wurde (empirische Verknüpfung).

## 2. Schuleffektivität und Schulentwicklung und ihre Erforschung

### 2.1 Schuleffektivität und Schuleffektivitätsforschung

Die Effektivität einer Schule wird gemessen an den Ergebnissen ihrer Schülerinnen und Schüler. Die Kriterien, die für die Untersuchung der Effektivität herangezogen werden, werden von Seidel (2008) unterschieden nach kognitiven, metakognitiven, motivational-affektiven und verhaltensbezogenen Kriterien. Systematisch und am häufigsten werden zur Untersuchung der Effektivität kognitive Kriterien herangezogen, und zwar Leistungen und Kompetenzen (Diedrich, 2008; Reynolds et al., 2011; Seidel, 2008; Van Damme et al., 2006).

Es gibt zwei zentrale Ansätze zur Bestimmung der Effektivität einer Schule. Nach dem *quality*-Ansatz ist eine Schule dann effektiv, wenn ein starker Zuwachs in den Ergebnissen der Schülerinnen und Schüler in den gewählten Kriterien erkennbar ist. Nach dem *equity*-Ansatz ist eine Schule dann effektiv, wenn die Schülerinnen und

Schüler (unabhängig von ihrem Hintergrund und ihren Lernvoraussetzungen) ähnliche Ergebnisse erzielen, sich die Ergebnisse der Schülerinnen und Schüler also annähern.

Neben der Untersuchung der Effektivität von Schule anhand der Ergebnisse der Schülerinnen und Schüler ist die Untersuchung des Zusammenhangs dieser Effektivität mit schulischen Merkmalen zentraler Untersuchungsgegenstand der Schuleffektivitätsforschung. Von Interesse ist, warum eine Schule eine andere Effektivität aufweist als eine andere Schule mit einer ähnlichen Schülerkomposition (Scheerens, 2000). Über mehrere Studien hinweg wurden auf diese Weise effektivitätssteigernde Faktoren identifiziert, die in Listen kategorisiert wurden (z.B. hohe Leistungserwartungen, Führungsqualität der Schulleitung, Schulklima, Evaluation, Konsens und Kooperation der Lehrkräfte, siehe z.B. Creemers & Reezigt, 1996; Scheerens & Bosker, 1997; Scheerens, Luyten, Steen & Luyten-de Thouars, 2007).

Die Methoden der Schuleffektivitätsforschung sind vorrangig quantitativer Natur. Der Effekt von Faktoren auf Schülerergebnisse wird meist über *value-added*-Ansätze untersucht. Bei solchen Ansätzen wird der Effekt eines Faktors unter Kontrolle von weiteren Merkmalen, die die Ergebnisse der Schülerinnen und Schüler (potenziell) beeinflussen (z.B. Lernvoraussetzungen des Schülers/der Schülerin und familiärer Hintergrund; Downes & Vindurampulle, 2007), analysiert. Ziel ist eine „faire“ Bestimmung des Anteils der Schule am Schülerergebnis, d.h. die Bestimmung des „Mehrerts“, den eine Schule zum Schülerergebnis beiträgt.

## 2.2 Schulentwicklung und Schulentwicklungsforschung

Schulentwicklung ist ein kontinuierlicher Prozess. Rolff (2007) schreibt: „Schulentwicklung geht von der Einzelschule als Einheit zielgerichteter systematischer und reflexiver Gestaltung aus, wobei UE [Unterrichtsentwicklung], OE [Organisationsentwicklung] und PE [Personalentwicklung] im Systemzusammenhang gesehen werden und die Entwicklung von Einzelschulen mit der Entwicklung des gesamten Schulsystems gekoppelt ist“ (ebd., S. 29). Die Einzelschule ist dabei der „Motor der Entwicklung“. Diese liegt somit im Verantwortungsbereich der Schule selbst und ist ein komplexer Prozess, der sich aus unterschiedlichen Komponenten zusammensetzt. Ziel von Schulentwicklungsprozessen im Verständnis der Schulentwicklungsforschung ist die Verbesserung der Selbststeuerungsfähigkeit der Schule und der Schulentwicklungsprozesse. Feldhoff, Bischof, Emmerich und Radisch (2015) formulieren den Fokus der Schulentwicklungsforschung wie folgt: „Forschungsleitend ist hierbei die Frage, wie eine nachhaltige Verbesserung der pädagogischen Interaktionspraxis durch Veränderungen von Handlungsstrukturen auf Organisationsebene der Schule erreicht werden kann. Forschungsziel ist entsprechend primär die Konzeption von Schulentwicklungsprogrammen und einzelnen Entwicklungsmaßnahmen bzw. die methodisch valide Evaluierung von Programmen und Maßnahmen sowie der hieran beteiligten Prozesse, die zum großen Teil steuerungspolitisch intendiert sind“ (ebd., S. 68).

Schulentwicklungsforschung wählt somit eine andere Perspektive auf Schule als Schuleffektivitätsforschung und sucht nach Faktoren, die positiv mit wünschenswerten Schulentwicklungsprozessen bzw. der Veränderungsfähigkeit von Schulen zusammenhängen, welche entsprechend die zentralen Ergebnisvariablen der Schulentwicklungsforschung darstellen. Ähnlich wie in der Schuleffektivitätsforschung wurden auch in der Schulentwicklungsforschung Listen von Faktoren erstellt (z.B. Schulleitungshandeln, Schulethos, Verantwortung der Schülerinnen und Schüler für ihr Lernen, Lehrerinnen und Lehrer als Lernerinnen und Lerner, Fokus auf die Qualität des Lernens und Lehrens an der Schule, Kooperation der Lehrerschaft, (systematische) Planung der Schulentwicklung; siehe z.B. Merrett, 2000; Stoll & Mortimore, 1997).

Noch in den 1990er-Jahren wurde die Schulentwicklungsforschung als eine Forschungsrichtung beschrieben, in der qualitative Methoden verwendet werden (z.B. Brown, Riddell & Duffield, 1996; Reynolds & Stoll, 1996). Die heutige Schulentwicklungsforschung bedient sich stärker auch quantitativer Methoden. Viele Erkenntnisse basieren jedoch weiterhin vor allem auf qualitativen Untersuchungen (vgl. Thillmann, 2012).

### 3. Verknüpfung von Schuleffektivität und Schulentwicklung

Die in Abschnitt 2 deutlich werdende Fokussierung der Forschungsrichtungen (z.B. hinsichtlich ihres Erkenntnisinteresses, ihrer Zielvariablen und ihrer Methoden) hat zur Folge, dass bestimmte Bereiche von Schule in den Blick genommen werden und andere Bereiche nicht. Die seit geraumer Zeit diskutierte Idee einer Verknüpfung der Ansätze und damit der Weitung der Perspektive (vgl. z.B. Brown et al., 1996; Clark, Lotto & Astuto, 1984; Creemers & Kyriakides, 2008; Creemers & Reynolds, 1989) ist durch die Vorstellung begründet, dadurch zu einer besseren theoretischen Beschreibung und effektiveren empirischen Untersuchung und damit zu einem besseren Verständnis von Schule und Schulentwicklung zu gelangen als es durch die einzelne Forschungsrichtung möglich ist (vgl. Bischof, 2017; Bensen, Büchter & van Ophuysen, 2004; Brown et al., 1996; Creemers, 2007; Gray et al., 1996; Harris, 2001; Reynolds et al., 1996; Stoll, 1994; Teddlie & Stringfield, 2007).

In diesem Kontext sind verschiedene Modelle entwickelt worden, wie eine Verknüpfung von Schuleffektivität und Schulentwicklung gelingen kann. Der Fokus der Modelle liegt entweder auf der Verknüpfung der beiden Forschungsrichtungen („Model Linking School Effectiveness and School Improvement“, Stoll & Fink, 1996; „Comprehensive Framework of Effective School Improvement“, z.B. Stoll et al., 2006) oder auf der Verknüpfung von Schuleffektivitätsforschung und Schulentwicklungspraxis; also auf der Übertragung von Erkenntnissen der Schuleffektivitätsforschung in die Praxis („Dynamic Model of Educational Effectiveness“ zusammen mit dem „Dynamic Approach to School Improvement“, z.B. Creemers & Kyriakides, 2008, 2012). In diesem Beitrag wird eine Verknüpfung von Schuleffektivitäts- und Schulentwicklungsforschung auf empirischer Ebene versucht. Der angesprochene Vorteil einer theoretischen und empirischen Verknüpfung soll hier kurz für eine Änderung des Designs

von Schuleffektivitätsstudien mit dem Ziel einer Öffnung für Schulentwicklungsfragestellungen erläutert werden.

