
Björn Hermstein/Ina Semper/Nils Berkemeyer/Lisa Mende

Thematisierungen von Bildungsmonitoring- instrumenten seitens der Bildungsforschung

Die Beispiele Bildungsberichterstattung, Schulinspektion und Vergleichsarbeiten

Zusammenfassung

Die Etablierung von Bildungsmonitoring wird von der empirischen Bildungsforschung kontinuierlich begleitet. Unter welchen Gesichtspunkten einzelne Instrumente forschungsseitig thematisiert werden, steht im Fokus dieses Beitrags. Hierzu werden Forschungen zu drei zentralen Elementen des deutschen „Monitoring-Paradigmas“ systematisiert vorgestellt. Erkennbar wird unter anderem eine Kongruenz zwischen den Analysemodellen und Fragestellungen der Bildungsforschung und den politisch formulierten Zwecksetzungen.

Schlüsselwörter: Bildungsmonitoring, Empirische Bildungsforschung, Bildungsberichterstattung, Schulinspektion, Vergleichsarbeiten

The Exploration of Educational Monitoring Instruments by Educational Research

**Taking the Examples of Educational Reporting, School Inspection,
and State-wide Comparative Tests**

Abstract

Empirical educational research has played a continuous role in accompanying the implementation of educational monitoring. This article examines the issues that educational research has foregrounded when studying three key instruments of the new German 'educational monitoring paradigm'. Our findings suggest that both the research questions and the analysis models that empirical educational research draws upon are congruent with politically formulated targets.

Keywords: educational monitoring, empirical educational research, educational reporting, school inspection, state-wide comparative tests

Einleitung

Die Ereignisse und Diskussionen der vergangenen Jahre um und über die deutschen Schulsysteme können als Ausweis eines zuvor nicht denkbaren Reformeifers (vgl. van Ackeren/Block 2009) beschrieben werden. Diese Diagnose erscheint plausibel, vergewärtigt man sich die vielfältigen bildungspolitischen Maßnahmen und regulativen Strategien, die im wissenschaftlichen Diskurs beispielsweise als *Modernisierung* (vgl. Brüsemeister/Eubel 2003), *Neue Steuerung* (vgl. Altrichter/Maag Merki 2010) oder auch *Ergebnisorientierte Steuerung* (vgl. Wacker/Maier/Wissinger 2012) begrifflich gefasst werden. Schon diese Beschreibungsversuche der veränderten Verhältnisse im Schulsystem dokumentieren die hohe Aufmerksamkeit der Bildungsforschung für Verfahren, die veränderte Koordinationsstrukturen zwischen politischen, administrativen und pädagogischen Systemelementen bewirken sollen und die unter solche Begriffsschemata subsumiert werden können.

Ausgehend von der Annahme, dass auch die Forschung durch ihre Zugriffsweisen an der diskursiven (Re-)Konstruktion der Steuerungsstrategien und ihrer Elemente beteiligt ist (analog zu den Analysen von Barlösius 2005), soll anhand von Forschungsbeiträgen gezeigt werden, in welcher Weise Reformgeschehnisse seitens der Forschung thematisiert und welche Zugriffsweisen dabei fokussiert werden. Dies geschieht anhand einer Rezeption ausgewählter Fachpublikationen, die sich entlang der Elemente Bildungsberichterstattung, Schulinspektion und Vergleichsarbeiten auf drei zentrale Bestandteile des deutschen „Monitoring-Paradigmas“ (vgl. Böttcher 2013) bezieht.

Das Instrument Bildungsberichterstattung im Spiegel der deutschen Bildungsforschung

Bildungsberichte wurden in den letzten Jahren als ein zentrales Instrument des Bildungsmonitorings in Deutschland etabliert. Dies wird vor allem dadurch ersichtlich, dass sie sowohl auf Bundes- und Landes- als auch auf regionaler und kommunaler Ebene erscheinen. Als zentrales Ziel von Bildungsberichten gilt die regelmäßige und indikatorengestützte Informationsgenerierung über Rahmenbedingungen, Verlaufsmerkmale, Ergebnisse und Erträge von Bildungsprozessen. Bildungsberichte dienen dem Anspruch nach der Rechenschaftslegung über das Gesamtsystem, sollen mögliche Fehlentwicklungen frühzeitig aufzeigen und entsprechende politisch-

administrative Steuerungsmaßnahmen fundieren (vgl. Konsortium Bildungsberichterstattung 2005).

