

Wie verändert sich die Unterrichts- und Schulqualität, die Leistung und die Schülerzusammensetzung nach der Schulinspektionsdiagnose „erheblicher Entwicklungsbedarf“?

Marina Wenger  · Holger Gärtner · Martin Brunner

Eingegangen: 19. Februar 2020 / Überarbeitet: 27. August 2021 / Angenommen: 27. Oktober 2021 /
Online publiziert: 15. Februar 2022
© Der/die Autor(en) 2022

Zusammenfassung Die Schulinspektion evaluiert Schulen mit dem Ziel der Qualitätssicherung von Unterrichts- und Schulqualität. Dies gilt insbesondere für Schulen, an denen „erheblicher Entwicklungsbedarf“ festgestellt wurde. Diese Schulen bekommen zusätzliche Unterstützung, erfahren aber auch zusätzlichen Druck durch diese Einordnung. Die weitere Entwicklung dieser Schulen ist bisher kaum erforscht. Diese Studie untersucht mit Daten der Schulinspektion, der amtlichen Statistik und Leistungsdaten von 333 Berliner Grundschulen Veränderungen in Indikatoren der Unterrichts- und Schulqualität, der Schulleistung, und der Zusammensetzung der Schülerschaft (SES und Anteil mit nicht-deutscher Herkunftssprache) nach der Diagnose „erheblicher Entwicklungsbedarf“. Die empirischen Analysen zeigten, dass sich bei diesen Schulen die Unterrichts- und Schulqualität nur geringfügig veränderte, sich der Leistungsabstand zu allen anderen Grundschulen nicht statistisch signifikant verringerte, und sich die Zusammensetzung der Schülerschaft hinsichtlich des sozioökonomischen Status (SES) nicht veränderte. Jedoch erhöhte sich der Anteil von Kindern mit nicht-deutscher Herkunftssprache statistisch signifikant.

Schlüsselwörter Schulinspektion · Entwicklungsbedarf · Schülerleistungen

Dr. Marina Wenger · Prof. Dr. Holger Gärtner
Institut für Schulqualität Berlin und Brandenburg e. V. (ISQ), Schwendenerstr. 31, 14195 Berlin, Deutschland
E-Mail: holger.gaertner@fu-berlin.de

Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Deutschland
E-Mail: marina.wenger@fu-berlin.de

Prof. Dr. Martin Brunner
Universität Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Deutschland
E-Mail: martin.brunner@uni-potsdam.de

How does instruction and school quality, school performance, and composition of the student body change after the school inspection diagnosis “considerable need for development”?

Abstract The School Inspectorate evaluates schools with the aim of quality assurance and quality development. This applies in particular to schools where “considerable need for development” has been identified. These schools receive additional support, but also experience additional pressure as a result. The further development of these schools has hardly been researched so far. Therefore, using data from the school inspectorate, official statistics and performance data from 333 Berlin primary schools, this study examines changes in indicators of instruction and school quality, school performance, and composition of the student body (SES and percentage of students from non-German-speaking homes) following the diagnosis “considerable need for development”. The empirical analyses showed that at these schools instruction and school quality changed only slightly, the achievement gap did not decrease significantly in statistical terms (in comparison to all other primary schools), and the composition of the student body did not change with regard to socio-economic status (SES). However, a statistically significant increase of the proportion of students from non-German-speaking homes was observed.

Keywords School inspection · Special measure · Student performance

1 Einleitung – Funktionen und Wirkannahmen der Schulinspektion

In vielen Ländern der Bundesrepublik Deutschland wurde im Zuge der Reformmaßnahmen nach der ersten Veröffentlichung der PISA-Ergebnisse Mitte der 2000er Jahre eine externe Evaluation von Schulen als neue Form der Qualitätssicherung eingeführt (im Folgenden einheitlich als Schulinspektion benannt, vgl. bspw. Döbert und Dederling 2008). Seitdem haben sich die Verfahrensweisen der Bundesländer unterschiedlich entwickelt: Von der Abschaffung der Schulinspektion, einer freiwilligen Teilnahme der Schulen bis hin zur weiterhin regelmäßigen Durchführung (Tarkian et al. 2019). Das übergeordnete Ziel der Schulinspektion im Sinne einer „Wirkungshoffnung“ (Dederling 2012) ist die Verbesserung der Unterrichts- und Schulqualität sowie insbesondere die Verbesserung der Leistungen von Schülerinnen und Schülern (Ehren und Visscher 2006). Generell beurteilt Schulinspektion schulische Qualität auf Grundlage normativer Festlegungen, die sich meist aus den landesspezifischen Qualitätsrahmen ergeben. Auch wenn sich diese Rahmenmodelle im Detail unterscheiden, besteht breiter Konsens darüber, Unterrichts- und Schulqualität als wissenschaftlich fundierte Unterrichts- und Schulpraxis zu verstehen, die die Entwicklung fachspezifischer, personaler und sozialer Kompetenzen der Schülerinnen und Schüler fördert und so insgesamt die soziale, kulturelle und ökonomische Teilhabe in einer demokratischen Gesellschaft und lebenslanges Lernen unterstützt (vgl. bspw. Senatsverwaltung für Bildung, Jugend und Wissenschaft 2013).

Mit Hilfe der Schulinspektion sollen mehrere Funktionen erfüllt werden. Die Beurteilungen der Schulinspektion liefern Wissen (Funktion der Wissensgewinnung),

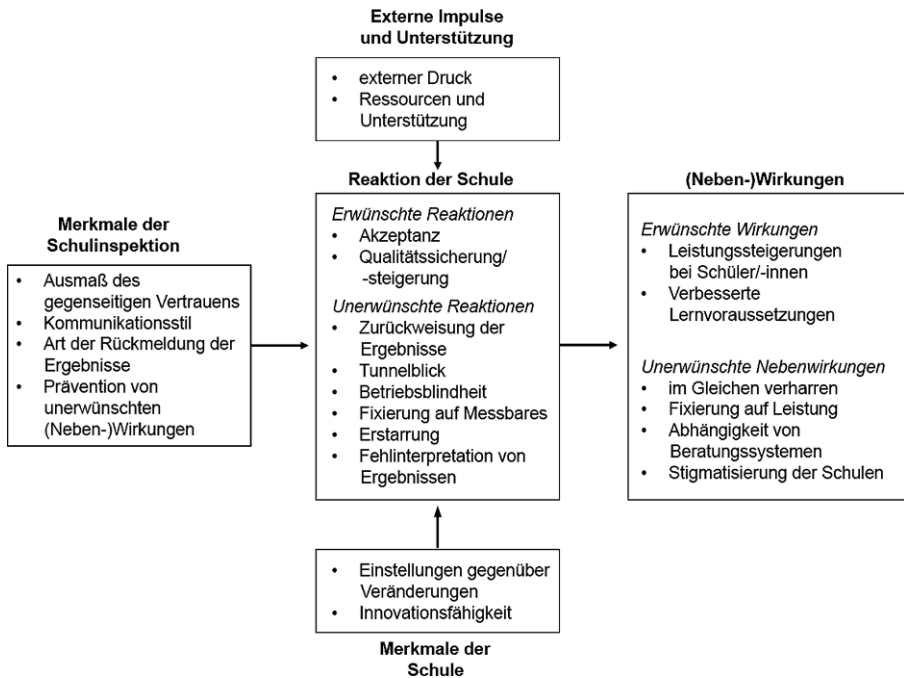


Abb. 1 Modell von Ehren und Visscher (2006; nach der Übersetzung aus Pietsch et al. 2014) zur multidimensionalen Wirkungsweise von Schulinspektion

das mehrere Folgeprozesse auslösen soll (Landwehr 2011). So sollen schulinterne Maßnahmen ausgelöst werden, die direkt zu einer Verbesserung der Unterrichts- und Schulqualität beitragen (Schulentwicklungsfunktion). Schulen nutzen Schulinspektionsergebnisse in diesem Sinne eigenständig, um aufgezeigten Entwicklungsbedarf zu bearbeiten und ihre Stärken zu sichern. Gleichzeitig können sowohl die Einzelschule als auch die Bildungsverwaltung mithilfe der Schulinspektionsergebnisse Rechenschaft über die Qualität der Schule ablegen (Rechenschaftsfunktion; Landwehr 2011). Im internationalen Raum (bspw. in England) wird insbesondere auf die Rechenschaftsfunktion zurückgegriffen, um die Wirkung von Schulinspektionsergebnissen zu erklären, denn schlechte Ergebnisse sind mit Sanktionen verbunden (Dedering 2012). Neben Sanktionen soll die Rechenschaftsfunktion aber auch zur Weiterentwicklung der Schul- und Unterrichtsqualität beitragen, beispielsweise durch von der Schulaufsicht angeordnete Maßnahmen für Schulen, bei denen Entwicklungsbedarf festgestellt wurde. Schließlich unterstützt die Schulinspektion die Implementation politisch veränderter Normen und Anforderungen (Funktion der Normendurchsetzung) (Landwehr 2011), indem sie diese in den Katalog der Bewertungskriterien aufnimmt, an die Schulen zur Vorbereitung auf eine Schulinspektion kommuniziert und entsprechend evaluiert.

Die Wirkung einer Evaluation durch die Schulinspektion ist multidimensional und multidirektional (Abb. 1). Sie wird bedingt durch Merkmale der Schulinspek-

tion, externe Impulse, Merkmale der Schule und die Reaktion der Schule auf die Schulinspektion (bspw. Ehren und Visscher 2006; Reezigt und Creemers 2005).

Wesentliche externe Faktoren zur Weiterentwicklung von Schulen sind Unterstützungssysteme (bspw. durch Beratungen der zuständigen Schulaufsicht; Senatsverwaltung für Bildung, Jugend und Wissenschaft 2014), aber auch der Druck, der beispielsweise durch die Veröffentlichung der Ergebnisse auf die Schulen entsteht (Altrichter und Kemethofer 2015; Reezigt und Creemers 2005; Abb. 1). Die Wirkung von externem Druck wird dabei einerseits als Anregung zur Qualitätsentwicklung gesehen, andererseits werden auch (unerwünschte und nichtintendierte) Nebeneffekte durch externen Druck diskutiert (bspw. Altrichter und Kemethofer 2015; Ehren und Visscher 2006; Kotthoff und Böttcher 2009; siehe auch Abschn. 2.1 und 2.2). Die externen Impulse und Unterstützungsmaßnahmen können entscheidend für die Wirkungskette der Schulinspektion sein, denn diese Faktoren wirken ebenso wie interne Faktoren der Schule (bspw. Einstellungen zu Veränderungen) darauf, wie die Schulen die Evaluationsergebnisse der Schulinspektion verarbeiten bzw. darauf reagieren. Die Reaktionen der Schulen auf die Schulinspektion unter ihren spezifischen Gegebenheiten reichen von Akzeptanz und Antrieb zur Qualitätsentwicklung bis hin zur Zurückweisung oder dem Missverstehen der Ergebnisse (Abb. 1). Dies kann gewünschte Wirkungen der Schulinspektion, wie eine verbesserte Unterrichts- und Schulqualität und dadurch vermittelt eine Leistungssteigerung der Schülerinnen und Schüler nach sich ziehen (Ehren und Visscher 2006; Abb. 1). Somit werden unter einer verbesserten Unterrichts- und Schulqualität Veränderungen der Schulpraxis verstanden, die als Reaktion der Schule auf die Erwartungen der Schulinspektion erstanden (Ehren und Visscher 2006).

Denkbar sind aber auch unerwünschte Folgen, wie beispielsweise die Stigmatisierung der Schulen (Böhm-Kasper et al. 2016) und des pädagogischen Personals, das dort arbeitet und der Schülerinnen und Schüler, die dort lernen (Abb. 1, Ehren und Visscher 2006). Dies betrifft insbesondere solche Schulen, die von der Schulinspektion als besonders schwach bewertet werden, also Schulen, deren Leistungen die Qualitätskriterien der Schulinspektion nicht genügend erfüllen. In Bayern werden solche Schulen zum Beispiel als Schulen mit „großer Schwäche“ bezeichnet, in Hamburg als „Fallkonferenzschulen“ und in Berlin als „Schulen mit erheblichem Entwicklungsbedarf“; in England sind es Schulen, die dem „special measure“-Programm zugeordnet werden. Im Folgenden verwenden wir die Bezeichnung „Schulen mit Entwicklungsbedarf“. Mit einer solchen Diagnose sind verstärkte Erwartungen verknüpft, die schulischen Prozesse nach der Schulinspektion rasch zu verbessern. Bisher gibt es international (bspw. Matthews und Sammons 2005) und für Deutschland (Dedering et al. 2016; Dedering 2018) nur sehr wenige Studien, die Veränderungen in Folge einer solchen Schulinspektionsdiagnose untersuchen. Dieser Beitrag widmet sich daher folgender Forschungsfrage: Wie verändert sich die Unterrichts- und Schulqualität, die Leistung und die Schülerzusammensetzung nach der Schulinspektionsdiagnose „erheblicher Entwicklungsbedarf“? Für die empirischen Analysen verwenden wir hierzu einen bislang einzigartigen Datensatz, der Daten der Schulinspektion, der amtlichen Statistik und Leistungsdaten von allen öffentlichen Berliner Grundschulen ($N = 333$) verknüpft.