Typische Effektivitätsstudien versuchen spezifische Fragestellungen zu beantworten wie „what makes a school effective?“. Dadurch nehmen sie einen bestimmten Bereich von Schule in den Blick, der anhand gängiger Methoden der Forschungsrichtung untersucht wird (z.B. quantitative Querschnittstudie); die Komplexität von Schule als Ganzes wird bewusst auf einzelne interessierende Bestandteile reduziert. Eine empirische Verknüpfung beider Ansätze hat nun den Vorteil, dass Konzepte beider Forschungsrichtungen berücksichtigt („What makes a school effective?“, „How can schools improve?“) und unterschiedliche Methoden (qualitative und quantitative Methoden in quer- und längsschnittlichem Design) herangezogen werden können, um einen umfassenderen Blick auf den Forschungsgegenstand Schule zu gewinnen. Diese Weitung der Perspektive kann, wie in der vorliegenden Untersuchung beispielhaft umgesetzt, durch eine nachträgliche Veränderung des Designs von Studien aus einer der beiden Forschungstraditionen erreicht werden. Im besten Fall wird sie jedoch bereits bei der Anlage der Studie mitgedacht. Die exemplarische Umsetzung im Rahmen des PISA-Schulpanels wird in Abschnitt 4.2 aufgegriffen.

Auf der konzeptuellen Ebene lässt sich für die bestehenden Modelle ein grundlegendes Problem pädagogischer Kontexte identifizieren: das der schulischen Arbeit immanente Technologiedefizit, d.h. das Fehlen klarer Wenn-Dann-Logiken in der Wirkung spezifischer Schul- und Unterrichtspraktiken (vgl. Bischof, 2017; Feldhoff et al., 2015). Die Schulentwicklungsforschung nimmt v.a. die Handlungsstrukturen auf Organisationsebene der Schule in den Blick und geht davon aus, dass deren Veränderung zu einer Verbesserung der pädagogischen Praxis führt. Damit ist für sie das Technologiedefizit auf der Schulebene zentral, d.h. das Fehlen klarer Technologien zur Entwicklung von Schule. In der Schuleffektivitätsforschung, die vorrangig an der Verbesserung der Leistungen der Schülerinnen und Schüler interessiert ist, liegt das Problem vor allem auf der Unterrichtsebene, d.h. im Fehlen klarer Technologien zur Transformation von Lehr- in Lernprozesse. Bei Feldhoff et al. (2015) wird festgehalten: „Weder komplexe Interaktionsprozesse des Lehrens und Lernens, noch Handlungsprozesse auf der Ebene der Schulorganisation [laufen] unter ‚kontrollierbaren‘ Kausalitätsbedingungen ab“ (ebd., S. 66). Die Schwierigkeit, für pädagogische Settings klare Technologien anzunehmen, wird beispielsweise deutlich bei Luhmann und Schorr (1979), Baumert und Kunter (2006), Helmke (2006), Tenorth (2006) und Fend (2008a; 2008b; vgl. auch Weick, 1976; 2008).

Geht man davon aus, dass in Schulen ein Technologiedefizit auf der Ebene der Schule (Fehlen von Technologien zur Entwicklung von Schule) wie auf der Ebene des Unterrichts (Fehlen von Technologien zur Transformation von Lehr- in Lernprozesse) existiert, so legt dies den Schluss nahe, dass eine Verknüpfung der beiden behandelten Forschungsrichtungen mit ihren spezifischen Beobachtungsgegenständen (Schuleffektivitätsforschung: lernwirksame Beeinflussbarkeit intrapsychischer Lernprozesse, Schulentwicklungsforschung: Beeinflussbarkeit kollektiver Handlungsprozesse) in entsprechenden Modellen zu einer Verdoppelung der Beobachtungsprobleme führt (vgl. Feldhoff et al., 2015).

Auf empirischer Ebene lässt sich das Problem des Technologiedefizits zwar nicht grundsätzlich lösen, doch kann den theoretischen Überlegungen zumindest Rechnung getragen werden, indem empirische Ergebnisse, z.B. über Zusammenhänge schulischer Faktoren, nicht kausal gemäß einer deterministischen Wenn-Dann-Logik interpretiert werden (z.B. Lehrkooperation hat einen positiven Einfluss auf Schülerleistungen). Stattdessen kann der Zusammenhang angemessener im Sinne einer erhöhten Wahrscheinlichkeit bestimmter Veränderungen als Folge spezifischer Maßnahmen und Entwicklungsprozesse beschrieben werden (die Kooperation von Lehrkräften erhöht die Wahrscheinlichkeit einer Verbesserung der Schülerleistung). Dementsprechend können wir auch in diesem Beitrag die vorzustellenden Ergebnisse nur vorsichtig interpretieren und als Hinweise verstehen.

Die deutlich gewordenen Herausforderungen und Schwierigkeiten im Hinterkopf behaltend, wird im Folgenden beispielhaft dargestellt, welche Aussagen über Schulentwicklung anhand großer Schulleistungsstudien möglich sein können und warum es gewinnbringend sein kann, die Verknüpfung von Schulentwicklung und Schuleffektivität weiter anzustreben.

## 4. Empirische Untersuchung von Schulentwicklung anhand von Daten aus der Schuleffektivitätsforschung

### 4.1 Die PISA-Studie und das PISA-Schulpanel

Die PISA-Studie („Programme for International Student Assessment“) ist eine zyklisch angelegte internationale Schulleistungsstudie, die mit dem Ziel durchgeführt wird, die Leistungsstände von Schülerinnen und Schülern kurz vor Ende der Pflichtschulzeit zu untersuchen.

PISA untersucht Kompetenzprofile 15-jähriger Schülerinnen und Schüler und vergleicht diese über Staaten hinweg. Daneben nimmt PISA Zusammenhänge zwischen schulischen und familiären Merkmalen und der Leistungen der Schülerinnen und Schüler in den Blick (vgl. dazu Jude & Klieme, 2010). Das Design der PISA-Studie ist so gestaltet, dass in den teilnehmenden Staaten in einem dreijährigen Turnus wiederholt Erhebungen stattfinden. Die Schulen werden dabei zufällig gezogen. Auch die Schülerschaft ist aufgrund des Fokus auf Fünfzehnjährige zu jedem Erhebungszeitpunkt eine andere. Deshalb lassen sich Veränderungen auf der Systemebene – im Sinne von Unterschieden zu vorherigen Kohorten – untersuchen. Einer Untersuchung der Entwicklung einzelner Schülerinnen und Schüler oder einzelner Schulen ist dadurch allerdings nicht (oder nur unsystematisch) möglich (vgl. Bischof, Hochweber, Hartig & Klieme, 2013; Klieme, Jude, Baumert & Prenzel, 2010; Klieme & Steinert, 2008).<sup>2</sup>

2 Ähnlich zu den im Folgenden dargestellten Ergebnissen wurden im Sammelband „PISA 2009 – Impulse für die Schul- und Unterrichtsforschung“ Klieme & Jude (2013) erste Ergebnisse des PISA-Schulpanels wiedergegeben. Der folgende Beitrag ergänzt die Ergebnisse durch eine Anpassung des Analysedesigns und durch die Einbettung in den Gesamtkontext einer Verknüpfung von Schuleffektivität und Schulentwicklung.

Im vorliegenden Beitrag wird die Frage aufgegriffen, inwiefern Effektivitätsstudien wie die PISA-Studie herangezogen werden können, um Schulentwicklung empirisch zu untersuchen. Durch eine Veränderung des üblichen PISA-Designs bot das Schulpanel die Möglichkeit, einige der Punkte, die im Rahmen der Verknüpfungsdiskussion genannt werden, umzusetzen (siehe Abschnitt 4.2) und dadurch eine Beantwortung von Fragen der Schulentwicklung zu versuchen.

Im Rahmen des PISA-Schulpanels nahmen 59 Gymnasien der regulären Stichprobe aus PISA 2000 aus allen Bundesländern mit Ausnahme Bremens im Jahr 2009 erneut an einer Erhebung teil. Ergänzend dazu wurden im Jahr 2010 retrospektive Interviews mit den Schulleitungen der Schulen durchgeführt. Inhalt der Interviews waren Veränderungen an den Schulen im Zeitraum 2000 bis 2009.<sup>3</sup> Die Teilnahme am PISA-Schulpanel war freiwillig. Durch die wiederholte Messung auf Schulebene konnten Veränderungsprozesse von Merkmalen der Schule und deren Auswirkungen auf der Ebene der Schülerinnen und Schüler untersucht werden (vgl. Klieme et al., 2010).

## 4.2 Empirische Verknüpfung von Schuleffektivität und Schulentwicklung im Rahmen des Schulpanels

Grundlegende Voraussetzung für die Messung von Schulentwicklung ist die *längsschnittliche Untersuchung von Schulen*, d.h. die wiederholte Erhebung von Daten auf Schulebene zu mindestens zwei Messzeitpunkten (vgl. Gustafsson, 2010). Diese ist im PISA-Schulpanel mit Erhebungen in den Jahren 2000 und 2009 gegeben.