Vornehmlich sind *technisch-methodische Arbeiten zur indikatorenbasierten Darstellung* interessierender Merkmale der Bildungssysteme veröffentlicht worden (vgl. Weishaupt/Zimmer 2013). Vertiefende *Diskussionen zur Konzeption von Bildungsberichten* und Optimierung der verwendeten Indikatorensysteme werden mehrfach angestellt (vgl. Döbert et al. 2009; Weishaupt 2009; Döbert/Klieme 2010). Herausforderungen an eine indikatorengestützte Bildungsberichterstattung werden in der Weiterentwicklung der Datenbasen und Indikatorensysteme, deren konzeptioneller Fundierung, ihrer empirisch belastbaren Plausibilität, der Überprüfung von Annahmen über Wirkungszusammenhänge und Interdependenzen (vgl. Döbert 2008) sowie der Modellierung systemischer Gerechtigkeit (vgl. Berkemeyer/Manitius 2013) gesehen. Rürup, Fuchs und Weishaupt (2010) zufolge konnten, inhaltlich gesehen, die Konzepte von Bildungsqualität, die sich in den Bildungsberichten niederschlugen, bis dato nicht überzeugen (siehe die Arbeit von Scheerens/Hendriks 2004). Dadurch können die zu einzelnen Indikatoren repräsentierten Daten zwar auf ein umfassendes Systemmodell des Bildungswesens referieren, jedoch werden Zusammenhänge zwischen einzelnen Wirkbereichen kaum plausibilisiert. Infolgedessen werden Bildungsberichte dem postulierten Anspruch eines System-Modeling aktuell nicht gerecht. Aus diesem Grund sehen Rürup, Fuchs und Weishaupt die Voraussetzungen dafür, dass Bildungsberichte konkrete Orientierungen für politische Entscheidungen vermitteln können, derzeit nicht realisiert. Forschungsdesiderata erkennen sie hinsichtlich der Validierung von Datensätzen sowie vor allem der politischen Rezeption und Informationsverarbeitung. Ferner soll auch der Einfluss von Bildungsberichten auf die bildungspolitische Agenda und Entscheidungsprozesse systematischer erforscht werden. Erste Hinweise auf steuerungsrelevante Indikatorensysteme liefert eine Studie von Siepke, Tegge und Egger (2014) zu kommunalen Bildungsberichten. Sie zeigen etwa, dass eine Kontext-Input-Prozess-Output-Systematik die Struktur der Berichtssysteme dominiert.

Ähnliche Bedenken in Bezug auf die postulierte *Steuerungsfunktion von Bildungsberichten* äußern Niedlich und Brüsemeister (2012). Sie diagnostizieren bei der Bildungsberichterstattung die Vernachlässigung der Steuerungsseite zugunsten einer auf Transparenz und öffentliche Rezeption setzenden Ausrichtung (vgl. ebd., S. 133). Sie thematisieren Bildungsberichte aus governanceanalytischer Perspektive als Produkte politischer Aushandlungsprozesse. Hieran anknüpfend konstatieren Hermstein und Manitius (2015, im Erscheinen) den Bildungsberichten ein bislang wenig beachtetes Gerechtigkeitspotenzial, da Bildungsberichterstattung aufgrund der spezifischen konstellationsbezogenen und methodisch-technischen Anlage Anlässe für Diskurse bereithält.

Erste Forschungsansätze zur *Rezeption, Nutzung und Wirksamkeit von Bildungsberichten* ließen sich schon 2006 in Deutschland finden (vgl. Weishaupt 2006; Lüders 2006), wurden jedoch nicht systematisch weitergeführt. Einen umfassenden (auch Forschung aus dem Ausland berücksichtigenden) Überblick zu empirischen Befunden geben Rürup, Fuchs und Weishaupt (2010). Abendroth, Sendzik und Järvinen (2014) fokussieren jüngst die Schulleitungen als Adressaten von Berichten und fragen nach ihren Nutzenerwartungen. Über das Rezeptions- und Verwendungsverhalten bildungspolitischer Instanzen existiert derzeit kein systematisch generiertes Wissen. Das mag mit einem geringen Interesse an der Überprüfung des Nutzens von Bildungsberichten in Bezug auf bildungspolitische Zielsetzungen sowie deren Nutzung durch die Steuerungsakteure zusammenhängen (vgl. Gärtner 2015).