2 Empirische Befunde

Im Folgenden wird der bisherige empirische Kenntnisstand zu den Folgen von Schulinspektion berichtet. Die Befunde werden hier zunächst generell (also für Schulen mit und ohne Entwicklungsbedarf) beschrieben, bevor anschließend auf Kenntnisse über Schulen mit Entwicklungsbedarf eingegangen wird.

2.1 Folgen der Schulinspektion für Schulen

Die bisherige Forschung zu generellen Folgen der Schulinspektion fokussierte auf drei zentrale Aspekte: (1) die Entwicklung der Unterrichts- und Schulqualität, (2) die Leistungsentwicklung der Schülerinnen und Schüler und (3) nichtintendierte Nebeneffekte.

1. Einige Untersuchungen zeigten, dass ein größerer Anteil von Lehrkräften in Folge der Schulinspektion etwas am Lehrstil und den Unterrichtsmethoden ändern wollte (bspw. Brimblecombe et al. 1996; in England: 33 %). Es konnte auch festgestellt werden, dass eher einfach umzusetzende Verbesserungen initiiert wurden, beispielsweise die Erhöhung der Zeit für den Sprachunterricht in den Niederlanden (Ehren und Visscher 2008). Gärtner et al. (2009) beschreiben, dass in Brandenburg durchschnittlich vier konkrete qualitätssichernde oder -verbessernde Maßnahmen in Folge auf die Rückmeldung der Schulinspektion eingeleitet wurden und 45 % der Schulleitungen einen Einfluss auf den Unterricht wahrnahmen. Auch bei Preuße et al. (2019) fand ein Großteil (65 %) der Berliner Schulleitungen, dass die Schulinspektion Prozesse der Schulentwicklung anregte. Auf der anderen Seite schätzten in dieser Studie 61 % der Schulleitungen den Nutzen der Schulinspektion für die Qualitätssicherung pädagogischer Prozesse als gering ein und nur 36 % sahen eine positive Langzeitwirkung. Befunde einer internationalen Studie aus sieben europäischen Ländern zeigen, dass – so wie theoretisch angenommen (Ehren und Visscher 2006; Reezigt und Creemers 2005) – externer Druck Schulen zur Qualitätsentwicklung anregte: Schulleitungen, die höheren externen Druck verspürten, berichteten über vermehrte Entwicklungsaktivitäten an ihren Schulen (Altrichter und Kemethofer 2015). Weiterhin führte die Androhung von Sanktionen in Florida in den USA dazu, dass mehr Geld für Fortbildungen und für die Unterrichtsentwicklung in leistungsschwachen Schulen eingesetzt wurde (Chiang 2009). Zudem wurden positive Effekte auf die Unterrichts- und Schulqualität durch die Veröffentlichung der Ergebnisse aus Schulinspektionen in den OECD-Ländern beschrieben und dadurch rückgeschlossen, dass externer Druck die Entwicklungstätigkeit erhöhen kann (Faubert 2009). Die Generalisierbarkeit dieser Effekte scheint jedoch eingeschränkt zu sein: So fanden zum Beispiel Gärtner et al. (2014) in einer Längsschnittstudie in Berlin und Brandenburg kaum relevante Veränderungen in verschiedenen Schulqualitätsindikatoren in Folge der Schulinspektion.
2. Zur Leistungsentwicklung von Schülerinnen und Schülern in Folge von Schulinspektion wurden keine eindeutigen Ergebnisse gefunden. In Brandenburg wurde keine positive Entwicklung der Schülerleistung im Nachgang der Schulinspekti-

- on gefunden (Gärtner und Pant 2011). Ähnliche Befunde berichten Cullingford (1999) und Shaw et al. (2003) für England, Rosenthal (2004) konnte sogar einen Leistungsabfall nach Schulinspektionen aufzeigen. Andererseits konnten Studien aus den USA sowie aus Hamburg in Deutschland auch positive Effekte berichten (Hanushek und Raymond 2005; Lee 2006; Pietsch et al. 2014). Manche Studien untersuchten differenzielle Effekte, die je nach Schulkontext, Bedingungen der Evaluation, Ressourcen, Untersuchungsmethode oder auch Fach unterschiedliche Auswirkungen aufzeigen (Hanushek und Raymond 2005; Lee 2006; Luginbuhl et al. 2009; Pietsch et al. 2014; Penninckx et al. 2016; Shaw et al. 2003).
3. Einige Studien weisen darauf hin, dass Steuerungsinstrumente im Bildungswesen zu nichtintendierten Effekten führen können (Bellmann und Weiß 2009). Mit Blick auf die Schulinspektion gehören hierzu das „window dressing“ bspw. durch Ausschließen bestimmter Schülerinnen und Schüler von Tests oder das „Aufhübschen“ der Dokumente in den Niederlanden (Ehren und Swanborn 2012; Kemehtofer und Helm 2017) sowie erhöhter Stress der Schulleitung bzw. des Kollegiums in den Niederlanden und in Berlin (Penninckx et al. 2016; Preuße et al. 2019). Weiterhin blieben aufgrund der Beanspruchung durch die Schulinspektion andere Aufgaben in Berlin liegen und das Arbeitsklima an Schulen wurde als angespannt beschrieben. Auch die zeitliche Belastung zur Vorbereitung und Durchführung der Schulinspektion wurde als hoch eingeschätzt (Preuße et al. 2019). Böhm-Kasper et al. (2016) benennen zudem eine „Stigmatisierung“ von Schulen in verschiedenen Bundesländern Deutschlands, die auch das Vertrauen in die Schule schwächen kann.

Insgesamt ist die Befundlage zu den Folgen von Schulinspektion nicht eindeutig: Die Schulinspektion scheint insgesamt nicht zwangsläufig zu einer Verbesserung von Unterrichts- und Schulqualität an den inspizierten Schulen sowie zu einer Leistungssteigerung der Schülerinnen und Schüler zu führen. Dabei stellt sich jedoch die Frage, inwiefern dies auch bei den Schulen mit Entwicklungsbedarf zutrifft. Denn diese Schulen erfahren durch diese Bewertung besondere Aufmerksamkeit durch zahlreiche Akteure, zum Beispiel durch die Bildungsverwaltung, Schulaufsicht und Eltern von aktuellen bzw. zukünftigen Schülerinnen und Schülern.

2.2 Folgen der Schulinspektion für „Schulen mit Entwicklungsbedarf“

Die Schulinspektion ist durch das Erfüllen ihrer Rechenschaftsfunktion vermutlich die einzige Institution, die auf einer breiten Kriterienbasis Schulen mit besonders schwacher Unterrichts- und Schulqualität identifizieren kann. Dabei nimmt das Feststellen eines „Entwicklungsbedarfs“ eine besondere Stellung ein, da eine solche Diagnose nur erfolgt, wenn zahlreiche Indikatoren als „schwach ausgeprägt“ bewertet werden. Die genauen Kriterien, wann Entwicklungsbedarf diagnostiziert wird, unterscheiden sich zwischen den Systemen der Bundesländer (und auch international). Unabhängig davon ist es das generelle Ziel der Diagnose Schule mit Entwicklungsbedarf, die Schulen herauszuheben, die besondere Schwächen aufweisen, um durch geeignete Interventionen die Unterrichts- und Schulqualität sowie die Schülerleistungen rasch zu verbessern. Dazu gehört meist eine weiterführende Kontrolle durch

die Schulinspektion, zum Beispiel eine zeitlich vorgezogene Nachinspektion der Unterrichts- und Schulqualität an diesen Schulen (Tarkian et al. 2019). Dazu zählen auch verstärkte Unterstützungsangebote an die Schulen, wie beispielsweise durch Schulberaterinnen und Schulberater oder die Schulaufsicht (bspw. Senatsverwaltung für Bildung, Jugend und Wissenschaft 2014; Böhm-Kasper et al. 2016). Es ist bei diesen Schulen also von „verschärften Bedingungen“ (Dedering et al. 2016, S. 204) auszugehen. Zudem ist anzunehmen, dass die Wirkmechanismen der Schulinspektionsdiagnose in dieser besonderen Gruppe von Schulen verstärkt von internen Prozessen und der Art der Unterstützungsangebote und des externen Drucks abhängen (Ehren und Visscher 2006). Wie bei Ehren und Visscher (2006) beschrieben (Abb. 1), kann es durch den verstärkten Druck in Kombination mit der Reaktion der Schule zu positiven Entwicklungen hinsichtlich der Unterrichts- und Schulqualität sowie der Leistungen kommen, aber auch zu nichtintendierten Nebeneffekten. Wie aus der Literatur zu datenbasierter Steuerung bekannt (Datnow und Hubbard 2016; Schildkamp et al. 2012), kann Druck nur unter spezifischen schulinternen Bedingungen zu positiven Wirkungen führen. Betrachtet man den Forschungsstand speziell für Schulen mit Entwicklungsbedarf, findet man bislang nur wenige Studien, die Wirkungen der Schulinspektionsdiagnose auf (1) die Unterrichts- und Schulqualität, (2) die Leistungsentwicklung der Schülerinnen und Schüler, oder (3) nichtintendierte Nebeneffekte für diese besondere Gruppe von Schulen untersuchten.

1. Die Berliner Schulinspektion konstatiert, dass in den Nachinspektionen deutliche Verbesserungen der Schulen mit Entwicklungsbedarf hinsichtlich ihrer Unterrichts- und Schulqualität zu beobachten sind (Senatsverwaltung für Bildung, Jugend und Wissenschaft 2014). In diesem Zusammenhang wird erwähnt, es entstehe durch die Diagnose ein externer Druck „auf Schule und Schulaufsicht, alle Anstrengungen zu mobilisieren, damit das Inspektionsergebnis nach zwei Jahren deutlich besser ausfällt“ (Senatsverwaltung für Bildung, Jugend und Wissenschaft 2014). Dies scheint auch für England zu gelten: So beschreiben zum Beispiel Matthews und Sammons (2005) eine Abnahme von schlechten Unterrichtsstunden als Folge der Einordnung in das special measure-Programm.
2. Weiterhin zeigte sich für niederländische Schulen die aufgrund ihres „hohen Risikos“ in einem risikobasierten Schulinspektionssystem häufiger inspiert wurden, dass sich die Leistungen von Schülerinnen und Schülern an diesen Schulen stärker verbesserten als an anderen Schulen (Ehren und Shackleton 2016).
3. Auf der anderen Seite können sich auch nichtintendierte Nebeneffekte der Schulinspektion verstärkt zeigen, insbesondere durch den Druck auf die Schulen mit Entwicklungsbedarf (Altrichter und Kemethofer 2015). Einerseits, da die Ergebnisse der Schulinspektion nicht ignoriert werden können, andererseits, da die schulinternen Voraussetzungen für eine sinnvolle Annahme der Ergebnisse und eine darauf basierende Weiterarbeit schwieriger sind als an anderen Schulen (Dedering et al. 2016). So zeigte sich in Niedersachsen, dass sich die pädagogischen Akteure an solchen Schulen häufig ungerecht behandelt fühlten. Der Fokus ihrer Arbeit lag dann häufig nicht auf einer Weiterentwicklung der Unterrichts- oder Schulqualität, sondern auf der Bearbeitung von Dokumenten (Dedering et al. 2016; Sommer 2010). Dies lag vermutlich auch daran, dass an diesen Schulen

nach Dederling et al. (2016) häufig das strategische Wissen bzw. die Strukturen fehlten (wie auch durch die Schulinspektion bescheinigt wurde), die erwarteten Veränderungen herbeizuführen. Dabei fühlten sich die pädagogischen Akteure an Schulen mit Entwicklungsbedarf auch häufig von der Schulaufsicht nicht in ausreichendem Maß unterstützt (Preuße et al. 2019).