Sowohl die Schuleffektivitäts- als auch die Schulentwicklungsforschung wurden wiederholt dafür kritisiert, vor allem effektive Schulen bzw. Schulen mit erfolgreichen Innovationen in den Blick zu nehmen. Dieser Kritik wird im PISA-Schulpanel Rechnung getragen, indem eine *Zufallsstichprobe aus den PISA-Schulen* gezogen und um Teilnahme ersucht wurde. Allerdings war die Teilnahme am Schulpanel freiwillig.

Ein häufig genanntes Vorgehen für eine empirische Verknüpfung von Schuleffektivität und Schulentwicklung ist die Überprüfung des Erfolgs einer Schule sowohl anhand eines *Effektivitätskriteriums* (z.B. *Leistung der Schülerinnen und Schüler*) als auch eines *Entwicklungskriteriums* (z.B. *Selbststeuerungsfähigkeit der Schule*). In vorliegender Studie wurde die Leseleistung der Schülerinnen und Schüler als *Effektivitätskriterium* verwendet. Dies entspricht dem klassischen Vorgehen der Schuleffektivitätsforschung, den Fokus auf kognitive Kriterien in Kernfächern zu legen. Die von den Schülerinnen und Schülern wahrgenommene Beziehungsqualität zu ihren Lehrerinnen und Lehrern wird als *Schulentwicklungskriterium* für das Schulklima verwendet (vgl. Baumert, Stanat & Demmrich, 2001; Hertel, Hochweber, Steinert & Klieme, 2010). Das Schulklima ist ein Prozessmerkmal der Schule. Es ist „sensibel für Veränderungen im Schulsystem bzw. Entwicklungsmaßnahmen einzelner Schulen“ (2006, S. 626) und gilt als zentrale Komponente von Schulentwicklung (vgl. Dellar, 1998; Lin-

3 Hier werden nur die für diesen Beitrag relevanten Änderungen des Designs verdeutlicht. Für eine ausführlichere Beschreibung vergleiche Bischof (2017).

dahl, 2011). Angaben zu beiden Kriterien liegen jeweils zum ersten (2000) als auch zum zweiten Messzeitpunkt (2009) vor.

In diesem Zeitraum von neun Jahren waren Schulen mit vielen Reformen konfrontiert und haben sich verändert. Die Geschehnisse in diesem Zeitraum sollten ebenfalls in den Blick genommen werden. Dazu wurden die Schulleitungen in einem retrospektiven Interview zu Veränderungen an ihrer Schule befragt. Anhand dieser Daten lässt sich die *Wirkung von Entwicklungsaktivitäten auf das Effektivitäts- und das Entwicklungskriterium* untersuchen.

### 4.3 Fragestellungen und Hypothesen

Im vorliegenden Beitrag legen wir den Fokus auf die Untersuchung des Effekts von Evaluation auf die Entwicklung von Schulen. Dies wird anhand des Einflusses auf Veränderungen in den zwei oben beschriebenen Kriterien untersucht, d.h. der Lesekompetenz und der wahrgenommenen Beziehungsqualität.

Evaluationen werden unterschieden in interne Evaluationen und externe Evaluationen. Hintergrund ist der unterschiedliche Grad der Beteiligung der Schule am Evaluationsprozess. Externe Evaluationen sind Verfahren, die von Personen/Institutionen außerhalb der Schule durchgeführt werden. Interne Evaluation zeichnet sich hingegen durch die in der Schule stattfindende Auseinandersetzung mit dem Evaluationsgegenstand (Selbstreflexion; vgl. Böttcher, 2008) aus.

Hinsichtlich der Wirksamkeit von Evaluationen liegen v. a. Ergebnisse aus dem internationalen Raum vor; im deutschsprachigen Raum ist die Erkenntnislage eher schwach (vgl. Ackeren van & Bellenberg, 2004; Berkemeyer & Müller, 2010; Böhm-Kasper & Selders, 2013; Dederling & Müller, 2011a; Groß Ophoff, 2013; Thiel & Thillmann, 2012). Ergebnisse aus dem internationalen Raum sind zumindest für externe Evaluationen nicht uneingeschränkt für Deutschland generalisierbar. Grund ist die unterschiedliche konzeptionelle Gestaltung externer Evaluation. So gehen beispielsweise Schulinspektionen in den Niederlanden und England mit wesentlich stärkeren Konsequenzen und größerem Druck für Schulen einher als dies in Deutschland der Fall ist. Auch ist durch die Gestaltung der Schulinspektion der Wettbewerbscharakter wesentlich stärker ausgeprägt („Ranking“). Ähnliches gilt auch für Vergleichsarbeiten (siehe z.B. Stufflebeam, Madaus & Kellaghan, 2000).

Deutschsprachige Autorengruppen kommen zu unterschiedlichen Ergebnissen hinsichtlich der Wirksamkeit externer Evaluation. Insgesamt wird die Erkenntnislage weiterhin als gering beschrieben (siehe Gärtner, 2013; vgl. Ackeren van & Bellenberg, 2004; Böttcher & Keune, 2011; Dederling, 2012; Groß Ophoff, 2013; Kotthoff & Böttcher, 2010). Pietsch, Janke und Mohr (2014) berichten, dass im deutschsprachigen Raum bisher keine Studie die Wirkung von Schulinspektionen<sup>4</sup> in Bezug auf Schüler-

---

4 In manchen Bundesländern werden andere Begriffe für die Verfahren der Schulinspektion genutzt (externe Evaluation, Qualitätsanalyse, Schulvisitation oder Fremdevaluation). Hier bezieht sich der Begriff der Schulinspektion auf die Verfahren aller Bundesländer unabhängig von deren Bezeichnung.

leistungen untersucht. Lamprecht und Rürup (2012) weisen darauf hin, dass Studien zur Wirksamkeit von Schulinspektionen vorrangig auf Einschätzungen der Schulleitung oder der Lehrkräfte beruhen und auf die Akzeptanz als einen zentralen Aspekt für die Wirksamkeit von Schulinspektionen fokussieren. Böttcher und Keune (2011) beschreiben Schulinspektionen eher als nicht wirksam für Schulentwicklung. Wurster und Gärtner (2013) erklären unterschiedliche Effekte von Schulinspektionen anhand unterschiedlicher „Typen“ von Schulen und ihren Wahrnehmungs- und Verarbeitungsprozessen. Auch bei Böhm-Kasper und Selders (2013) wird die Uneinheitlichkeit der Ergebnisse herausgearbeitet. Sie stellen in ihrem Vergleich zwischen vier Bundesländern Deutschlands zudem heraus, dass sich die Akzeptanz des Verfahrens über die Bundesländer hinweg, trotz der teilweise bestehenden Andersartigkeit der Umsetzung, kaum unterscheidet, und dass Schulleitungen bundeslandunabhängig anscheinend kaum „auf die Ergebnisse einer Schulinspektion bezogene Schul- und Unterrichtsentwicklung“ initiieren (ebd., S. 150). Pietsch et al. (2014) berichten für die Schulinspektion Hamburg „sowohl Effekte auf Lernzuwächse als auch Leistungstrends von Schülerinnen und Schülern“ (S. 466). Dederling und Müller (2011b) fassen die Ergebnisse verschiedener Studien zur Wirksamkeit von Schulinspektionen als uneinheitlich, aber eher positiv zusammen.

Hinsichtlich der Vergleichsarbeiten ist das Bild ebenfalls uneinheitlich. Maier (2008) kritisiert den mangelnden Fokus auf Effekte von Vergleichsarbeiten auf die Schul- und Unterrichtsentwicklung. Schwerpunktmäßig untersuchen Studien zur Wirkung von Vergleichsarbeiten ähnlich wie Studien zur Schulinspektion die Nutzung und Akzeptanz von Ergebnissen von Vergleichsarbeiten. Auch hierfür lassen sich verschiedene „Typen“ von Schulen mit unterschiedlichem Umgang mit Vergleichsarbeiten identifizieren (vgl. Groß Ophoff, Hosenfeld & Koch, 2007; Stamm, 2003). Auch bei Maier, Ramsteck und Frühwacht (2013) wird der unterschiedliche Umgang mit Rückmeldungen aus Vergleichsarbeiten deutlich.

Patton (1998) benennt den Prozessnutzen als entscheidende Komponente für die Wirksamkeit von Evaluationen (vgl. auch Koch, 2011; Russ-Eft & Preskill, 2009). Nach Patton ist ein Verständnis von Nutzen im Sinne des (Be-)Nutzens von Ergebnissen zu kurz gefasst. Prozessnutzen definiert er als den Nutzen einer Person/einer Institution durch die Beteiligung am Evaluationsprozess (ebd., S. 56). Über den Nutzen von Evaluationsergebnissen hinaus hat dieser Nutzen den Mehrwert, dass evaluatives Denken gelernt werden kann, welches auch über die konkrete Evaluation hinaus anwendbar ist.