Trotz der breiten Etablierung findet eine empirische Erforschung der Bildungsberichterstattung bis dato kaum statt. Die Mehrzahl der Beiträge sind konzeptualisierender Art mit dem Fokus auf die technisch-methodische Weiterentwicklung und Optimierung der Berichtssysteme. Auch findet man häufiger Arbeiten, die Erwartungen und wünschenswerte Wirkungen von Bildungsberichten skizzieren (vgl. u.a. Hüfner 2006). Empirische Forschungen stellen die Ausnahme dar, insbesondere zu Thematisierungen, die außerhalb bildungspolitischer Erwartungen liegen.

Das Instrument Schulinspektion im Spiegel der deutschen Bildungsforschung

Die Schulinspektion ist in den Schulsystemen der Bundesländer seit Mitte der 2000er-Jahre das zentrale Verfahren der externen Evaluation von Einzelschulen hinsichtlich definierter Qualitätsstandards von Schulorganisation und Unterricht. Es gilt als ein wesentliches Moment evidenzbasierter Steuerung sowie des Systemmonitorings, welches derzeit aber nicht mehr in allen Bundesländern Anwendung findet.

Im Rahmen *systematisierender Bestandsaufnahmen deutscher Inspektionssysteme* wurden detailreiche Aufarbeitungen von Verfahrensbedingungen sowie Zielen und Zwecken erstellt (vgl. Bos et al. 2007). Auf Basis einer Befragung der Kultusministerien konnte Rürup (2008) etwa länderübergreifende Merkmale von Schulinspektion herausarbeiten sowie Typen von Strategien, Verfahrensarten und Funktionszuschreibungen (z.B. Gegenüberstellung von eher *wahrheits-* und eher *angemessenheitsorientierten* Verfahren) beschreiben.

Die Typisierungen von Inspektionsverfahren sind im Zusammenhang mit Beiträgen zu sehen, die *Bestimmungen von Funktionsweisen des Instruments Schulinspektion* aufweisen. Neben den politischen Akteuren beteiligt sich die Bildungsforschung ebenfalls daran, die potenziellen und erwarteten Verfahrensfunktionen zu konkre-

tisieren. Böttcher und Kotthoff (2010) unterscheiden grundsätzlich die Kontroll- und die Entwicklungsfunktion. Maritzen (2006) schreibt der Schulinspektion einen Funktionsmix zu, wozu auch die Impulsfunktion und die Erkenntnisfunktion gehören. Landwehr (2011) betont die Normendurchsetzungsfunktion. Die funktionellen Zuschreibungen sind auch, einmal mehr und einmal weniger explizit, Gegenstand empirischer Forschungen zur Schulinspektion. Dederling, Fritsch und Weyer (2012) zeigen, dass Schulleitungen und Lehrkräfte die Schulinspektion eher als Entwicklungs- denn als Kontrollinstrument wahrnehmen.

Die Funktionsbestimmungen strukturieren die Blickrichtung auf Schulinspektion als Forschungsgegenstand. Forschungen, die an vorgängig formulierte Zielsetzungen anschließen und somit als Wirksamkeitsforschung (vgl. Husfeldt 2011) bezeichnet werden, befassen sich einerseits mit dem *Ziel der Schul- und Unterrichtsentwicklung* und andererseits mit dem *Generalziel der Förderung von Schülerleistungen*. Hinsichtlich der entwicklungsbezogenen Zielsetzungen veranlasst die bisherige Befundlage Böttcher und Keune (2012) zu der Einschätzung, dass wenig Grund bestehe, der von politischer Seite behaupteten Schulentwicklung durch Inspektion Glauben zu schenken, obgleich sie ein dahingehendes Potential der Schulinspektion erkennen. Ergebnisse einer Studie von Gärtner, Hüsemann und Pant (2009) auf Basis einer Schulleiterbefragung stützen dieses Zwischenfazit, da sie keinen Zusammenhang zwischen den rückgemeldeten Inspektionsergebnissen und der Beschaffenheit der berichteten Entwicklungsmaßnahmen nachweisen konnten. In weiterführenden Studien konnten auf Grundlage einer Befragungsstudie diverse Reaktionsmuster auf Schulebene typisiert werden (vgl. Wurster/Gärtner 2013), was auf differenzielle Verarbeitungsweisen hinweist. Die Output-Kategorie Schülerleistungen wurde bislang, auch aufgrund von Schwierigkeiten im Hinblick auf Datenlage und Forschungsmethodik, eher selten im Zusammenhang mit der Schulinspektion betrachtet. Trotz der niedrigen Forschungsfrequenz ist offenbar die Position virulent, dass sich die Schulinspektion daran messen lassen muss, „ob sie zu positiven Auswirkungen auf das Lernen der Schüler und deren Kompetenzaufbau beiträgt.“ (Gärtner/Hüsemann/Pant 2009, S. 16) Obgleich Studien aus dem In- und Ausland eher keine dahingehenden Wirkungen festgestellt haben, ist es Pietsch, Janke und Mohr (2014) mit einem *Difference-in-Difference*-Ansatz gelungen, für das Hamburger Inspektionsmodell Hinweise auf Effekte hinsichtlich der Lernzuwächse und Leistungstrends zu finden.