Wie beschrieben, kann eine unerwünschte Nebenwirkung von Schulinspektion auch die Stigmatisierung der Schulen sein, die durch eine Veröffentlichung der Ergebnisse befördert wird (Böhm-Kasper et al. 2016; Ehren und Visscher 2006). Betrifft dies verstärkt Schulen mit Entwicklungsbedarf, würde dies nahelegen, dass sich auch Änderungen des Schulwahlverhaltens von Eltern gegenüber diesen Schulen zeigen. Eltern nutzen zumindest zur Wahl der weiterführenden Schule ihrer Kinder öffentliche Schulinspektionsergebnisse und Kennzahlen zur Schülerschaft (Böhm-Kasper et al. 2016; Clausen 2006; Jurczok 2019). Zum elterlichen Schulwahlverhalten in Zusammenhang mit Schulen mit der Schulinspektionsdiagnose „Entwicklungsbedarf“ liegen bislang jedoch keine Befunde vor.

3 Forschungsfragen

Die Diagnose Schule mit Entwicklungsbedarf geht mit gesteigerten Erwartungen an die Verbesserung von Unterrichts- und Schulqualität einher (Perryman 2010). Der Forschungsstand zu Wirkungen der Schulinspektion im Allgemeinen zeigt, dass in Folge der Schulinspektion geringe Verbesserungen der Unterrichts- und Schulqualität zu beobachten sind. Die Ergebnisse zur Leistungsentwicklung von Schülerinnen und Schülern sind nicht eindeutig. Schulinspektion kann wiederum auch nichtintendierte Nebeneffekte herbeiführen. Die internationalen Forschungsergebnisse speziell zu Schulen mit Entwicklungsbedarf zeigen tendenziell positivere Wirkungen (Ehren und Shackleton 2016; Matthews und Sammons 2005). Über Schulen mit Entwicklungsbedarf in Deutschland ist jedoch abgesehen von einzelnen qualitativen Studien zur Verarbeitung von Schulinspektionsergebnissen in Niedersachsen (Dederling et al. 2016; Dederling 2018; Sommer 2010) so gut wie nichts bekannt. Die Erwartungen der Politik an die Schulinspektionsdiagnose „Schule mit Entwicklungsbedarf“ sind (zumindest in Berlin) jedoch hoch (Senatsverwaltung für Bildung, Jugend und Wissenschaft 2014). Auf Basis der Daten aller Berliner Grundschulen sollen mit dieser Studie daher erstmals Forschungsfragen zu Veränderungen zentraler Schulmerkmale nach der Schulinspektionsdiagnose als Schule „mit Entwicklungsbedarf“ untersucht werden: Welche Veränderungen zeigen sich an Schulen mit Entwicklungsbedarf hinsichtlich (a) ihrer Unterrichts- und Schulqualität, (b) ihren Schulleistungen und (c) der Zusammensetzung ihrer Schülerschaft?

Die Diagnose Schule mit Entwicklungsbedarf ist in Berlin mit einem Bündel an Maßnahmen verbunden, die zahlreiche Wirkmechanismen auslösen können. Es ist zu erwarten, dass die Schulen durch den Druck, der durch die Veröffentlichung der Ergebnisse entsteht, großes Interesse haben, bei einer Nachinspektion Verbesserungen zu zeigen. Zudem kommt es zu einer verstärkten Kontrolle durch die Schulaufsicht, die fordert, Qualitätsdefizite abzubauen. Hierzu können Schulen mit

Entwicklungsbedarf zusätzliche Unterstützung in Form von unterschiedlichen Beratungsangeboten in Anspruch nehmen (Senatsverwaltung für Bildung, Jugend und Wissenschaft 2014). In der Summe können diese Faktoren zu einer Veränderung der Schulpraxis und damit im Sinne der Schulinspektion zu einer Verbesserung der Unterrichts- und Schulqualität führen und darüber vermittelt zu besseren Schülerleistungen (Forschungsfragen a und b). Andererseits ist es denkbar, dass durch eine Stigmatisierung der Schulen (Ehren und Visscher 2006; Abb. 1) aufgrund der Diagnose einer Grundschule als Schule mit Entwicklungsbedarf, Eltern versuchen ihre Kinder (trotz der in Berlin geltenden Einzugsgebietsregelung) auf andere Grundschulen zu geben (Fincke und Lange 2012; Noreisch 2007). Dabei ist ausgehend von der Forschung zum Elternverhalten bei der Schulwahl für ihre Kinder anzunehmen, dass Eltern verstärkt versuchen, Schulen mit Entwicklungsbedarf zu meiden. Für die Suche nach der passenden weiterführenden Schule wird von der Senatsverwaltung explizit auf die öffentlichen Schulinspektionsergebnisse im Schulverzeichnis hingewiesen (Senatsverwaltung für Bildung, Jugend und Familie 2017) und Eltern nutzen solche Quellen bei der Wahl der weiterführenden Schule (Böhm-Kasper et al. 2016; Clausen 2006; Jurczok 2019). Es kann daher angenommen werden, dass Eltern solche Quellen auch für die „Wahl“ der Grundschule nutzen. Berücksichtigt man Ergebnisse zu Unterschieden der Schulwahl abhängig vom sozialen Hintergrund und der Migrationserfahrung der Familien (Jurczok 2019; Riedel et al. 2010), lässt sich vermuten, dass sich insbesondere Eltern aus bestimmten Gruppen gegen Schulen mit Entwicklungsbedarf entscheiden, und zwar Eltern, die sich *nicht* in sozialer, finanzieller oder bildungsbezogener Risikolage befinden. So wäre zu erwarten, dass sich durch die Diagnose Schule mit Entwicklungsbedarf die Zusammensetzung der Schülerschaft an diesen Schulen verändert und sich der Anteil von Kindern aus sozioökonomisch benachteiligten Familien oder Familien mit Migrationshintergrund erhöht (Forschungsfrage c).

4 Methode

4.1 Stichprobe Berliner Grundschulen

Die Schulinspektion in Berlin begutachtet seit 2005 etwa alle fünf Jahre alle öffentlichen Berliner Schulen. In der sogenannten zweiten Runde der Berliner Schulinspektion wurden zwischen 43 und 67 Grundschulen pro Jahr inspiziert (Tab. 1), die

Tab. 1 Anzahl an Schulinspektionen an Berliner Grundschulen 2011–2017

Diagnose der Schulinspektion	Schuljahr						Summe
	11/12	12/13	13/14	14/15	15/16	16/17	
Schulen ohne erheblichen Entwicklungsbedarf	55	38	51	63	43	61	311
Schulen mit erheblichem Entwicklungsbedarf	1	5	6	4	2	4	22
Summe	56	43	57	67	45	65	333

Tab. 2 Deskriptive Merkmale der Berliner Grundschulen ($N=333$)

	<i>N</i>	<i>M</i>	<i>SD</i>	Min	Max
Schulleistung ^a im Jahr der SI	333	0,06	0,97	-2,36	2,23
Schulleistung ^a ein Jahr nach SI	268	0,13	0,97	-2,56	2,57
Schulleistung ^a zwei Jahre nach SI	223	0,07	0,99	-2,37	2,08
Schulleistung ^a drei Jahre nach SI	155	0,08	1,01	-2,45	1,98
Nicht-deutsche Herkunftssprache (%) im Jahr der SI	333	40,53	28,09	0,31	98,59
Nicht-deutsche Herkunftssprache (%) ein Jahr nach SI	268	40,87	27,46	0,00	95,70
Nicht-deutsche Herkunftssprache (%) zwei Jahre nach SI	223	41,67	27,78	0,27	96,26
Nicht-deutsche Herkunftssprache (%) drei Jahre nach SI	156	39,98	27,25	0,61	96,11
SES (%) im Jahr der SI	333	37,76	25,29	0,26	98,79
SES (%) ein Jahr nach SI	268	36,61	24,52	2,19	91,98
SES (%) zwei Jahre nach SI	223	37,93	24,85	1,32	91,95
SES (%) drei Jahre nach SI	156	35,88	24,67	0,27	96,26

Min Minimum, *Max* Maximum, *SES* (%) 100 %- schul-spezifischer Anteil der Schülerinnen und Schüler, die lernmittelzuzahlungsbefreit sind (d. h. ein SES-Wert von 0 % steht für ein sehr niedriges Niveau des SES der Schülerschaft dieser Schule), *SI* Schulinspektion

^az-standardisierte mittlere Schulleistung

zu inspizierenden Schulen wurden in jedem Jahr zufällig ausgewählt. Die Analysen in dieser Studie basieren auf den Daten aller öffentlichen Berliner Grundschulen ($N=333$), die im Zeitraum zwischen 2011 und 2017 im Rahmen der zweiten Runde der Schulinspektion evaluiert wurden. Die deskriptiven Merkmale dieser Grundschulen sind in Tab. 2 dargestellt¹. Von diesen 333 Grundschulen wurden 22 öffentliche Grundschulen als Schulen mit Entwicklungsbedarf eingeordnet (6,6 % aller Grundschulen). An 10 der 22 Grundschulen fanden im Zeitraum bis 2017 die Nachinspektionen statt (2–3 Jahre nach der Regelinspektion im Rahmen der zweiten Runde). Eine von diesen nachinspizierten Schulen wies dabei wiederum Entwicklungsbedarf auf. Die anderen 12 Grundschulen werden im Rahmen der dritten Runde der Schulinspektion erneut inspiziert werden.

4.2 Verfahrensweise der Berliner Schulinspektion

Im Vorfeld der Schulinspektion werden von jeder Schule Dokumente angefordert (u. a. Gremienprotokolle und das schulinterne Curriculum), anschließend besucht ein Schulinspektionsteam aus mehreren Personen mehrere Tage die Schule, um sie zu evaluieren. Im Anschluss daran wird jeder Schule Rückmeldung zu den Evaluationsergebnissen in Form eines Berichts gegeben. Eine Kurzversion des Berichts wird im Internet im Rahmen des Schulprofils veröffentlicht. Der Bericht enthält eine Zusammenfassung der Stärken und Schwächen sowie das Qualitätsprofil mit

¹ Im Online Supplement sind die deskriptiven Merkmale zudem separat für Schulen mit und ohne Entwicklungsbedarf dargestellt.

den Bewertungen zu den einzelnen Indikatoren der Schulinspektion (Senatsverwaltung für Bildung, Wissenschaft und Forschung 2009). Wenn eine Schule Entwicklungsbedarf aufweist, der „weit über dem Durchschnitt liegt“ (Senatsverwaltung für Bildung, Wissenschaft und Forschung 2009), dann wird sie als „Schule mit erheblichem Entwicklungsbedarf“ eingestuft. Die Diagnose erfolgt in Berlin wenn Schulen eines oder mehrere der folgenden Kriterien aufweisen (Stand 2019, in Bezug auf Senatsverwaltung für Bildung, Wissenschaft und Forschung 2009): (1) Schlechte Ergebnisse in Schulleistungsuntersuchungen bleiben ohne schulische Konsequenzen, (2) Mängel im Schul- bzw. Konflikt- und Beschwerdemanagement, (3) für besondere schulspezifische Problemlagen werden keine geeigneten Gegenmaßnahmen ergriffen, (4) es bestehen gravierende Mängel im Personal- und Ressourcenmanagement, oder (5) es liegt eine durchwegs als schwach bewertete Unterrichtsqualität vor.

Hauptkonsequenz der Diagnose als Schule mit Entwicklungsbedarf ist, dass etwa zwei Jahre später eine vorzeitige „Nachinspektion“ der Schule stattfindet, um zu prüfen, ob sich die Unterrichts- und Schulqualität bedeutsam verbessert hat. Zudem wurde die prozessbegleitende Schulberatung „proSchul“ eingerichtet, um insbesondere Schulen mit Entwicklungsbedarf zu unterstützen. Ferner sollen Absprachen mit der Schulaufsicht getroffen werden, um zu dokumentieren, wie der festgestellte Entwicklungsbedarf überwunden werden kann (Senatsverwaltung für Bildung, Jugend und Wissenschaft 2012).

4.3 Instrumente

Die Analysen Berliner Grundschulen in dieser Studie basieren auf der Kombination der Daten der (a) Schulinspektion, (b) amtlichen Statistik sowie (c) Vergleichsarbeiten der dritten Jahrgangsstufe (VERA 3).