Die Wirksamkeit interner Evaluation wird in deutschsprachigen Studien insgesamt positiver eingeschätzt, so zum Beispiel von Müller (2002) oder Thiel und Thillmann (2012). Jedoch verweisen Thiel und Tillmann auf mögliche Probleme der Generalisierung der Ergebnisse, die nach ihrer Sichtung vorrangig in qualitativen Studien in Pilotprojekten entstanden sind (S. 47).

Aus den Ergebnissen bisheriger Studien ergaben sich vier Hypothesen, für die im Folgenden die Analysen und Ergebnisse vorgestellt werden:

- H1: Veränderungen hin zu mehr Evaluationsaktivitäten haben einen positiven Effekt auf Veränderungen im Leseleistungsniveau der Schule.
- H2: Veränderungen hin zu mehr Evaluationsaktivitäten haben einen positiven Effekt auf Veränderungen in der mittleren Beziehungsqualität der Schule.
- H3: Der Effekt interner Evaluation auf die Entwicklung von Schulen (gemessen anhand der Veränderungen im Leseleistungsniveau) ist größer als der Effekt externer Evaluation.
- H4: Der Effekt interner Evaluation auf die Entwicklung von Schulen (gemessen anhand der Veränderungen der Beziehungsqualität) ist größer als der Effekt externer Evaluation.

Die ersten beiden Hypothesen ergeben sich aufgrund der zusammengefassten uneinheitlichen, aber eher positiven empirischen Ergebnisse zur Wirkung bzw. Wirksamkeit von Evaluation. Die letzten beiden Hypothesen resultieren aus den Annahmen zum Prozessnutzen von Evaluationen, der in diesem Kontext angesprochen wurde. Während externe Evaluationen weitgehend ohne Selbstreflexion der Schulen stattfinden können, erfordern interne Evaluationen eine aktive Umsetzung durch die Schule, was mit einer stärkeren Notwendigkeit der Selbstreflexion einhergeht und einen stärkeren Prozessnutzen erwarten lässt.

Die Hypothesen werden unter Kontrolle des Migrationsstatus, des sozioökonomischen Status und des Geschlechts der Schülerinnen und Schüler überprüft. Dies ist dadurch begründet, dass für alle drei Variablen Zusammenhänge mit den beiden Kriterien nachgewiesen wurden (s. dazu z.B. Walter, Ramm, Zimmer, Heidemeier und Prenzel, 2006; Naumann, Artelt, Schneider & Stanat, 2010; Drechsel & Artelt, 2007; Baumert, Watermann & Schümer, 2003; Ehmke & Jude, 2010; Müller & Ehmke, 2013; Wentzel, Battle, Russell & Looney, 2010; Eder, 1996).

#### 4.4 Stichprobe und Instrumente

Im Rahmen des PISA-Schulpanels wurde eine Zielstichprobe von 60 Schulen angestrebt. Von den insgesamt 90 nach Bundesland stratifiziert zufällig gezogenen Gymnasien (inklusive Ersatzschulen), die aufgrund ihrer Teilnahme an PISA 2000 angeschrieben und gebeten wurden, an den erneuten Erhebungen im Rahmen des PISA-Schulpanels im Jahr 2009 teilzunehmen, nahmen letztlich 59 Gymnasien teil. Die Befragung im Jahr 2000 basiert damit auf den regulären PISA-Instrumenten (PISA-2000-Elternfragebogen, PISA-2000-Schülerfragebogen, PISA-2000-Lesekompetenztest). Im Jahr 2009 wurden in Teilen ebenfalls die regulären PISA-Instrumente verwendet (PISA-2009-Elternfragebogen). Der Lesekompetenztest wurde in 2009, um Unterschiede im Testheftdesign zwischen beiden Erhebungen kontrollieren zu kön-

nen, sowohl in der Ausführung aus dem Jahr 2000 als auch in der Ausführung aus dem Jahr 2009 verwendet. Der PISA-2009-Schülerfragebogen wurde für eine bessere Vergleichbarkeit der Angaben der Schülerinnen und Schüler in beiden Jahren an den Fragebogen aus 2000 angepasst. Für das Jahr 2000 liegen Daten von  $N = 1493$  Schülerinnen und Schülern vor, für das Jahr 2009 von  $N = 1273$ .

Die Befragung zum Zeitraum zwischen diesen beiden Messzeitpunkten wurde mit Hilfe eines retrospektiven Schulleitungsinterviews im Jahr 2010 umgesetzt, welches auf entwicklungsbezogene Aktivitäten, Veränderungen bei den Akteurinnen und Akteuren und in den Leitgedanken der Schule fokussierte. Interviewdaten liegen für insgesamt 47 der 59 Schulen vor. Im Folgenden werden kurz die Quellen dargestellt, aus denen die in die Analysen eingehenden Daten gewonnen wurden.

### **Lesekompetenztest**

Die Lesekompetenz der Schülerinnen und Schüler wurde anhand von 36 gemeinsamen Items (Link-Items) der Lesekompetenztests 2000 und 2009 ermittelt. Lesekompetenz wird hier entsprechend der Definition in PISA „als basales Kulturwerkzeug“ „für eine befriedigende Lebensführung“ verstanden (Naumann, Artelt, Schneider & Stanat, 2010, S. 24). In PISA wird Lesekompetenz beschrieben als die Fähigkeit, „geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potenzial weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ (Artelt, Stanat, Schneider & Schiefele, 2001, S. 80). Da die wiederholte Messung der Lesekompetenz im PISA-Schulpanel auf den Link-Items beruht und durch die Rotation der Aufgaben über die Testhefte nicht alle Schülerinnen und Schüler alle 36 Link-Items bearbeitet haben, ist die Anzahl der Schülerinnen und Schüler, für die Leistungswerte vorliegen, im Vergleich zu den PISA-Gesamtstichproben der Schülerinnen und Schüler leicht reduziert. Für das Jahr 2000 liegen pro Schule durchschnittlich  $N = 19.6$  Leistungswerte vor ( $Min = 9$ ;  $Max = 22$ ), im Jahr 2009 durchschnittlich  $N = 14.4$  Leistungswerte ( $Min = 6$ ;  $Max = 19$ ). Die Leistungswerte wurden für 2000 und 2009 gemeinsam skaliert und gehen als Punktschätzer (WLE; weighted likelihood estimates) in die Analysen ein.

### **Schülerfragebogen**

Mithilfe des Schülerfragebogens wurden Daten zur wahrgenommenen Beziehungsqualität der Schülerinnen und Schüler zu ihren Lehrkräften, zum Migrationshintergrund der Schülerinnen und Schüler, ihrem sozioökonomischen Status und ihrem Geschlecht gewonnen.

Die wahrgenommene Beziehungsqualität wurde über drei Items bestimmt. Hohe Werte in der Skala stehen für eine hohe Beziehungsqualität und somit (auf Schulebene aggregiert) für ein positives Schulklima. Ein Beispielitem dieser Skala ist „Die meisten meiner Lehrer/innen interessieren sich für das, was ich zu sagen habe“. Die Items konnten jeweils über die Kategorien „stimme überhaupt nicht zu“ (1), „stimme eher nicht zu“ (2), „stimme eher zu“ (3) und „stimme ganz zu“ (4) beantwortet werden. Zur Beziehungsqualität liegen Daten von insgesamt  $N = 1451$  (2000) bzw.  $N = 1247$  (2009) Schülerinnen und Schülern vor.

Der sozioökonomische Status der Schülerinnen und Schüler wurde vorrangig anhand der Angaben der Erziehungsberechtigten zu ihrem Beruf und den damit verbundenen Tätigkeiten bestimmt. Bei fehlenden Angaben wurden diese durch die Angaben der Schülerinnen und Schüler ergänzt. Die Ermittlung des sozioökonomischen Status der Schülerinnen und Schüler wird im Abschnitt zum Elternfragebogen (siehe unten) beschrieben.

Der Migrationshintergrund wurde bestimmt anhand der Angaben der Schülerinnen und Schüler zur ihrem Geburtsort und dem ihrer Eltern. Schülerinnen und Schüler wurden anhand dessen, dem Vorgehen aus PISA 2009 entsprechend, als „mit Migrationshintergrund“ oder „ohne Migrationshintergrund“ kategorisiert. Kein Migrationshintergrund wurde dann festgelegt, wenn beide Elternteile in Deutschland geboren wurden. Schülerinnen und Schüler mit mindestens einem im Ausland geborenen Elternteil wurden als Schülerinnen und Schüler mit Migrationshintergrund definiert. Bei fehlenden Werten erfolgte keine Einteilung in eine der beiden Kategorien.