Eine spezifische Form der Wirkungsforschung hat sich jüngst um das *Educational-Governance*-Paradigma formiert. Hierbei wird abseits der formulierten Wirkungshoffnungen nach *Folgen der Etablierung des Akteurs Schulinspektion auf Ebene der regionalen Governance-Strukturen* gefragt. Die Befundlage dieser Interviewstudien ist bisher als uneindeutig zu bezeichnen. Zwar werden die traditionellen Akteurskonstellationen, Interdependenzen und Koordinationen wohl nicht grundsätzlich berührt, doch gibt es Anzeichen für neue Bündnisbildungen zwischen Inspektion und

Schulleitung, wodurch die Schulleitungen dem Lehrerkollegium entrückt werden (vgl. Brüsemeister/Preuß/Wissinger 2014).

Seitens der Forschung werden *schulinterne Verarbeitungsprozesse* eher vernachlässigt (vgl. Husfeldt 2011). Neben Ansätzen der Überprüfung von Modellierungen der Wirkungsannahmen (vgl. Altrichter et al. 2014) liegen mittlerweile einzelne Forschungen vor, die die Prozesshaftigkeit und damit *Schulinspektion als soziale Praxis* thematisieren. Dazu gehören neben empirischen Studien zu Wahrnehmung und Akzeptanz der Schulinspektion, die bei Schulleitungen ausgeprägter zu sein scheint als bei Lehrkräften (vgl. Schwank/Sommer 2012) und eher nicht mit verfahrensmäßigen Variationen zusammenhängen (vgl. Böhm-Kasper/Selders 2013), auch Vorhaben, die auf die akteursbezogenen Tätigkeiten (vgl. Sowada/Dedering 2014) und Reaktionen (vgl. Dedering/Fritsch/Weyer 2012) sowie auf Interaktionen abstellen. Beispielsweise können Katenbrink und Schaffer (2015, im Erscheinen) anhand von Fallrekonstruktionen von als „failing schools“ bewerteten Schulen in Niedersachsen auf die Konflikt- und Wertgeladenheit von Schulinspektionen hinweisen. Dietrich und Lambrecht (2012) konnten gezielte Versuche des „*evidence-makings*“ seitens der Inspektoren in Bezug auf die erzeugten Ergebnisse nachzeichnen.

Daneben hat sich ein weiterer Forschungsstrang herausgebildet, der stärker die *technisch-methodische Seite der Inspektionsprozesse* untersucht. So fanden Perels und Zahn (2013) im Vergleich von 20- und 45-minütigen Unterrichtsbeobachtungen keine wesentlichen Bewertungsunterschiede. Weitere Forschungen konzentrieren sich auf die Relevanz von Merkmalen der Inspektoren sowie der eingesetzten Instrumente für die Beobachtungsergebnisse (vgl. Müller/Pietsch 2011). Pietsch und Tosana (2008) fanden zwar nur geringe Urteilsfehler, empfehlen aber, Forschungen zu Objektivität und Validität der durch Schulinspektionen generierten Datensätze zu forcieren. Dies sei sowohl für die Ableitung wirksamer Maßnahmen als auch für die weitergehende Nutzung der Daten zu Zwecken der Bildungsforschung relevant (vgl. Wurster/Gärtner 2013; Gärtner/Pant 2011).