4.3.1 Unterrichts- und Schulqualität

Die Schulinspektion erfasst 19 Merkmale zu Unterrichts- und Schulqualität für alle Berliner Schulen. Diese Merkmale wurden von der Berliner Schulinspektion als Evaluationskriterien entwickelt und sind aus dem „Handlungsrahmen Schulqualität in Berlin“ (Senatsverwaltung für Bildung, Jugend und Wissenschaft 2013) abgeleitet. Ein Schulinspektionsteam besteht aus vier geschulten Personen. Bei der Schulinspektion werden zur Bewertung der Unterrichtsqualität mindestens 70 % aller Lehrkräfte jeweils 20 min in ihrem Unterricht besucht. Hieraus wird das „Unterrichtprofil der Schule“ erstellt, Einzelbewertungen auf Klassenebene werden nicht rückgemeldet. Die Bewertung der Schulqualitätsindikatoren basiert auf verschiedenen Datengrundlagen: Es erfolgen Interviews mit der Schulleitung, mit Schülerinnen und Schülern, mit Lehrkräften und mit Eltern, es werden Schuldokumente analysiert und schriftliche Fragebögen mit verschiedenen Personengruppen eingesetzt. Die Qualitätsindikatoren der Unterrichts- und Schulqualität werden jeweils auf einer 4-stufigen Bewertungsskala von 1 = „schwach ausgeprägt“, 2 = „eher schwach ausgeprägt“, 3 = „eher stark ausgeprägt“ und 4 = „stark ausgeprägt“ eingeschätzt. Die Ergebnisse von Reliabilitätsanalysen der Daten aus der Brandenburger Schulinspektion (Visitation) mit einem sehr ähnlichen Verfahren, legen nahe, dass die

Tab. 3 Unterrichts- und Schulqualität: Indikatoren der Schulinspektion

	<i>M</i>	<i>SD</i>	α	Qualitätsindikatoren der Schulinspektion	Beispielitem
<i>Unterrichtsqualität</i>					
Effiziente Klassenführung	3,78	0,13	0,64	Lehr-Lernzeit	Der Anteil an Warte- und Leerlaufzeiten für die Schüler/innen ist gering
				Verhalten der SuS im Unterricht	Schülerinnen und Schüler stören nicht den Unterricht
Kognitive Aktivierung	2,51	0,24	0,87	Methoden- und Medienwahl	Die Lehrkraft gestaltet den Unterricht anregend und motivierend
				Selbstständiges Lernen	Die Schüler/innen organisieren Lernprozesse/Unterrichts- bzw. Arbeitsabläufe selbstständig
				Problemorientiertes Lernen	Im Unterricht werden ergebnisoffene Frage- und Problemstellungen behandelt
				Leistungs- und Anstrengungsbereitschaft	Die Leistungsanforderungen sind herausfordernd
				Kooperatives Lernen	Teamorientierte Aufgabenstellungen werden im Unterricht gestellt und behandelt
Konstruktive Unterstützung	3,10	0,18	0,77	Pädagogisches Klima im Unterricht	Die Lehrkraft sorgt für eine angstfreie Lernatmosphäre
				Förderung von Selbstvertrauen und Selbsteinschätzung	Das Selbstvertrauen der Schüler/innen wird gefördert (z. B. Anerkennung, Ermunterung, Lob)
				Innere Differenzierung	Für die Schüler/innen bestehen Wahlmöglichkeiten entsprechend ihren Interessen und Neigungen
				Strukturierung und transparente Zielausrichtung	Die Unterrichtsschritte sind nachvollziehbar und klar strukturiert
<i>Schulqualität</i>					
Schulkultur	2,77	1,01	–	Fortschreibung des Schulprogramms	In der Schule wird nachvollziehbar an der Umsetzung der Schwerpunkte des Schulprogramms gearbeitet

Merkmale zur Unterrichts- und Schulqualität reliabel auf Schulebene erfasst werden können (Wurster und Gärtner 2013).

Die von der Schulinspektion evaluierten Merkmale wurden im Rahmen dieser Studie einschlägigen Dimensionen der Unterrichts- und Schulqualität zugeordnet (Tab. 3). In der Forschung zur Unterrichtsqualität besteht breiter Konsens, drei Basisdimensionen zu differenzieren: Effiziente Klassenführung, kognitive Aktivierung und konstruktive Unterstützung (Klieme et al. 2001; Kunter et al. 2011). In der Forschung zu Schulqualität werden häufig die folgenden Faktoren unterschieden (Ditton 2000; Scheerens und Bosker 1997): Schulkultur, Schulmanagement, Kooperation/Koordination, Personal.

Tab. 3 (Fortsetzung)

	<i>M</i>	<i>SD</i>	α	Qualitätsindikatoren der Schulinspektion	Beispielitem
Schulmanagement	3,23	0,67	0,69	Schulleitungshandeln und Schulgemeinschaft	Die Schulleiterin/der Schulleiter wird von den Lehrkräften in ihrem bzw. seinem Führungsverhalten anerkannt
				Schulleitungshandeln und Qualitätsmanagement	Die Schulleiterin/der Schulleiter sorgt für die Entwicklung einer schulspezifischen Steuerungsstruktur zur Qualitätsentwicklung und -sicherung
				Evaluation schulischer Entwicklungsvorhaben	Die Schule wählt aus ihren Entwicklungsvorhaben Schwerpunkte zur internen Evaluation aus
Kooperation/Koordination	3,31	0,52	0,52	Systematische Unterstützung, Förderung und Beratung	Die Schule stimmt sich im Hinblick auf Fördermaßnahmen mit internen und/oder externen Fachleuten ab
				Beteiligung der SuS und der Eltern	Die Schüler/innen beteiligen sich aktiv an der Schulentwicklung
				Arbeits- u. Kommunikationskultur im Kollegium	Es gibt klare Teamstrukturen, in die eine bedeutsame Anzahl des Kollegiums eingebunden ist
Personalentwicklung	2,84	0,85	–	Personalentwicklung und Personaleinsatz	Strategien zur Professionalisierung von Mitarbeiter/innen werden in konkreten Maßnahmen umgesetzt

Die Qualitätsindikatoren der Unterrichts- und Schulqualität werden auf einer 4-stufigen Bewertungsskala von 1= „schwach ausgeprägt“, 2= „eher schwach ausgeprägt“, 3= „eher stark ausgeprägt“, 4= „stark ausgeprägt“ durch die Schulinspektion eingeschätzt

4.3.2 Leistungen der Schülerinnen und Schüler

Als Indikator der Schulleistung einer Grundschule wurden die Ergebnisse der Schülerinnen und Schüler bei den Vergleichsarbeiten der dritten Jahrgangsstufe (VERA 3) genutzt. Die Vergleichsarbeiten basieren auf den deutschen Bildungsstandards, die Kompetenzen in verschiedenen Domänen (zum Beispiel Deutsch und Mathematik) definieren. Die Teilnahme an den Vergleichsarbeiten war für alle öffentlichen Berliner Grundschulen verpflichtend. Um einen Indikator der Schulleistung zu ermitteln, wurde zunächst der Mittelwert der Testergebnisse in Mathematik und Deutsch für jeden Schüler/jede Schülerin berechnet (Korrelationen der beiden Testwerte lagen auf Schülerebene zwischen $r=0,85$ und $r=0,91$). Die gemittelten individuellen Leistungswerte wurden über alle Schülerinnen und Schüler an einer Schule aggregiert, um für jede Schule einen Schulleistungswert zu erhalten. Schließlich wurden die aggregierten Schulleistungswerte für jedes Schuljahr jeweils z-standardisiert mit Mittelwert $M=0$ und $SD=1$. Da die Grundschulen in einem bestimmten Jahr zufällig von der Schulinspektion ausgewählt werden und nicht alle Grundschulen im gleichen Jahr inspiziert werden, erlaubt die z-Standardisierung einen Vergleich der Schulleistungen über die Jahre hinweg. Ein Schulleistungswert von 0 bedeutet bei-

spielsweise, dass das mittlere Leistungsniveau der Schülerinnen und Schüler an einer bestimmten Grundschule im Vergleich zu anderen Grundschulen in Berlin durchschnittlich ausgefallen ist. Ein Schulleistungswert von 1 bedeutet, dass die mittlere Schulleistung dieser Schule eine Standardabweichung über dem Durchschnitt aller anderen Berliner Grundschulen lag.

4.3.3 Zusammensetzung der Schülerschaft

Als Indikatoren der Zusammensetzung der Schülerschaft einer Schule dienen zwei Berlinspezifische Merkmale: die nicht-deutsche Herkunftssprache und die Lernmittelzuzahlungsbefreiung. Die nicht-deutsche Herkunftssprache dient als Indikator des Migrationshintergrunds, die aus der amtlichen Statistik stammt und bildet in einem Prozentanteil ab, in wie vielen Familien einer Schule die Verkehrssprache deutsch bzw. nicht deutsch ist (Senatsverwaltung für Bildung, Jugend und Familie 2020). Der Indikator reicht von 0–100 %: 0 % bedeutet, dass kein Schüler und keine Schülerin an der Schule in der Familie hauptsächlich eine andere Sprache als deutsch sprechen; 100 % bedeutet, dass alle Schülerinnen und Schüler der Schule in ihren Familien nicht deutsch sprechen. Die Lernmittelzuzahlungsbefreiung dient als Indikator für den sozioökonomischen Hintergrund (SES) der Schülerinnen und Schüler. Dieses Merkmal entstammt ebenso der amtlichen Statistik und erfragt, ob die Familie von der Zuzahlung zu Lernmitteln für die Schule befreit wurde. Eltern müssen dafür nachweisen, dass sie finanzielle staatliche Unterstützung bekommen. Damit hohe Werte einen hohen SES einer Schule abbilden, wurde ein Indikator erstellt, der sich wie folgt darstellt: 100 % abzüglich des schulspezifischen prozentualen Anteils der Schülerinnen und Schüler, die von der Lernmittelzuzahlung befreit wurden. Daher

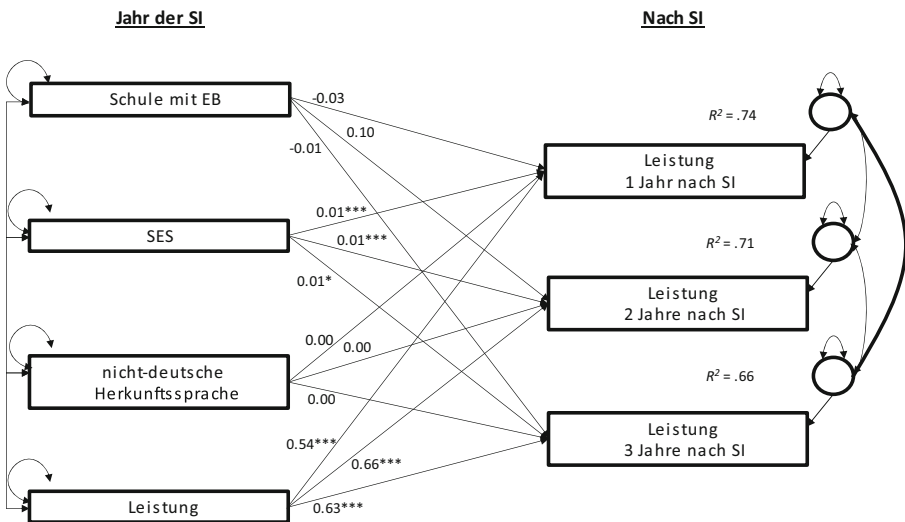


Abb. 2 Strukturgleichungsmodell mit abhängigen Variablen Leistung (unstandardisierte Koeffizienten) (Schule mit EB Schule mit Entwicklungsbedarf (kodiert= 1), Schule ohne Entwicklungsbedarf (kodiert= 0), *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$)

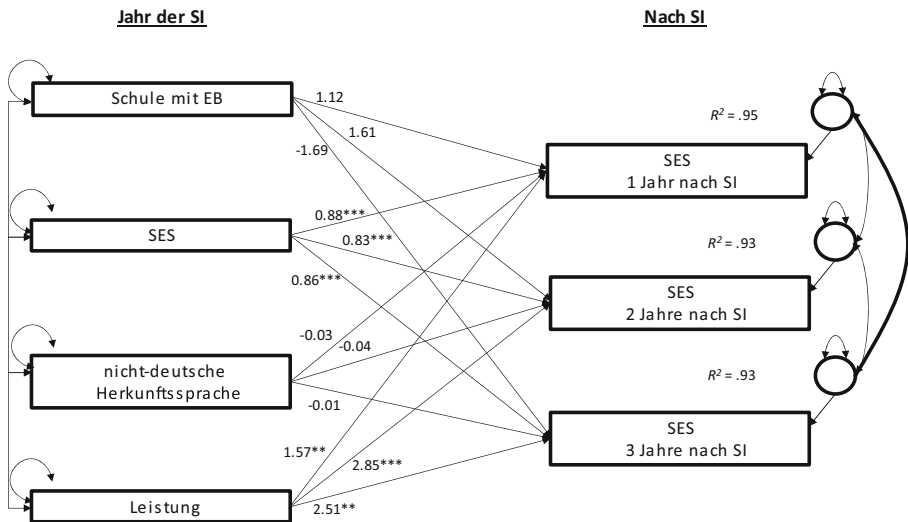


Abb. 3 Strukturgleichungsmodell mit abhängigen Variablen SES (unstandardisierte Koeffizienten) (Schule mit EB Schule mit Entwicklungsbedarf (kodiert= 1), Schule ohne Entwicklungsbedarf (kodiert= 0), *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$)

impliziert ein Wert des sozioökonomischen Hintergrunds von 0 %, dass alle Schülerinnen und Schüler dieser Schule von der Lernmittelzuzahlung befreit wurden, während ein Wert von 100 % bedeutet, dass an der Schule keine Schülerinnen und Schüler von der Lernmittelzuzahlung befreit wurden. In anderen Studien zeigte sich dieser Berlinspezifische Indikator als hoch korreliert mit anderen etablierten Indikatoren für SES (bspw. der Highest International Socio-Economic Index of Occupational Status [HISEI]; Maaz et al. 2016).