Hinsichtlich des Geschlechts der Schülerinnen und Schüler wurden ihre Angaben aus dem Schülerfragebogen herangezogen. Dazu liegen in 2000  $N = 1469$  und in 2009  $N = 1270$  Angaben vor.

### **Elternfragebogen**

Der Elternfragebogen wurde herangezogen zur Bestimmung des sozioökonomischen Status der Schülerinnen und Schüler. Dazu wurden die Angaben der Eltern zu ihrem Beruf und den damit verbundenen Tätigkeiten verwendet. Im Fall fehlender Angaben wurden die Angaben aus dem Schülerfragebogen ergänzend herangezogen. Zur Bestimmung des sozioökonomischen Status der Schülerin/des Schülers wurden die Angaben zunächst anhand der ISCO-88 (International Standard Classification of Occupation) klassifiziert und dann in den internationalen sozioökonomischen Index (ISEI; International Socio-Economic Index of Occupational Status; vgl. Ganzeboom & Treiman, 1996) übersetzt. Der höchste sozioökonomische Index der Familie, der HISEI, galt dann als Kennwert für den sozioökonomischen Status der Schülerin bzw. des Schülers.

### **Retrospektives Interview**

Im retrospektiven Interview wurden die Schulleitungen zu Veränderungen und Entwicklungsaktivitäten an ihrer Schule zwischen 2000 und 2009 befragt. Ein Bestandteil bezog sich auf Veränderungen in Evaluationsaktivitäten. Dabei wurde unterschieden zwischen internen und externen Evaluationen. In 33 Fällen liegen Angaben zum Stattfinden interner Evaluationen und in 30 Fällen zum Stattfinden externer Evaluationen vor. Das Stattfinden von Evaluation wurde mit „1“ kodiert; wurde für den Zeitraum zwischen 2000 und 2009 angegeben, dass keine Evaluation stattfand, wurde mit „0“ kodiert.

### 4.5 Analysemethoden

Die Hypothesen wurden mit Hilfe von Cross-lagged-panel-Analysen untersucht. Da Messungen bei denselben Schülerinnen und Schülern nicht, wie in gängigen längsschnittlichen Mehrebenenmodellen erforderlich, mehrfach durchgeführt wurden, sondern nur eine Messwiederholung auf Schulebene vorliegt (vgl. Abschnitt 4.1), verwenden wir hier Mehrebenenpfadmodelle, die für eine Messwiederholung auf Schulebene angepasst wurden (Hochweber & Hartig, 2015). In unserem Modell beschränken wir uns auf die Schüler- und die Schulebene; die Klassenebene konnte nicht spezifiziert werden, da die Schülerinnen und Schüler in PISA nur aufgrund ihrer Schul- und nicht ihrer Klassenzugehörigkeit gezogen wurden. Um für mögliche Drittvariablen zu kontrollieren, wurden Migrationsstatus, Geschlecht und sozioökonomischer Status im Jahr 2000 ebenfalls in das Modell aufgenommen. In Abbildung 1 wird das Analysemodell dargestellt.

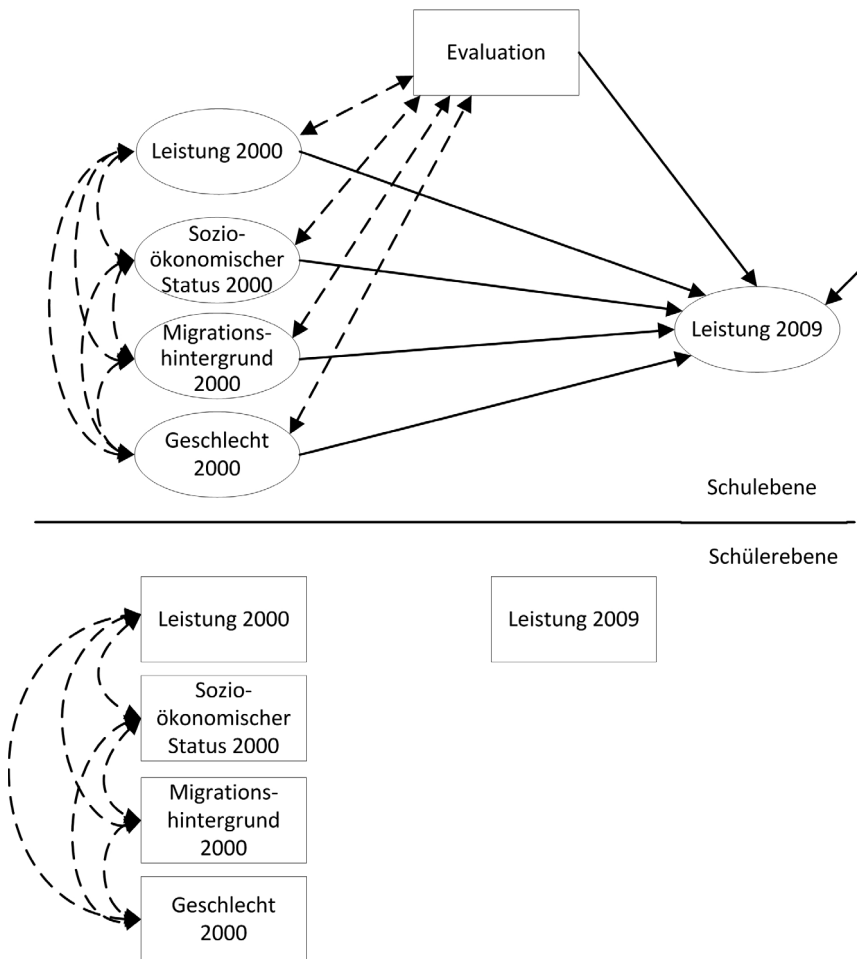


Abbildung 1: Analysemodell mit Evaluation als Prädiktor

Spezifizierte korrelative Zusammenhänge werden über gestrichelte Linien, Regressionspfade über durchgezogenen Linien dargestellt. Für angenommene Beziehungen zwischen Variablen, die nicht im Fokus der Untersuchung stehen, wurden Korrelationen spezifiziert; für die Pfade, die für die Fragestellungen/Hypothesen bedeutsam sind und für die eine gerichtete Beziehung angenommen wird, wurden Regressionspfade spezifiziert (vgl. Kline, 2011). Auf der Schülerebene wurden aufgrund der beschriebenen fehlenden Messwiederholung keine Beziehungen zwischen den Zeitpunkten spezifiziert.

#### 4.6 Ergebnisse zum Effekt interner und externer Evaluation

Im Folgenden werden die Ergebnisse zu den vier oben formulierten Hypothesen wiedergegeben. Für deren Prüfung wurden die Analysen jeweils für die interne und die externe Evaluation durchgeführt. Da in den Hypothesen jeweils eine Richtung für den Effekt angenommen wurde, wurden die zugehörigen Regressionspfade einseitig getestet.

Tabelle 1: Darstellung der relevanten standardisierten Regressionskoeffizienten der Cross-lagged-panel-Analysen auf der Schulebene<sup>5</sup>

Prädiktor	Effekt auf die Beziehungsqualität		Effekt auf die Leseleistung	
	$\beta$	(SE)	$\beta$	(SE)
Interne Evaluation	.43	(.32)	.58*	(.32)
Externe Evaluation	-.42	(.35)	.20	(.43)

\*  $p < .05$

##### Ergebnisse zum Effekt von Evaluation auf das Effektivitätskriterium

Die Ergebnisse zeigen für die interne Evaluation den vermuteten positiven Effekt auf Veränderungen im Leseleistungsniveau. Der Effekt der externen Evaluation auf Veränderungen im Leseleistungsniveau ist nicht signifikant. Die erste Hypothese kann somit nur für die interne Evaluation bestätigt werden. Interne Evaluation geht mit einer erhöhten Wahrscheinlichkeit der Verbesserung des Leseleistungsniveaus einher. Der signifikante Effekt der internen Evaluation bei einem nicht-signifikanten Effekt der externen Evaluation entspricht den in der dritten Hypothese formulierten Erwartungen.

##### Ergebnisse zum Effekt von Evaluation auf das Entwicklungskriterium

Die zweite Hypothese kann weder in den Analysen zum Effekt der internen Evaluation noch zum Effekt der externen Evaluation bestätigt werden. Der Effekt der internen Evaluation auf die Beziehungsqualität findet sich maximal im Trend ( $p = .089$ ). Die Ergebnisse zum Effekt der externen Evaluation sind nicht signifikant. Somit kann die

<sup>5</sup> Die Pfade wurden auf der Schulebene geprüft und resultieren unter Kontrolle von Migrationshintergrund, Geschlecht und sozioökonomischem Status. Aufgrund der gerichteten Annahmen wurden die Pfade einseitig getestet.

Hypothese 4 zum größeren Effekt von interner Evaluation gegenüber externer Evaluation ebenfalls nicht bestätigt werden.