4.4 Vorgehen/Analysen

Die Untersuchung von Forschungsfrage (a) zur Veränderung von Aspekten der Unterrichts- und Schulqualität in Folge der Diagnose „erheblicher Entwicklungsbedarf“ erfolgte durch einen Vergleich der Werte der Regelinspektion mit denen der Nachinspektion an diesen Schulen. Für diese Forschungsfrage konnten daher lediglich Daten der 10 Schulen genutzt werden, die im Rahmen der zweiten Runde der Schulinspektion zweimal evaluiert wurden. Obwohl für Forschungsfrage (a) nur wenige Daten zur Verfügung stehen, soll dieser Forschungsfrage nachgegangen werden, um einen Eindruck von den Veränderungen in den Schulinspektionsergebnissen der Schulen mit Entwicklungsbedarf zu erlangen. Dafür wurden mit dem Programm ESCI (Cumming und Calin-Jageman 2017) neben Mittelwertdifferenzen auch standardisierte Mittelwertdifferenzen (Hedges g) sowie deren 95 %-Konfidenzintervalle für alle Unterrichts- und Schulqualitätsindikatoren berechnet. Dieser korrigiert für die Unterschätzung des Populationswerts bei kleinen Stichproben (Cumming und Calin-Jageman 2017).

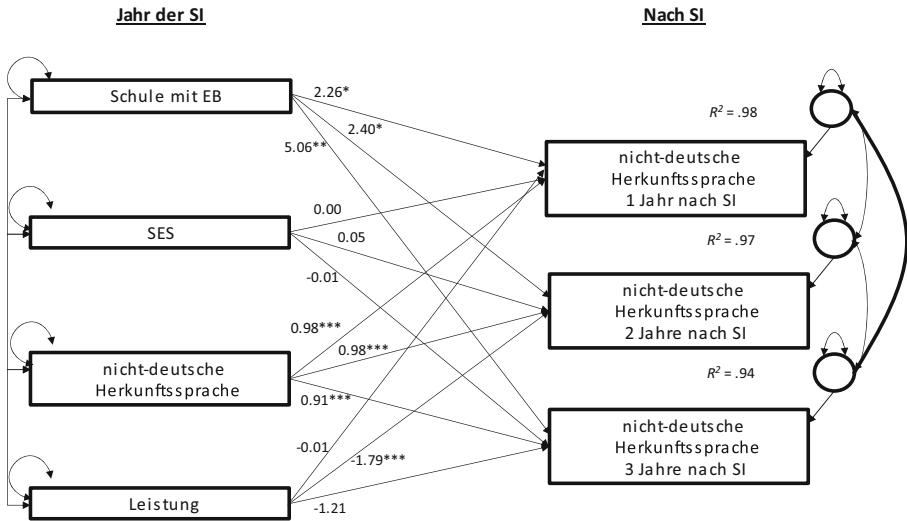


Abb. 4 Strukturgleichungsmodell mit abhängigen Variablen nicht-deutsche Herkunftssprache (unstandardisierte Koeffizienten) (*Schule mit EB* Schule mit Entwicklungsbedarf (kodiert= 1), Schule ohne Entwicklungsbedarf (kodiert= 0), *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$)

Die Beantwortung der Forschungsfragen zu Veränderungen (b) der Schulleistungen und (c) der Schülerzusammensetzung in Folge der Diagnose Schule mit Entwicklungsbedarf basierte auf den Daten von allen 333 Berliner Grundschulen. Für die Analysen wurden mittels lavaan (Rosseele 2012) drei Strukturgleichungsmodelle zur Vorhersage von drei Kriterien spezifiziert: (1) Schulleistung, (2) SES der Schülerschaft und (3) schulspezifischer Anteil der Kinder mit nicht-deutscher Herkunftssprache (s. Abb. 2, 3 und 4). Da der Zeitpunkt der Schulinspektion zwischen den Schulen variierte (Tab. 1), wurde die Zeitdimension relativ zum Zeitpunkt der Schulinspektion parametrisiert (Driver et al. 2017). In allen drei Modellen wurden hierzu die drei Kriterien ein Jahr nach der Schulinspektion, zwei Jahre nach der Schulinspektion, und drei Jahre nach der Schulinspektion vorhergesagt (in den Abb. 2, 3 und 4 unter „Nach SI“ dargestellt). Als Prädiktor diente eine Indikatorvariable (in den Abb. 2, 3 und 4 bezeichnet mit „Schule mit EB“), die angab, ob eine Schule mit Entwicklungsbedarf eingeordnet wurde (kodiert als 1) oder nicht (kodiert als 0). Zusätzlich wurden als Kovariaten der im Jahr der Schulinspektion ermittelte SES und der Anteil von Kindern mit nicht-deutscher Herkunftssprache sowie die Schulleistung im Modell hinzugefügt (in den Abb. 2, 3 und 4 unter „Jahr der SI“ dargestellt) um für diese bei der Vorhersage zu kontrollieren. Der Anteil fehlender Werte lag zwischen 0 und 53 %, durchschnittlich bei 12,6 %² (siehe Fallzahlen Tab. 2). Zur Schätzung aller Modellparameter wurde daher das „Full Information Maximum Likelihood“-Verfahren „MLR“ in lavaan genutzt (Enders 2010), um die

² Grund dafür ist, dass die Anzahl der verfügbaren Daten der Schulen mit jedem Jahr nach der Schulinspektion geringer werden, da diese dann für einen Teil der Schulen über das Ende der zweiten Welle der Schulinspektion hinausreichen und daher nicht zur Verfügung standen.

Standardfehler der Modellparameter dafür zu korrigieren, dass die Verteilungen der Residuen der abhängigen Variablen von einer Normalverteilung abweichen können.

5 Ergebnisse

Ein Vergleich der Mittelwertdifferenzen wies auf differenzielle Veränderungsprozesse der Unterrichts- und Schulqualität (Forschungsfrage a) hin (Tab. 4). Die stärksten positiven Veränderungen zeigten sich für die kognitive Aktivierung ($g=0,39$) und konstruktive Unterstützung ($g=0,41$). Nach Cohen (1988) sind diese Veränderungen als kleine bis mittlere Effekte einzuordnen. Im Bereich der Klassenführung war keine bedeutsame Veränderung festzustellen. Mit Blick auf die Schulqualität erhöhten sich die Werte für Kooperation/Koordination ($g=0,22$) und den Bereich Personal ($g=0,27$) leicht (kleiner Effekt). Bei den anderen beiden Qualitätsdimensionen (Schulkultur und -management) zeigten sich keine Veränderungen.

Abb. 5a zeigt die mittlere Leistung von Schulen mit bzw. ohne Entwicklungsbedarf. Es zeigt sich, dass Schulen mit Entwicklungsbedarf schon vor der Schulinspektion im Mittel niedrigere Leistungen erzielten als Schulen ohne Entwicklungsbedarf. Dieser Trend setzte sich auch nach der Schulinspektion fort. Die Leistungen von Schulen mit Entwicklungsbedarf sind zudem *im* Jahr der Schulinspektion niedriger als zuvor. Weiter zeigte sich, dass sich bereits vor der Schulinspektion die Schulen mit bzw. ohne Entwicklungsbedarf in der Zusammensetzung der Schülerschaft im Mittel unterschieden: Der mittlere SES war an Schulen mit Entwicklungsbedarf niedriger (Abb. 5b) und der Anteil von Kindern mit nicht-deutscher Herkunftsspra-

Tab. 4 Forschungsfrage a: Veränderung der Evaluationsergebnisse bei nachinspizierten Schulen ($N=10$)

	Regelinspektion <i>M (SD)</i>	Nachinspektion <i>M (SD)</i>	Mittelwertdifferenz (95 %-KI)	Standardisierte Mittelwertdifferenz Hedges <i>g</i> (95 %-KI)
<i>Unterrichtsqualität</i>				
Klassenführung	3,71 (0,15)	3,72 (0,17)	0,01 (-0,13; 0,14)	0,05 (-0,68; 0,78)
Kognitive Aktivierung	2,28 (0,26)	2,40 (0,29)	0,12 (-0,21; 0,45)	0,39 (-0,64; 1,47)
Konstruktive Unterstützung	2,89 (0,21)	3,01 (0,30)	0,12 (-0,19; 0,43)	0,41 (-0,61; 1,49)
<i>Schulqualität</i>				
Schulkultur	2,20 (1,14)	2,30 (1,16)	0,10 (-1,09; 1,29)	0,08 (-0,82; 0,98)
Schulmanagement	2,42 (0,93)	2,47 (1,03)	0,05 (-1,03; 1,13)	0,05 (-0,90; 1,00)
Kooperation/ Koordination	2,60 (0,54)	2,73 (0,58)	0,13 (-0,50; 0,76)	0,22 (-0,75; 1,21)
Personal	1,90 (0,99)	2,20 (1,03)	0,30 (-0,71; 1,31)	0,27 (-0,59; 1,17)

95 %-KI 95 %-Konfidenzintervall

Standardisierte Mittelwertdifferenz Hedges *g*: Nach Cohen (1988) steht $g=0,20$ für einen kleinen, $g=0,50$ für einen mittleren und $g=0,80$ für einen großen Effekt

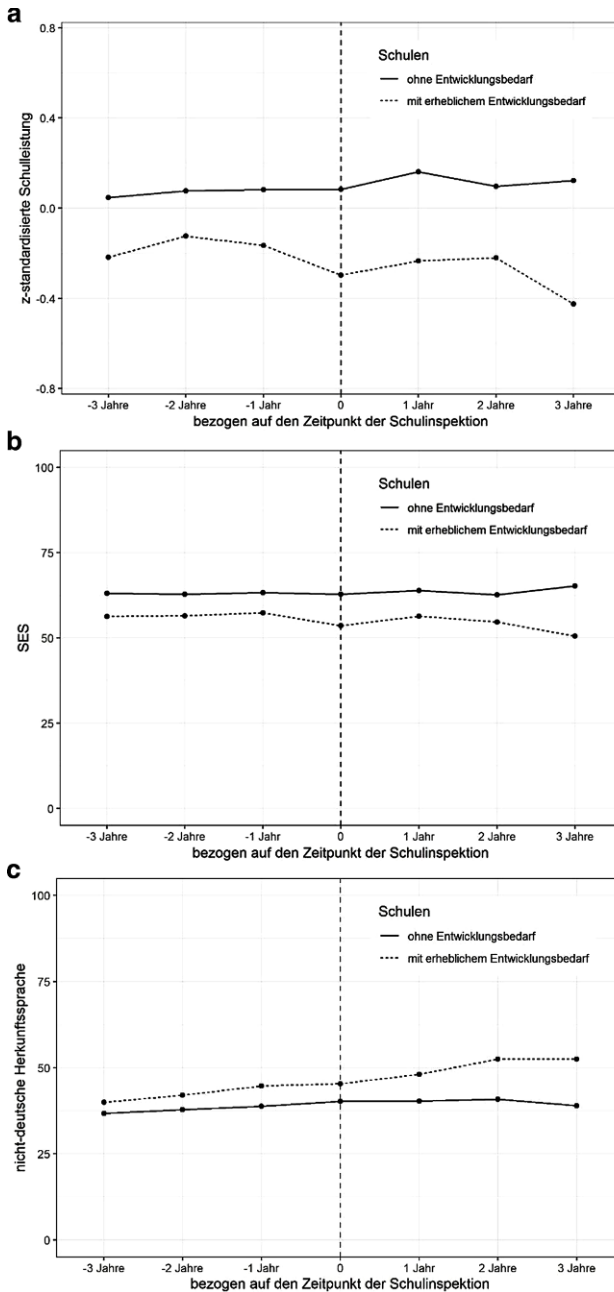


Abb. 5 Entwicklung der Schulen vor/nach der Schulinspektion: **a** Schulleistung (mittlere Schülerleistung aggregiert auf Schulebene), **b** SES (Prozentualer Anteil der Schülerinnen und Schüler ohne Lernmittelzahlungsbefreiung) und **(c)** nicht-deutsche Herkunftssprache (Prozentualer Anteil der Schülerinnen und Schüler mit nicht-deutscher Herkunftssprache)