## 5. Diskussion der empirischen Verknüpfung von Schuleffektivität und Schulentwicklung

Die Ergebnisse zeigen für diese Untersuchung unterschiedliche Effekte für interne und externe Evaluation auf Veränderungen in den beiden Kriterien auf. Während interne Evaluation zumindest in Teilen mit einer erhöhten Wahrscheinlichkeit von Schulentwicklung im Sinne einer Verbesserung der mittleren Leseleistung an der Schule einhergeht, findet sich für externe Evaluationen keine erhöhte Wahrscheinlichkeit der Verbesserung des Leseleistungsniveaus der Schule oder des Schulklimas. Der fehlende Effekt externer Evaluation könnte darauf zurückgehen, dass viele Schulen zwar im Vorfeld der Evaluation Aktivitäten initiieren, die Rückmeldung im Rahmen der externen Evaluation es jedoch nicht vermag, in signifikantem Ausmaß Schulentwicklungsprozesse anzustoßen (vgl. z.B. Böhm-Kasper & Selders, 2013; Böttcher & Keune, 2011; Groß Ophoff et al., 2007; van Ackeren et al., 2013). Dies könnte verschiedene Ursachen haben, z.B. mangelndes Vertrauen in die Aussagekraft der Rückmeldung/des Berichts, fehlende Einsicht in die Notwendigkeit von Entwicklungsprozessen sowie fehlende Schulentwicklungskompetenzen, solche Veränderungen anzustoßen und umzusetzen (vgl. Dederling, Fritsch & Weyer, 2012; Ehren & Visscher, 2006; Landwehr, 2011; Sommer, 2011; Wurster, Richter, Schliesing & Pant, 2013).

Bei der internen Evaluation dagegen ist davon auszugehen, dass Schulen, die sich selbst evaluieren, über die Bereitschaft zur kritischen Reflexion der eigenen Entwicklung und damit möglicherweise über fundiertere Erfahrungen und Kompetenzen in der Schulentwicklung verfügen. Unsere Ergebnisse könnten also darauf hindeuten, dass Evaluationen mit stärkerer Beteiligung der Schulen einen größeren Nutzen besitzen als Evaluationen, bei denen Schulen nur eine passive Rolle spielen. Denkbar wäre dabei, dass der von Patton (1998) beschriebene Prozessnutzen eine entscheidende Rolle bei der Wirksamkeit von Evaluationen spielt. Auch aktuelle Ansätze gehen von einer zentralen Bedeutung der schulinternen Prozesse aus (siehe Dederling et al., 2012; Ehren & Visscher, 2006; Landwehr, 2011; Sommer, 2011). Durch eine aktive Beteiligung von Schulen am Evaluationsprozess werden diese zur Auseinandersetzung mit der eigenen schulischen Situation und zur Selbstreflexion, und damit zu kritisch-reflektiertem Denken, angeregt. Entsprechend erscheint es begrüßenswert, dass einige Bundesländer eine stärkere Beteiligung der Schulen am Prozess der Schulinspektion planen bzw. bereits umsetzen (siehe z.B. Niedersachsen oder Nordrhein-Westfalen).

Das Beispiel des PISA-Schulpanels zeigt anschaulich die Möglichkeit und den Mehrwert einer empirischen Verknüpfung von Schuleffektivität und Schulentwicklung. Schuleffektivitätsstudien lassen sich durch eine Erweiterung des Designs gut nutzen, um zu weitreichenderen Aussagen im Bereich der Entwicklung von Schulen zu kommen. Die Grenzen im vorliegenden Beispiel liegen vor allem in der Auswahl und begrenzten Anzahl von für Schulentwicklung relevanten Kriterien. Dadurch, dass es

sich um eine nachträgliche Erweiterung des PISA-Designs handelt, waren die meisten Indikatoren durch die Fragebögen und Tests der Erhebung in 2000 gesetzt und konnten nur in geringem Umfang durch die zusätzlichen Interviews erweitert werden. Aus diesem Grund konnten auch keine weiteren Informationen zu den Schulentwicklungsaktivitäten der Schulen miteinbezogen werden. Auch der Zeitpunkt der Erhebung war durch das PISA-Design gesetzt. Wird eine Erweiterung von Effektivitätsstudien jedoch bereits beim Design der Studie mitgedacht, können entsprechende Indikatoren gleich in der ersten Erhebung gezielt mitaufgenommen werden. Damit kann sich die Aussagekraft und der Nutzen der Studie erhöhen, wenn konkrete Faktoren identifiziert werden, die sowohl einen Einfluss auf die Selbststeuerungsfähigkeit bzw. Entwicklung als auch auf die Effektivität von Schulen besitzen. In Bezug auf die Wirkung spezifischer Reformen können solche Erweiterungen gezielte Evaluations- und/oder Interventionsstudien nicht ersetzen, aber zumindest ergänzen.

Eine zufriedenstellende Verknüpfung von Schuleffektivität und Schulentwicklung auf theoretischer Ebene steht jedoch weiterhin aus. Hintergrund ist die in Abschnitt 3 angedeutete Komplexität von Schule, die in Arbeiten wie denen von Fend (Rekontextualisierung, 2008a; 2008b), Weick (lose Kopplung, 1976; 1982a; 1982b) und Luhmann und Schorr (Technologiedefizit, 1979) deutlich wird. Der Aspekt der Komplexität von Schulen und von Prozessen der Schulentwicklung ist in den oben angesprochenen Modellen jedoch nicht berücksichtigt. Entsprechend gewinnbringend könnte es sein, diese in die weitere Untersuchung der (fehlenden) Verknüpfung von Schuleffektivität und Schulentwicklung einzubeziehen. Cunningham (2001) weist darauf hin, dass sich nicht nur für die Schuleffektivitätsforschung neue Perspektiven auf Schulen ergeben würden, auch für die Schulentwicklungsforschung könnten sich neue Forschungsfragen auftun. Beziehen Forscherinnen und Forscher Konzepte wie jene des Technologiedefizits, der losen Kopplung und der Rekontextualisierung (und damit der Komplexität) in ihre Theorien und Modelle mit ein, müsste dies einen Einfluss auf die Wahl der Methoden, Designs, Fragestellungen und das Verständnis von Schulen haben. Durch eine Beschäftigung mit diesen Themen könnte sich auch das Selbstverständnis von Schuleffektivitäts- und Schulentwicklungsforschung ändern und damit eine neue Grundlage für ihre Verknüpfung entstehen.

## Literatur

- Ackeren van, I. & Bellenberg, G. (2004). Parallelarbeiten, Vergleichsarbeiten und zentrale Abschlussprüfungen. Bestandsaufnahme und Perspektiven. In H.-G. Holtappels, K. Klemm, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 13, S. 125–159). Weinheim und München: Juventa.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–138). Opladen: Leske + Budrich.

- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Baumert, J., Watermann, R. & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6 (1), 46–71.
- 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.
- Bischof, L. M. (2017). *Schulentwicklung und Schuleffektivität. Ihre theoretische und empirische Verknüpfung*. Wiesbaden: VS Verlag.
- Bischof, L. M., Hochweber, J., Hartig, J. & Klieme, E. (2013). Schulentwicklung im Verlauf eines Jahrzehnts – Erste Ergebnisse des PISA-Schulpanels. In E. Klieme & N. Jude (Hrsg.) *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung. Zeitschrift für Pädagogik*, 172–199 [Themenheft]. Weinheim.
- Böhm-Kasper, O. & Selders, O. (2013). „Schulinspektionen sollten regelmäßig durchgeführt werden“? Ländervergleichende Analyse der Wahrnehmung und Akzeptanz von Schulinspektionsverfahren. In I. van Ackeren, M. Heinrich & F. Thiel (Hrsg.) *Evidenzbasierte Steuerung im Bildungssystem. Die Deutsche Schule*. (12), 121–153 [Themenheft].
- Bonsen, M., Büchter, A. & van Ophuysen, S. (2004). Im Fokus: Leistung. Zentrale Aspekte der Schulleistungsforschung und ihre Bedeutung für die Schulentwicklung. In H.-G. Holtappels, K. Klemm, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 13, S. 187–223). Weinheim und München: Juventa.
- Böttcher, W. (2008). Evaluation. In T. Coelen & H.-U. Otto (Hrsg.), *Grundbegriffe Ganztagsbildung. Das Handbuch* (1. Aufl, S. 889–898). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Böttcher, W. & Keune, M. (2011). Wirkungsanalyse der Schulinspektion. Empirische Ergebnisse aus Deutschland. In C. Quesel, V. Husfeldt, N. Landwehr & P. Steiner (Hrsg.), *Wirkungen und Wirksamkeit der externen Schulevaluation* (S. 123–143). Bern: hep.
- Brown, S., Riddell, S. & Duffield, J. (1996). Possibilities and problems of small-scale studies to unpack the findings of large-scale studies of school effectiveness. In J. Gray, D. Reynolds, C. T. Fitz-Gibbon & D. Jesson (Hrsg.), *Merging traditions. The future of research on school effectiveness and school improvement* (S. 93–120). London: Cassell.
- Clark, D. L., Lotto, L. S. & Astuto, T. A. (1984). Effective schools and school improvement: A comparative analysis of two lines of inquiry. *Educational Administration Quarterly*, 20 (3), 41–68.
- Creemers, B. P. M. (2007). *A comprehensive framework for effective school improvement Briefing Paper 27*. Verfügbar unter: <http://www.pjb.co.uk/npl/>
- Creemers, B. P. M. & Kyriakides, L. (2008). *The dynamics of educational effectiveness. A contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Creemers, B. P. M. & Kyriakides, L. (2010). Using the Dynamic Model to develop an evidence-based and theory-driven approach to school improvement. *Irish Educational Studies*, 29 (1), 5–23.

- Creemers, B. P. M. & Kyriakides, L. (2012). *Improving quality in education. Dynamic approaches to school improvement*. Abingdon: Routledge.
- 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. & Reezigt, G. J. (1997). School effectiveness and school improvement: Sustaining links. *School Effectiveness and School Improvement*, 8 (4), 396–429.
- Creemers, B. P. M. & Reynolds, D. (1989). The future development of school effectiveness and school improvement. In B. P. M. Creemers, T. Peters & D. Reynolds (Hrsg.), *School effectiveness and school improvement* (S. 379–383). Amsterdam/Lisse: Swets & Zeitlinger.
- Cunningham, R. (2001, September). *Chaos, complexity and the study of education communities*, University of Leeds. Verfügbar unter: [www.leeds.ac.uk/educol/documents/00001895.doc](http://www.leeds.ac.uk/educol/documents/00001895.doc)
- Dedering, K. (2012). Schulinspektion als wirksamer Weg der Systemsteuerung? *Zeitschrift für Pädagogik*, 58 (1), 69–88.
- Dedering, K., Fritsch, N. & Weyer, C. (2012). Die Ankündigung von Schulinspektionen und deren innerschulische Effekte – hektisches Treiben oder genügsame Gelassenheit? In S. Hornberg & M. Parreira do Amaral (Hrsg.), *Deregulierung im Bildungswesen* (S. 205–222). Münster: Waxmann.
- Dedering, K. & Müller, S. (2011a). School improvement through inspections? First empirical insights from Germany. *Journal of Educational Change*, 12 (3), 301–322.
- Dedering, K. & Müller, S. (2011b). Wirkung von externen Evaluationen/Schulinspektionen. *Journal für Schulentwicklung*, 15 (1), 17–24.
- Dellar, G. B. (1998). School climate, school improvement and site-based management. *Learning Environments Research*, 1 (3), 353–367.
- Diedrich, M. (2008). *Demokratische Schulkultur. Messung und Effekte*. Münster: Waxmann.
- Downes, D. & Vindurampulle, O. (2007). *Value-added measures for school improvement* (Paper No. 13). East Melbourne: Department of Education and Early Childhood Development. Verfügbar unter: [edulibrary/public/publ/research/publ/value-added-measures-report.pdf](http://edulibrary/public/publ/research/publ/value-added-measures-report.pdf)
- Drechsel, B. & Artelt, C. (2007). Lesekompetenz. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme et al. (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 225–247). Münster: Waxmann.
- Eder, F. (1996). *Schul- und Klassenklima. Ausprägung, Determinanten und Wirkungen des Klimas an höheren Schulen* (Studien zur Bildungsforschung & Bildungspolitik, 8). Innsbruck: Studien-Verlag.
- Eder, F. (2006). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 622–631). Weinheim: Beltz.
- Ehmke, T. & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231–254). Münster: Waxmann.
- Ehren, M. C. M. & Visscher, A. J. (2006). Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, 54 (1), 51–72.
- Feldhoff, T., Bischof, L. M., Emmerich, M. & Radisch, F. (2015). Was nicht passt, wird passend gemacht! Zur Verbindung von Schuleffektivität und Schulentwicklung. In H. J. Abs, T. Brüsemeister, M. Schemmann & J. Wissinger (Hrsg.), *Governance im Bildungswesen – Analysen zur Mehrebenenperspektive, Steuerung und Koordination*. Wiesbaden: Springer VS.

- Fend, H. (2008a). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen* (2. Aufl.). Wiesbaden: VS Verlag.
- Fend, H. (2008b). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag.
- Ganzeboom, H. B. & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research*, 25 (3), 201–239.
- Gärtner, H. (2013). Praxis und Verhältnis interner und externer Evaluation im Schulsystem im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 16 (4), 693–712.
- Gray, J., Reynolds, D., Fitz-Gibbon, C. T. & Jesson, D. (Hrsg.). (1996). *Merging traditions. The future of research on school effectiveness and school improvement*. London: Cassell.
- Groß Ophoff, J. (2013). *Lernstandserhebungen: Reflexion und Nutzung* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 85). Münster u.a.: Waxmann.
- Groß Ophoff, J., Hosenfeld, I. & Koch, U. (2007). Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. *Empirische Pädagogik*, 21 (4), 411–427.
- Gustafsson, J. E. (2010). Longitudinal designs. In B. P. M. Creemers, L. Kyriakides & P. Sammons (Hrsg.), *Methodological advances in educational effectiveness research* (S. 77–101). Milton Park: Routledge.
- Harris, A. (2001). Contemporary perspectives on school effectiveness and school improvement. In A. Harris & N. Bennett (Hrsg.), *School effectiveness and school improvement. Alternative perspectives* (S. 7–25). London: Continuum.
- Harris, A. & Bennett, N. (Hrsg.). (2001). *School effectiveness and school improvement. Alternative perspectives*. London: Continuum.
- Helmke, A. (2006). *Unterrichtsqualität. Erfassen, Bewerten, Verbessern*. Seelze: Kallmeyersche Verlagsbuchhandlung.
- 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 et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 113–151). Münster: Waxmann.
- Hochweber, J. & Hartig, J. (2015). Analyzing organizational growth in repeated multistage sampling designs. Manuscript submitted for publication.
- Hoeben, W. T. J. G. (Hrsg.). (1998). *Effective school improvement: State of the art. Contribution to a discussion*. Groningen: GION, Institute for Educational Research, University of Groningen.
- Hopkins, D. (1990). The international school improvement project (ISIP) and effective schooling: Towards a synthesis. *School Organisation*, 10 (2 & 3), 179–194.
- Hopkins, D. (1996). Towards a theory for school improvement. In J. Gray, D. Reynolds, C. T. Fitz-Gibbon & D. Jesson (Hrsg.), *Merging traditions. The future of research on school effectiveness and school improvement* (S. 30–50). London: Cassell.
- 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 et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 11–21). Münster: Waxmann.
- Klieme, E. & Jude, N. (Hrsg.). (2013) *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung* [Themenheft]. *Zeitschrift für Pädagogik*. Weinheim.
- 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 et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277–300). Münster: Waxmann.