Tab. 5 Forschungsfrage b und c, unstandardisierte Regressionskoeffizienten aus Strukturgleichungsmodellen

Prädiktoren aus dem Jahr der Schulinspektion	B 1 Jahr nach SI [95 %- KI]	B 2 Jahre nach SI [95 %- KI]	B 3 Jahre nach SI [95 %- KI]	Differenz B SI+1 – SI+3 [95 %-KI]	Differenz B SI+1 – SI+2 [95 %-KI]	Differenz B SI+2 – SI+3 [95 %-KI]
<i>AV1: Schulleistung</i>						
Schule mit EB	-0,03 [-0,24; 0,18]	0,10 [-0,26; 0,45]	-0,01 [0,55; 0,78]	-0,02 [-0,36; 0,33]	-0,12 [-0,52; 0,27]	0,11 [-0,28; 0,50]
SES	0,01*** [0,01; 0,02]	0,01*** [0,00; 0,02]	0,01* [0,00; 0,02]	0,00 [-0,01; 0,01]	0,00 [0,00; 0,01]	0,00 [-0,01; 0,01]
Nicht-deutsche Herkunftssprache	0,00 [-0,01; 0,00]	0,00 [0,00; 0,00]	0,00 [-0,01; 0,00]	0,00 [-0,01; 0,00]	0,00 [-0,01; 0,00]	0,00 [0,00; 0,01]
Leistung	0,54*** [0,44; 0,63]	0,66*** [0,55; 0,78]	0,63*** [0,47; 0,79]	-0,09 [-0,26; 0,07]	-0,13* [-0,24; -0,01]	0,03 [-0,11; 0,18]
–	$R^2 = 74\%$	$R^2 = 71\%$	$R^2 = 66\%$	–	–	–
<i>AV2: SES</i>						
Schule mit EB	1,12 [-0,89; 3,13]	1,61 [-2,00; 5,23]	-1,69 [-5,18; 1,79]	2,81 [-1,29; 6,92]	-0,50 [-4,13; 3,14]	3,31 [-0,91; 7,53]
SES	0,88*** [0,82; 0,94]	0,83*** [0,76; 0,89]	0,86*** [0,77; 0,94]	0,02 [-0,07; 0,11]	0,05 [-0,01; 0,11]	-0,03 [-0,10; 0,05]
Nicht-deutsche Herkunftssprache	-0,03 [-0,06; 0,00]	-0,04 [-0,08; 0,00]	-0,01 [-0,07; 0,06]	-0,02 [-0,08; 0,04]	0,01 [-0,03; 0,05]	-0,03 [-0,08; 0,02]
Leistung	1,57** [0,36; 2,78]	2,85*** [1,15; 4,55]	2,51** [0,66; 4,37]	-0,95 [-2,45; 0,56]	-1,28* [-2,47; -0,09]	0,34 [-1,01; 1,68]
–	$R^2 = 95\%$	$R^2 = 93\%$	$R^2 = 93\%$	–	–	–
<i>AV3: nicht-deutsche Herkunftssprache</i>						
Schule mit EB	2,26* [0,09; 4,42]	2,40* [0,03; 4,76]	5,06** [1,35; 8,78]	-2,81 [-6,90; 1,28]	-0,14 [-1,95; 1,67]	-2,67 [-5,98; 0,65]
SES	0,00 [-0,04; 0,04]	0,05 [0,00; 0,11]	-0,01 [-0,10; 0,08]	0,01 [-0,08; 0,10]	-0,05* [-0,11; 0,00]	0,07* [0,00; 0,13]
Nicht-deutsche Herkunftssprache	0,98*** [0,95; 1,01]	0,98*** [0,95; 1,01]	0,91*** [0,84; 0,97]	0,08* [0,01; 0,14]	0,00 [-0,03; 0,04]	0,07** [0,02; 0,12]
Leistung	-0,01 [-0,84; 0,82]	-1,79*** [-3,03; -0,55]	-1,21 [-3,16; 0,75]	1,19 [-0,70; 3,09]	1,78** [0,69; 2,87]	-0,58 [-2,00; 0,83]
–	$R^2 = 98\%$	$R^2 = 97\%$	$R^2 = 94\%$	–	–	–

B unstandardisierte Regressionskoeffizienten, *Schule mit EB* Schule mit Entwicklungsbedarf (kodiert = 1), Schule ohne Entwicklungsbedarf (kodiert = 0), *95 %-KI* 95 %-Konfidenzintervall, *SI* Schulinspektion, *AV* Abhängige Variable

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

che war höher (Abb. 5c). Auch diese Trends setzten sich nach der Schulinspektion fort.

Bei der Frage nach möglichen Veränderungen in Folge der Diagnose als Schule mit Entwicklungsbedarf wurden adjustierte Mittelwertunterschiede zwischen Schulen mit bzw. ohne Entwicklungsbedarf berechnet. Hierzu spezifizierten wir drei Strukturgleichungsmodelle für die drei abhängigen Variablen Leistung, SES und nicht-deutsche Herkunftssprache. Zur Adjustierung der Mittelwertunterschiede wurden in jedem dieser Strukturgleichungsmodelle für die vor der Schulinspektion bestehenden Unterschiede (hinsichtlich Leistung, SES und Anteil von Kindern mit nicht-deutscher Herkunftssprache) zwischen Schulen mit bzw. ohne Entwicklungsbedarf statistisch kontrolliert.

Bezüglich Veränderungen der Schulleistung (Forschungsfrage b) zeigten sich keine signifikanten (adjustierten) Mittelwertunterschiede der Schulen mit Entwicklungsbedarf gegenüber den Schulen ohne Entwicklungsbedarf (Abb. 5). Dies gilt sowohl für die Schulleistungen ein Jahr, zwei Jahre und drei Jahre nach der Schulinspektion. Die adjustierten Mittelwertunterschiede veränderten sich in ihrer Größe auch nicht im Zeitverlauf; entsprechende Differenzen zwischen den Regressionsgewichten waren statistisch nicht signifikant (Tab. 5, Zeile „Schule mit EB“ unter AV1: Schulleistung).

Zur Veränderung des SES der Schulen (Abb. 3) zeigt sich ebenfalls kein Unterschied zwischen Schulen mit und ohne Entwicklungsbedarf. Auch hier gilt dies für alle Zeitpunkte (ein Jahr, zwei Jahre und drei Jahre nach der Schulinspektion). Die Differenzen der Regressionsgewichte bei Schulen mit diagnostiziertem Entwicklungsbedarf unterscheiden sich im Zeitverlauf ebenfalls nicht signifikant (Tab. 5, Zeile „Schule mit EB“ unter AV2: SES).

Bezüglich der Veränderung des Anteils der Schülerinnen und Schüler mit nicht-deutscher Herkunftssprache in Folge der Diagnose Schule mit Entwicklungsbedarf (Abb. 4) zeigen sich signifikante Unterschiede: Relativ zu Schulen ohne Entwicklungsbedarf (mit vergleichbaren Ausgangsbedingungen bezüglich Schulleistung, SES und nicht-deutscher Herkunftssprache vor dem Zeitpunkt der Schulinspektion) stieg bei Schulen mit Entwicklungsbedarf der Anteil der nicht-deutschen Herkunftssprache in der Schülerschaft nach der Schulinspektion an. Dieser Effekt wird statistisch signifikant zum Messzeitpunkt ein Jahr, zwei Jahre und drei Jahre nach der Schulinspektion. Die Regressionsgewichte unterscheiden sich dabei nicht statistisch signifikant zwischen den Messzeitpunkten (Tab. 5, Zeile „Schule mit EB“ unter AV3: nicht-deutsche Herkunftssprache), was auf die zeitliche Stabilität dieses Zusammenhangs hinweist.

6 Diskussion

Die Schulinspektion evaluiert Schulen um Maßnahmen auszulösen, die zu einer Verbesserung der Unterrichts- und Schulqualität sowie der Leistungen der Schülerinnen und Schüler beitragen sollen. Gleichzeitig sind durch die Evaluation selbst sowie durch die Veröffentlichung der Ergebnisse auch nichtintendierte Nebenwirkungen möglich, zum Beispiel ein verändertes Schulwahlverhalten der Eltern in

Folge der Ergebnisse und damit verbunden eine Veränderung der Zusammensetzung der Schülerschaft. Von besonderer Relevanz sind hierbei diejenigen Schulen, für die im Rahmen der Schulinspektion Entwicklungsbedarf festgestellt wurde. Für diese Gruppe von Schulen ist national aber auch international wenig bekannt. In dieser Studie wurde daher erstmals explorativ untersucht, welche Veränderungen die Schulinspektionsdiagnose „Schule mit Entwicklungsbedarf“ nach sich ziehen kann, und zwar für Indikatoren der Unterrichts- und Schulqualität, der Schulleistung und der Zusammensetzung der Schülerschaft (SES und Anteil von Kindern mit nicht-deutscher Herkunftssprache).

6.1 Veränderung spezifischer Indikatoren der Berliner Grundschulen mit Entwicklungsbedarf in den Jahren nach der Schulinspektion

Mit der Diagnose Schule mit Entwicklungsbedarf ist die Vorstellung verbunden, dass sich die Identifizierung dieser Schulen durch die Rückmeldung und durch die daraus folgende Unterstützung positiv auf die weitere Entwicklung der Schulen, insbesondere in den Bereichen Unterrichts- und Schulqualität und Schulleistungen, auswirkt. Der Forschungsstand zu Folgen der Schulinspektion im Allgemeinen lässt keine großen Effekte erwarten, die Ergebnisse zu Veränderungen von Schulen mit Entwicklungsbedarf aus Großbritannien und den Niederlanden verweisen jedoch auf tendenziell positive Effekte (Ehren und Shackleton 2016; Matthews und Sammons 2005).

Bezüglich der Unterrichts- und Schulqualität in Schulen mit Entwicklungsbedarf (a) konnten in dieser Studie tendenziell positive Veränderungen in den Jahren nach der Schulinspektion aufgezeigt werden, bspw. in der kognitiven Aktivierung und der konstruktiven Unterstützung. Diese können als Anpassungen an die Erwartungen der Schulinspektion betrachtet werden. In der Klassenführung oder im Schulmanagement zeigten sich dagegen keine Veränderungen. Insgesamt kann auf Basis der geringen Datenbasis der vorliegenden 10 Schulen mit Entwicklungsbedarf jedoch nur eingeschränkt von einer tatsächlichen Verbesserung der Unterrichts- und Schulqualität gesprochen werden. Weiterhin ist zu bedenken, dass die Wirkung von Schulinspektion nach dem Modell von Ehren und Visscher (2006, Abb. 1) von vielen Bedingungen abhängig ist. Beispielsweise von Merkmalen des Schulinspektionsverfahrens und der Schule selbst, aber auch von externen Faktoren, die in dieser Studie nicht berücksichtigt werden konnten. Gemeinsam wirken diese Faktoren auf die Reaktion der Schulen: Sie sind möglicherweise unterschiedlich mit den Schulinspektionsergebnissen umgegangen, je nach Typ der Schule (aktiv, reaktiv, (selbst)-zufrieden, aktiv unzufrieden, passiv unzufrieden, Wurster und Gärtner 2013). Daraus entstehende differenzielle Entwicklungen können auf Basis der für diese Fragestellung zur Verfügung stehenden 10 Schulen nicht abgebildet werden. Weitere mögliche Erklärungen für ein Ausbleiben deutlicher Verbesserungen können Forschungserkenntnisse zur datenbasierten Steuerung von Bildungssystemen liefern, die zeigen, dass spezifische schulinterne Bedingungen vorliegen müssen, um positive Veränderungen im Sinne der Schulinspektion beobachten zu können (bspw. Vertrauen in die Daten, Fähigkeiten, mit diesen umzugehen, Zusammenarbeit im Kollegium

und Unterstützung von der Schulleitung, Datnow und Hubbard 2016; Schildkamp et al. 2012).

Die vorliegenden Ergebnisse zeigen, dass sich die Leistungen der Schülerinnen und Schüler (b) an Schulen mit Entwicklungsbedarf nicht positiver entwickeln als an allen anderen Grundschulen mit vergleichbaren Ausgangsbedingungen hinsichtlich Leistung oder Zusammensetzung der Schülerschaft (SES und nicht-deutscher Herkunftssprache). Die Diagnose Schule mit Entwicklungsbedarf kann die Entwicklung der Schulleistung für die drei Jahre nach der Schulinspektion nicht signifikant vorhersagen. Das bedeutet, die Diagnose zeigt keine signifikante Veränderung im verwendeten Indikator der Schulleistung. Eine Ursache dafür kann darin liegen, dass alle Schulen ein Feedback durch die Schulinspektion bekommen, welches Veränderungen auslösen könnte und sich Leistungsveränderungen in Schulen mit Entwicklungsbedarf nicht deutlich von anderen Schulen unterscheiden. Eine weitere Möglichkeit ist, dass eine Veränderung der Schulleistungen in Folge der Diagnose auch nach drei Jahren noch nicht sichtbar wird, sondern es eine längere Zeit dauert, bis sich entsprechende Veränderungsmechanismen in der Schule sichtbar auswirken. Schließlich ist zu fragen, ob die positiven Wirkerwartungen realistisch waren: Möglicherweise hat allein die Diagnose zu wenig Effekt und die verbundenen Unterstützungsangebote werden von den Schulen zu wenig genutzt. Dies bringt die Frage hervor, welche Unterstützungsmaßnahmen für eine positive Entwicklung relevant sind. Studien zu „failing schools“/Schulen mit großen Schwierigkeiten liefern Hinweise, wie positive Veränderungsprozesse trotz schwieriger Bedingungen initiiert werden können, bspw. durch den Fokus auf die Verbesserung des Unterrichts und eine starke Führung sowie eine eigenständige Schulentwicklungsplanung der Schulen mit Unterstützung von Externen (bspw. Meyers und Murphy 2007; Muijs et al. 2004).

Veränderungen der Schülerzusammensetzung (c) in Folge der Schulinspektionsdiagnose „Schule mit Entwicklungsbedarf“ wurden in dieser Studie betrachtet, um Hinweise darauf zu erlangen, inwieweit die Diagnose nichtintendierte Nebeneffekte im Bereich der elterlichen Schulwahl nach sich zieht: Die Annahme, dass Eltern mit höherem SES bei der Schulwahl eher Schulen meiden, die Entwicklungsbedarf aufweisen und sich dadurch die Schülerzusammensetzung verändern würde, konnte hier nicht bestätigt werden.

Bezüglich des Anteils der nicht-deutschen Herkunftssprache an den Schulen zeigten Schulen mit Entwicklungsbedarf jedoch einen signifikanten Anstieg um 2,3 % (ein Jahr später) bis 5,1 % (3 Jahre nach der Schulinspektion) pro Schule. Unter der Annahme, dass es nicht zu differenziellen, migrationsbezogenen Zu- und Abgängen von Schülerinnen und Schülern in den Grundschulen kam, bedeutet das für Schülerinnen und Schüler der ersten Jahrgangsstufe, dass bei durchschnittlich 50 Kindern pro Jahrgangsstufe 1, zwischen 4 (1 Jahr später) und 8 Kindern (3 Jahre später) mehr als zum Zeitpunkt der Schulinspektion eine andere als die deutsche Sprache in ihren Familien sprechen. Da die Schulinspektionsberichte im Internet veröffentlicht werden, lässt dieses Ergebnis vermuten, dass sich durch die Diagnose der Schulinspektion die Schülerschaft ändern könnte und insbesondere Eltern ohne Migrationshintergrund andere Schulen bevorzugen könnten. Dieser Befund geht mit Ergebnissen zur elterlichen Schulwahl einher, dass Eltern sich je nach Migrations-

hintergrund in ihrer Schulwahl unterscheiden (bspw. Riedel et al. 2010). Es könnte also durch die Schulinspektion zu einer stärkeren Segregation der Schülerschaft kommen, was die Schwierigkeiten an diesen Schulen beispielsweise hinsichtlich der Leistungsentwicklung zusätzlich erhöhen könnte (van Ewijk und Slegers 2010; Mickelson et al. 2013). Diesen Schwierigkeiten müsste die Schule dann zusätzlich begegnen, was wiederum Ressourcen der Schule in Anspruch nimmt. Diese möglichen Wirkungen der Schulinspektionsdiagnose auf die Schülerschaft sollten daher in den Unterstützungsangeboten an diese Schulen berücksichtigt werden.

6.2 Grenzen der Studie

Zur Untersuchung der Entwicklung von Schulen mit Entwicklungsbedarf analysierte diese Studie erstmals einen umfangreichen Datensatz aus verschiedenen Quellen zu allen öffentlichen Berliner Grundschulen. Dennoch bestehen Grenzen der Aussagekraft der vorliegenden Studie. (1) Möglicherweise werden positive Entwicklungsprozesse durch die hier verwendeten Variablen nicht sichtbar, da sich andere Outcome-Variablen verändert haben: Beispielsweise könnten sich das Wohlbefinden der Schülerinnen und Schüler sowie der Lehrkräfte durch angestoßene Prozesse nach der Schulinspektion verbessert haben, die Leistungen bleiben aber (noch) unverändert. Zwei weitere Punkte schließen sich hier an: Ab wann kann man von stattfindender Veränderung sprechen? Und wann ist der richtige Zeitpunkt, um die Entwicklung zu messen? Als Zeitpunkt für die weiteren Entwicklungen haben wir drei Messzeitpunkte nach der Schulinspektion gewählt (ein Jahr, zwei Jahre, drei Jahre danach). (2) Die Betrachtung von Veränderungen der Unterrichts- und Schulqualität (a) kann leider nur für die Schulen mit Entwicklungsbedarf erfolgen ($N=10$), die in der zweiten Welle der Schulinspektion bereits nachinspiziert wurden. Zudem kann hier nicht mit Schulen ohne Entwicklungsbedarf verglichen werden, da dort erst in der nächsten Regelinspektion (fünf Jahre später) wieder Daten erhoben werden. (3) Es liegen keine konkreten Angaben aus den Schulen vor, welche Unterstützungsleistungen sie nach der Schulinspektion nutzen und für wie wirksam sie die Unterstützung wahrnahmen. Dies kann jedoch die Reaktion der Schulen auf die Schulinspektionsergebnisse und folglich auch die Wirkung der Schulinspektionsdiagnose bedingen (Ehren und Visscher 2006). In weiteren Studien sollten die Unterstützungsangebote, die Schulen mit diagnostiziertem Entwicklungsbedarf gemacht werden, sowie deren Wirkungen gezielter in den Blick genommen werden. (4) Daten der Schulinspektion werden in dieser Studie für den Zweck der empirischen Bildungsforschung verwendet. Dadurch werden auch die durch die Schulinspektion gewählten und normativ gesetzten Vorstellungen von Unterrichts- und Schulqualität übernommen. Dies bringt verschiedene Schwierigkeiten mit sich, wie beispielsweise die theoretische Zuordnung der Schulinspektionsindikatoren zu wissenschaftlichen Konstrukten der Unterrichts- und Schulqualität. Weiterhin wurde die Übereinstimmung der Beobachtungen des Unterrichts durch externe Personen sowie deren Reliabilität bereits in einigen bisherigen Studien diskutiert. Dennoch zeigten sich diese Aspekte der Schulinspektionsdaten überwiegend als angemessen (Müller und Pietsch 2011; Pietsch und Tosana 2008; Wurster und Gärtner 2013). Weiterhin ermöglichen Schulinspektionsdaten Zugang zu Informationen zu Unterrichts- und

Schulqualität in einer Breite, die sonst kaum erreichbar wäre (Artelt et al. 2019). Nichtsdestotrotz wäre eine Vertiefung der hier aufgezeigten Fragestellungen mit Hilfe anderer Daten hilfreich und nützlich, um die Generalisierbarkeit der Befunde zu untersuchen. (5) Die Verknüpfung der elterlichen Schulwahl der Grundschule mit der Diagnose Schule mit Entwicklungsbedarf wurde bislang noch nicht untersucht. Sie erscheint uns jedoch aufgrund der in Abschn. 2.2 bzw. 3 angeführten Logik und angesichts der Tatsache, dass die Schulinspektionsergebnisse veröffentlicht werden, als hochrelevant. Um diese Frage statistisch zu analysieren, haben wir mit Rückgriff auf einschlägige methodische Empfehlungen (Hallberg et al. 2018; Steiner et al. 2010) drei Kovariaten gewählt (Leistung, SES, nicht-deutsche Herkunftssprache), die mit den untersuchten Ergebniskriterien (substanziell) assoziiert sind. Insbesondere gibt es in der kausalanalytischen Forschung mit nicht-experimentellen Daten starke empirische Evidenz dafür, dass die Nutzung von Daten zur Ausgangslage („Prätest“) Verzerrungen durch Selektionseffekte deutlich reduzieren kann (Hallberg et al. 2018). Damit geben die vorliegenden Ergebnisse erste empirische Hinweise auf mögliche Veränderungen der Schülerschaft in Form von Segregationsprozessen in Folge der Diagnose als Schule mit Entwicklungsbedarf. Dennoch sollte diese Forschungsfrage in zukünftigen Studien weiter untersucht werden, um die Robustheit und insbesondere die kausale Belastbarkeit dieses Befunds zu untersuchen sowie um die Hintergründe des Befunds näher zu beleuchten. (6) Schließlich wurde in dieser Studie ausschließlich die Situation in Berlin dargestellt. Inwiefern die Ergebnisse auf andere Bundesländer generalisiert werden können, ist durch die größere Konkurrenz zwischen Schulen in einem Stadtstaat, sowie durch die öffentliche Verfügbarkeit der Schulinspektionsberichte in Berlin eine offene Forschungsfrage.

6.3 Schlussfolgerungen

Mit Hilfe der Kombination verschiedener Daten zu 333 Berliner Grundschulen aus den Jahren 2011 bis 2017 konnte erstmals die Veränderung bedeutsamer Schulmerkmale von Schulen mit Entwicklungsbedarf untersucht werden. Es konnten Hinweise auf erwartete positive Wirkungen der Schulinspektion, aber auch Nebeneffekte aufgezeigt werden. Nach der Theorie, dass zusätzlicher Druck durch die Schulinspektion positive Entwicklungen nach sich zieht, müsste man vermehrte Handlungen im Anschluss an die Schulinspektion erwarten, die sich auch auf die Leistungen der Schülerinnen und Schüler auswirken. Unsere Studie gab erste Hinweise darauf, dass sich positive Veränderungen im Sinne der Schulinspektion im Bereich der Unterrichtsqualität abzeichnen. Dass sich die Schülerleistung an Schulen mit Entwicklungsbedarf im Anschluss an eine Schulinspektion besser entwickelt als an anderen Schulen, kann nicht nachgewiesen werden. Stattdessen deuteten die Ergebnisse auf eine zunehmende Segregation zwischen Schulen hin: Bei vergleichbarer Ausgangslage der Schulleistung und Zusammensetzung der Schülerschaft nahm an Schulen mit diagnostiziertem Entwicklungsbedarf der Anteil an Schülerinnen und Schülern mit nicht-deutscher Herkunftssprache in den ersten drei Jahren nach der Schulinspektion zu. Dies legt den Schluss nahe, dass Eltern ohne Migrationshintergrund Schulen mit Entwicklungsbedarf eher meiden.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s11618-022-01066-3>) enthalten.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Interessenkonflikt M. Wenger, H. Gärtner und M. Brunner geben an, dass kein Interessenkonflikt besteht.