- Klieme, E. & Steinert, B. (2008). Schulentwicklung im Längsschnitt. Ein Forschungsprogramm und erste explorative Analysen. In M. Prenzel & J. Baumert (Hrsg.) Vertiefende Analysen zu PISA 2006. *Zeitschrift für Erziehungswissenschaft*. (Sonderheft 10), 221–238 [Themenheft]. Wiesbaden: VS Verlag.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3. Aufl.). New York, London: Guilford Press.
- Koch, U. (2011). *Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten? Datenkompetenz von Lehrkräften und die Nutzung von Ergebnissrückmeldungen aus Vergleichsarbeiten*. Münster: Waxmann.
- Kotthoff, H.-G. & Böttcher, W. (2010). Neue Formen der „Schulinspektion“: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. Alt-richter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 295–325). Wiesbaden: VS Verlag.
- Lambrecht, M. & Rürup, M. (2012). Bildungsforschung im Rahmen einer *evidence based policy*: Das Beispiel „Schulinspektion“. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung* (S. 57–77). Wiesbaden: VS Verlag.
- Landwehr, N. (2011). Thesen zur Wirkung und Wirksamkeit der externen Schulevaluati-on. In C. Quesel, V. Husfeldt, N. Landwehr & P. Steiner (Hrsg.), *Wirkungen und Wirk-samkeit der externen Schulevaluation* (S. 35–69). Bern: hep.
- Lindahl, R. A. (2011). The crucial role of assessing the school’s climate and culture in planning for school improvement. *Education Research Complete*, 20 (1), 16–30.
- Luhmann, N. & Schorr, K.-E. (1979). Das Technologiedefizit der Erziehung und die Päd-agogik. *Zeitschrift für Pädagogik*, 25 (3), 345–365.
- Maier, U. (2008). Rezeption und Nutzung von Vergleichsarbeiten aus der Perspektive von Lehrkräften. *Zeitschrift für Pädagogik*, 54 (1), 95–117.
- Maier, U., Ramsteck, C. & Kuper, H. (2013). Lehr-lerntheoretische Argumentations-muster bei der Interpretation und Nutzung von Vergleichsarbeitenrückmeldungen durch Gymnasiallehrkräfte. In I. Ackeren van (Hrsg.), *Evidenzbasierte Steuerung im Bil-dungssystem? Befunde aus dem BMBF-SteBis-Verbund* (Die deutsche Schule: Beiheft, Bd. 12, S. 74–96). Münster: Waxmann.
- Merrett, G. (2000). The factors and strategies which contribute to school improvement. *Improving Schools*, 3 (3), 33–37.
- Müller, K. & Ehmke, T. (2013). Soziale Herkunft als Bedingung der Kompetenzentwick-lung. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschrit-te und Herausforderungen in Deutschland* (S. 245–274). Münster/New York/München/Berlin: Waxmann.
- Müller, S. (2002). *Schulinterne Evaluation. Gelingensbedingungen und Wirkungen*. Dort-mund: IFS.
- 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 et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23–71). Münster: Waxmann.
- Patton, M. Q. (1998). Die Entdeckung des Prozeßnutzens. In M. Heiner (Hrsg.), *Experi-mentierende Evaluation. Ansätze zur Entwicklung lernender Organisationen* (S. 55–66). Weinheim: Juventa.
- Pietsch, M., Janke, N. & Mohr, I. (2014). Führt Schulinspektion zu besseren Schülerlei-stungen? Difference-in-difference-Studien zu Effekten der Schulinspektion Hamburg auf Lernzuwächse und Leistungstrends. *Zeitschrift für Pädagogik*, 60 (3), 446–470.
- Reezigt, G. J. (2001). *A framework for effective school improvement*: GION, Institute for Educational Research, University of Groningen.

- Reezigt, G. J. & Creemers, B. P. M. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement*, 16 (4), 407–424.
- Reynolds, D., Bollen, R., Creemers, B. P. M., Hopkins, D., Stoll, L. & Lagerweij, N. (Hrsg.). (1996). *Making good schools. Linking school effectiveness and school improvement*. London: Routledge.
- Reynolds, D., Hopkins, D. & Stoll, L. (1993). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement*, 4 (1), 37–58.
- Reynolds, D., Sammons, P., De Fraine, B., Townsend, T. & Van Damme, J. (2011). *Educational effectiveness research. State of the art review*. Paper presented at the annual meeting of the international congress for school effectiveness and improvement, Cyprus.
- Reynolds, D. & Stoll, L. (1996). Merging school effectiveness and school improvement. The knowledge bases. In D. Reynolds, R. Bollen, B. P. M. Creemers, D. Hopkins, L. Stoll & N. Lagerweij (Hrsg.), *Making good schools. Linking school effectiveness and school improvement* (S. 94–112). London: Routledge.
- Reynolds, D., Teddlie, C., Hopkins, D. & Stringfield, S. (2000). Linking School Effectiveness and School Improvement. In C. Teddlie & D. Reynolds (Hrsg.), *The international handbook of school effectiveness research* (S. 206–231). London: Falmer Press.
- Rolf, H.-G. (2007). *Skizzen zu einer Theorie der Schulentwicklung*. Köln. Verfügbar unter: [http://www.netzwerk-schulentwicklung.de/Skizzen zu einer Theorie der Schulentwicklung.pdf](http://www.netzwerk-schulentwicklung.de/Skizzen%20zu%20einer%20Theorie%20der%20Schulentwicklung.pdf)
- Russ-Eft, D. F. & Preskill, H. S. (2009). *Evaluation in organizations. A systematic approach to enhancing learning, performance, and change* (2. Aufl.). New York: Basic Books.
- Scheerens, J. (2000). *Improving school effectiveness*. Paris: Unesco, International Institute for Educational Planning.
- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Scheerens, J., Luyten, H., Steen, R. & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Twente: University of Twente, Department of Educational Organization and Management.
- Seidel, T. (2008). Stichwort: Schuleffektivitätskriterien in der internationalen empirischen Forschung. *Zeitschrift für Erziehungswissenschaft*, 11 (3), 348–367.
- Sommer, N. (2011). Wie beurteilen schulische Gruppen die erlebte Schulinspektion? Ergebnisse einer Befragung. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht* (S. 137–164). Münster: Waxmann.
- Stamm, M. (2003). *Evaluation und ihre Folgen für die Bildung. Eine unterschätzte pädagogische Herausforderung*. Münster: Waxmann.
- Stoll, L. (1994). School effectiveness and school improvement: A meeting of two minds. In D. H. Hargreaves & D. Hopkins (Hrsg.), *Development planning for school improvement* (School development series, S. 129–140). London, New York: Cassell.
- Stoll, L., Creemers, B. P. M. & Reezigt, G. J. (2006). Effective school improvement. Similarities and differences in improvement in eight European countries. In A. Harris & J. H. Chrispeels (Hrsg.), *Improving schools and educational systems. International perspectives* (1. Aufl., S. 90–104). London [u.a.]: Routledge.
- Stoll, L. & Fink, D. (1996). *Changing our schools. Linking school effectiveness and school improvement*. Buckingham: Open University Press.

- Stoll, L. & Mortimore, P. (1997). School effectiveness and school improvement. In J. White & M. Barber (Hrsg.), *Perspectives on school effectiveness and school improvement* (S. 9–24). London: University of London, Institute of Education.
- Stufflebeam, D. L., Madaus, G. F. & Kellaghan, T. (Hrsg.). (2000). *Evaluation models. Viewpoints on educational and human services evaluation* (2. Aufl.). Boston: Kluwer Academic.
- Teddlie, C. & Stringfield, S. (2007). A history of school effectiveness and improvement research in the USA focusing on the past quarter century. In T. Townsend (Hrsg.), *International handbook of school effectiveness and improvement* (S. 131–166). Dordrecht: Springer.
- Tenorth, H.-E. (2006). Professionalität im Lehrerberuf. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 580–597.
- Thiel, F. & Thillmann, K. (2012). Interne Evaluation als Instrument der Selbststeuerung von Schulen. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung* (S. 35–55). Wiesbaden: VS Verlag.
- Thillmann, K. (2012). *Schulentwicklung und Schulorganisation. Eine empirische Untersuchung schulischer Organisationsgestaltung vor dem Hintergrund der Neuen Steuerung im Bildungssystem*. Dissertation, Freie Universität Berlin. Berlin. Verfügbar unter: [http://www.diss.fu-berlin.de/diss/receive/FUDISS\\_thesis\\_000000037514](http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000037514)
- van Ackeren, I., Binnewies, C., Clausen, M., Demski, D., Dormann, C., Koch, A. et al. (2013). Welche Wissensbestände nutzen Schulen im Kontext von Schulentwicklung? Theoretische Konzepte und erste deskriptive Befunde des EviS-Verbundprojektes im Überblick. In I. Ackeren van (Hrsg.), *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund* (Die deutsche Schule: Beiheft, Bd. 12, S. 51–72). Münster: Waxmann.
- Van Damme, J., Opdenakker, M., 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., Ramm, G., Zimmer, K., Heidemeier, H. & Prenzel, M. (2006). PISA 2003 – Kompetenzen von Jungen und Mädchen mit Migrationshintergrund in Deutschland. Ein Problem ungenutzter Potentiale? *Unterrichtswissenschaft*, 34 (2), 146–169.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1–19.
- Weick, K. E. (1982a). Administering education in loosely coupled schools. *Education Digest*, 48 (4), 28–36.
- Weick, K. E. (1982b). Administering education in loosely coupled schools. *Phi Delta Kappan*, 63 (10), 673–676.
- Weick, K. E. (2008). Bildungsorganisationen als lose gekoppelte Systeme. In S. Koch & M. Schemmann (Hrsg.), *Neoinstitutionalismus in der Erziehungswissenschaft. Grundlegende Texte und empirische Studien* (1. Aufl., S. 85–109). Wiesbaden: VS Verlag.
- Wentzel, K. R., Battle, A., Russell, S. L. & Looney, L. B. (2010). Social supports from teachers and peers as predictors of academic and social motivation. *Contemporary Educational Psychology*, 35 (3), 193–202.
- Wurster, S. & Gärtner, H. (2013). Schulen im Umgang mit Schulinspektion und deren Ergebnissen. *Zeitschrift für Pädagogik*, 59 (3), 425–445.
- Wurster, S., Richter, D., Schliesing, A. & Pant, H. A. (2013). Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. In I. Ackeren van (Hrsg.), *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund* (Die deutsche Schule: Beiheft, Bd. 12, S. 19–50). Münster: Waxmann.