Literatur

- Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, 26(1), 32–56.
- Artelt, C., Bug, M., Kleinert, C., Maaz, K., & Runge, T. (2019). Nutzungspotenziale amtlicher Statistik in der Bildungsforschung. Ein Überblick zu Erreichtem, möglichen Chancen und anstehenden Herausforderungen. In D. Fickermann & H. Weishaupt (Hrsg.), *Bildungsforschung mit Daten der amtlichen Statistik* (Die Deutsche Schule: Beiheft 14, S. 21–37). Münster: Waxmann.
- Bellmann, J., & Weiß, M. (2009). Risiken und Nebenwirkungen Neuer Steuerung im Schulsystem. Theoretische Konzeptualisierung und Erklärungsmodelle. *Zeitschrift für Pädagogik*, 55(2), 286–308.
- Böhm-Kasper, O., Selders, O., & Lambrecht, M. (2016). Schulinspektion und Schulentwicklung – Ergebnisse der quantitativen Schulleitungsbefragung. In Arbeitsgruppe Schulinspektion (Hrsg.), *Schulinspektion als Steuerungsimpuls? Ergebnisse aus Forschungsprojekten* (Educational governance, Bd. 25, S. 1–50). Wiesbaden: Springer VS.
- Brimblecombe, N., Shaw, M., & Ormston, M. (1996). Teachers' intention to change practice as a result of Ofsted school inspections. *Educational Management & Administration*, 24(4), 339–354.
- Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, 93(9), 1045–1057.
- Clausen, M. (2006). Warum wählen Sie genau diese Schule? Eine inhaltsanalytische Untersuchung elterlicher Begründungen der Wahl der Einzelschule innerhalb eines Bildungsgangs. *Zeitschrift für Pädagogik*, 52(1), 69–90.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Erlbaum.
- Cullingford, C. (1999). *An inspector calls*. London: Routledge.
- Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the new statistics. Estimation, open science, and beyond*. New York: Routledge.
- Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: a literature review of international research. *Journal of Educational Change*, 17(1), 7–28.
- Dederig, K. (2012). Schulinspektion als wirksamer Weg der Systemsteuerung? *Zeitschrift für Pädagogik*, 58(1), 69–88.
- Dederig, K. (2018). Consultancy in 'failing schools': emerging issues. *Improving Schools*, 21(2), 141–157.
- Dederig, K., Katenbrink, N., Schaffer, G., & Wischer, B. (2016). Veränderung unter Druck. In Arbeitsgruppe Schulinspektion (Hrsg.), *Schulinspektion als Steuerungsimpuls? Ergebnisse aus Forschungsprojekten* (Educational governance, Bd. 25, S. 201–226). Wiesbaden: Springer VS.

- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. In A. Helmke, W. Hornstein & E. Terhart (Hrsg.), *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule* (Zeitschrift für Pädagogik: Beiheft 41, S. 73–92). Weinheim: Beltz.
- Döbert, H., & Dederich, K. (2008). Externe Evaluation von Schulen in Deutschland. Die Konzepte der Bundesländer, ihre Gemeinsamkeiten und Unterschiede. In H. Döbert (Hrsg.), *Externe Evaluation von Schulen. Historische, rechtliche und vergleichende Aspekte* (S. 63–151). Münster: Waxmann.
- Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, 77(5), 1–35.
- Ehren, M. C. M., & Shackleton, N. (2016). Risk-based school inspections: impact of targeted inspection approaches on Dutch secondary schools. *Educational Assessment, Evaluation and Accountability*, 28(4), 299–321.
- Ehren, M. C. M., & Swanborn, M. S. L. (2012). Strategic data use of schools in accountability systems. *School Effectiveness and School Improvement*, 23(2), 257–280.
- 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.
- Ehren, M. C. M., & Visscher, A. J. (2008). The relationship between school inspections, school characteristics and school improvement. *British Journal of Educational Studies*, 56(2), 205–227.
- Enders, C. K. (2010). *Methodology in the social sciences: applied missing data analysis*. New York: Guilford.
- van Ewijk, R., & Sleegers, P. J. C. (2010). The effect of peer socioeconomic status on student achievement: a meta-analysis. *Educational research review*, 5, 134–150.
- Faubert, V. (2009). *School evaluation: current practices in OECD countries and a literature review*. Bd. 42.
- Fincke, G., & Lange, S. (2012). Segregation an Grundschulen: Der Einfluss der elterlichen Schulwahl. https://www.svr-migration.de/wp-content/uploads/2014/11/Segregation_an_Grundschulen_SVR-FB_WEB.pdf. Zugegriffen: 19. Feb. 2020.
- Gärtner, H., & Pant, H. A. (2011). How valid are school inspections? Problems and strategies for validating processes and results. *Studies in Educational Evaluation*, 37(2/3), 85–93.
- Gärtner, H., Hüsemann, D., & Pant, H. A. (2009). Wirkungen von Schulinspektion aus Sicht betroffener Schulleitungen. Die Brandenburger Schulleiterbefragung. *Empirische Pädagogik*, 23(1), 1–18.
- Gärtner, H., Wurster, S., & Pant, H. A. (2014). The effect of school inspections on school improvement. *School Effectiveness and School Improvement*, 25(4), 489–508.
- Hallberg, K., Cook, T. D., Steiner, P. M., & Clark, M. H. (2018). Pretest measures of the study outcome and the elimination of selection bias: evidence from three within study comparisons. *Prevention Science*, 19(3), 274–283.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297–327.
- Jurczok, A. (2019). *Schulwahl unter „gleichwertigen“ Einzelschulen*. Wiesbaden: Springer.
- Kemethofer, D., & Helm, C. (2017). Effekte durch Rechenschaftsdruck im Kontext von Schulinspektionen: Ein Vergleich von Schweden und Österreich. In M. Pietsch & I. Hosenfeld (Hrsg.), *Inspektionsbasierte Schul- und Unterrichtsentwicklung* (Empirische Pädagogik, Bd. 2, S. 138–157). Landau in der Pfalz: Verlag Empirische Pädagogik e. V.
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur“ und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). Bonn: Bundesministerium für Bildung und Forschung.
- Kotthoff, H.-G., & Böttcher, W. (2009). Neue Formen der „Schulinspektion“: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. Altrichter (Hrsg.), *Neue Steuerung im Schulsystem. Ein Handbuch* (Educational governance, Bd. 7, S. 295–325). Wiesbaden: VS.
- Kunter, M., Baumert, J., Blum, W., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Landwehr, N. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation. In C. Quesel (Hrsg.), *Wirkungen und Wirksamkeit der externen Schulevaluation* (S. 35–70). Bern: Hep.
- Lee, J. (2006). Input-guarantee versus performance-guarantee approaches to school accountability: cross-state comparisons of policies, resources, and outcomes. *Peabody Journal of Education*, 81(4), 43–64.
- Luginbuhl, R., Webbink, D., & de Wolf, I. (2009). Do inspections improve primary school performance? *Educational Evaluation and Policy Analysis*, 31(3), 221–237.

- Maaz, K., Böse, S., & Neumann, M. (2016). BONUS-Studie. Wissenschaftliche Begleitung und Evaluation des Bonus-Programms zur Unterstützung von Schulen in schwieriger Lage in Berlin. Zwischenbericht über die erste Schulleiterbefragung aus dem Schuljahr 2013/2014.
- Matthews, P., & Sammons, P. (2005). Survival of the weakest: the differential improvement of schools causing concern in England. *London Review of Education*, 3(2), 159–176.
- v. Meyers, C., & Murphy, J. (2007). Turning around failing schools: an analysis. *Journal of School Leadership*, 17(5), 631–659.
- Mickelson, R. A., Bottia, M. C., & Lambert, R. (2013). Effects of school racial composition on K–12 mathematics outcomes. *Review of Educational Research*, 83(1), 121–158.
- Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas? A review of research evidence. *School Effectiveness and School Improvement*, 15(2), 149–175.
- Müller, S., & Pietsch, M. (2011). Was wir messen, wenn wir Unterrichtsqualität messen. Inter-Beurteilerübereinstimmung und -Reliabilität bei Unterrichtsbeobachtungen im Rahmen von Schulinspektion. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht* (S. 33–56). Münster: Waxmann.
- Noreisch, K. (2007). School catchment area evasion: the case of Berlin, Germany. *Journal of Education Policy*, 22(1), 69–90.
- Penninckx, M., Vanhoof, J., de Maeyer, S., & van Petegem, P. (2016). Explaining effects and side effects of school inspections: a path analysis. *School Effectiveness and School Improvement*, 27(3), 333–347.
- Perryman, J. (2010). Improvement after inspection. *Improving Schools*, 13(2), 182–196.
- Pietsch, M., & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 11(3), 430–452.
- Pietsch, M., Janke, N., & Mohr, I. (2014). Führt Schulinspektion zu besseren Schülerleistungen? Difference-in-Differences-Studien zu Effekten der Schulinspektion Hamburg auf Lernzuwächse und Leistungstrends. *Zeitschrift für Pädagogik*, 60(3), 446–470.
- Preuß, D., Pohl, J., & Gärtner, H. (2019). Wahrgenommene Auswirkungen der Schulinspektion aus Sicht von Schulleitungen und Schulaufsicht in Berlin. https://www.isq-bb.de/wordpress/wp-content/uploads/2019/09/Ergebnisbericht_Berlin_2019-07.pdf. Zugegriffen: 1. Okt. 2019.
- Reezigt, G. J., & Creemers, B. P. M. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement*, 16(4), 407–424.
- Riedel, A., Schneider, K., Schuchart, C., & Weishaupt, H. (2010). School choice in German primary schools: How binding are school districts? *Journal for Educational Research Online / Journal für Bildungsforschung Online*, 2(1), 94–120.
- Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review*, 23(2), 143–151.
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness* (1. Aufl.). Oxford: Pergamon.
- Schildkamp, K., Rekers-Mombarg, L. T. M., & Harms, T. J. (2012). Student group differences in examination results and utilization for policy and school development. *School Effectiveness and School Improvement*, 23(2), 229–255.
- Senatsverwaltung für Bildung, Jugend und Familie (2020). Blickpunkt Schule Bericht. Schuljahr 2019/2020. <https://www.berlin.de/sen/bildung/schule/bildungsstatistik/>. Zugegriffen: 25. Nov. 2020.
- Senatsverwaltung für Bildung, Jugend und Familie (2017). *Berliner Schulwegweiser. Wohin nach der Grundschule? Schuljahr 2018/2019*. Berlin: Senatsverwaltung für Bildung, Jugend und Familie.
- Senatsverwaltung für Bildung, Jugend und Wissenschaft (2012). *Berliner Schule. Zweite Runde Schulinspektion in Berlin*. Berlin: Senatsverwaltung für Bildung, Jugend und Wissenschaft.
- Senatsverwaltung für Bildung, Jugend und Wissenschaft (2013). *Handlungsrahmen Schulqualität in Berlin. Qualitätsbereiche und Qualitätsmerkmale*. Berlin: Senatsverwaltung für Bildung, Jugend und Wissenschaft.
- Senatsverwaltung für Bildung, Jugend und Wissenschaft (2014). *7 Jahre Schulinspektion in Berlin*. Berlin: Senatsverwaltung für Bildung, Jugend und Wissenschaft.
- Senatsverwaltung für Bildung, Wissenschaft und Forschung (2009). *Bildung für Berlin. Handbuch Schulinspektion*. Berlin: Senatsverwaltung für Bildung, Wissenschaft und Forschung.
- Shaw, I., Newton, D. P., Aitkin, M., & Darnell, R. (2003). Do OFSTED inspections of secondary schools make a difference to GCSE results? *British Educational Research Journal*, 29(1), 63–75.

- Sommer, N. (2010). Schulen mit „gravierenden Mängeln“. In W. Böttcher (Hrsg.), *Evaluation, Bildung und Gesellschaft. Steuerungsinstrumente zwischen Anspruch und Wirklichkeit* (S. 209–227). Münster: Waxmann.
- Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological methods*, 15(3), 250–267.
- Tarkian, J., Lankes, E.-M., & Thiel, F. (2019). Externe Evaluation – Konzeption und Implementation in den 16 Ländern. In F. Thiel, J. Tarkian, E.-M. Lankes, N. Maritzen, T. Riecke-Baulecke & A. Kroupa (Hrsg.), *Datenbasierte Qualitätssicherung und -entwicklung in Schulen. Eine Bestandsaufnahme in den Ländern der Bundesrepublik Deutschland* (S. 105–183). Wiesbaden: Springer VS.
- Wurster, S., & Gärtner, H. (2013). Schulen im Umgang mit Schulinspektion und deren Ergebnissen. *Zeitschrift für Pädagogik*, 59(3), 425–445.