Insgesamt hat sich seit Einführung der Schulinspektion in Deutschland ein recht breites Forschungsfeld herausgebildet, wobei sich die Forscher zum Teil auch aus den Schulministerien unterstellten Instituten rekrutieren. Lambrecht und Rürup (2012) ist tendenziell zuzustimmen, wenn sie eine Dominanz von wirksamkeitsbezogenen Forschungen identifizieren. Zwar wird das Instrument der Schulinspektion zumeist im Zusammenhang mit Fragen zur Qualitätsentwicklung (vgl. Dedering 2012) betrachtet, aber unterdessen bilden die bildungspolitisch formulierten und in Wirkungsmodellen auffindbaren Wirksamkeitsannahmen nicht mehr die alleinige Orientierung für Untersuchungen. Der Blick richtet sich nun verstärkt auch auf soziale Praktiken, Verarbeitungsprozesse sowie Veränderungen der Akteurskonstellationen, worin in den Jahren zuvor noch die größten Forschungsdesiderate ausgemacht wurden. Die Rechenschaftsflegungsfunktion wurde bislang kaum erforscht (vgl. Husfeldt

2011), was sich aber perspektivisch ändern könnte, wird doch das Verhältnis von Schulinspektion und Schulaufsicht zumindest vereinzelt thematisiert (vgl. Maritzen 2006; Heinrich 2012).

Das Instrument Vergleichsarbeiten im Spiegel der deutschen Bildungsforschung

Vergleichsarbeiten zielen nicht auf ein Monitoring auf Schulsystemebene ab, sondern beobachten die Ebenen der Einzelschule und der Einzelklasse. Das Hauptanliegen der Vergleichsarbeiten besteht darin, die Prozesse der Schul- und Unterrichtsentwicklung durch Informationen zu den Leistungsständen von Schulklassen zu unterstützen (vgl. KMK 2006; KMK 2010).

Mit Beginn der Einführung länderübergreifender Vergleichsarbeiten¹ etablierte sich zunächst eine Forschungslinie im Anschluss an die Implementation der Tests in den Bundesländern. Im Mittelpunkt dieser Forschungen stehen *Fragen der Akzeptanz, der Rezeption, der Bewertung von Handhabbarkeit, Verständlichkeit und Nützlichkeit der Instrumente und Rückmeldeformate* (vgl. Maier/Kuper 2012). Diese Rezeptionsforschung steht im Kontext von Evaluationen des Instruments, die sich zum Teil auf einzelne Länder konzentrieren und auf Befragungen fußen (vgl. Groß Ophoff/Hosenfeld/Koch 2007; Nachtigall/Hellrung 2013; Bonsen/Büchter/Peek 2006; Kühle/Peek 2007). Diese Studien ergaben folgende Befunde: Mit der Durchführung und Auswertung der Tests hatten die Lehrkräfte überwiegend keine oder nur wenige Probleme. Befragt zur Einschätzung der Nützlichkeit zeigte sich, dass die Rückmeldungen überwiegend als diagnostische Information über den erreichten Lernstand rezipiert werden, Reflexion und Entwicklung des eigenen Unterrichtes bleiben zumeist aus. Es gibt Hinweise auf einen Zusammenhang zwischen wahrgenommener Nützlichkeit und tatsächlicher Nutzung der Rückmeldungen.

Neben diesen Evaluationsforschungen finden sich Studien, die ebenfalls Fragen der Rezeption und Wirkung von Vergleichsarbeiten untersuchen, deren Fragestellungen aber ein stärkeres Erkenntnisinteresse an der Überprüfung der Steuerungseffekte auszeichnet. Quantitative Arbeiten zielen zumeist auf die Erforschung der Determinanten von Rezeption und Nutzung der Ergebnissrückmeldungen, wobei die allgemeine Bereitschaft zur Auseinandersetzung und Nutzung und die tatsächlich erfolgte Auseinandersetzung bzw. die daraus abgeleiteten Maßnahmen unterschieden werden können. Mehrere Studien kommen zu dem Ergebnis, dass die allgemeine Bereitschaft zur Rezeption überwiegend positiv (vgl. Maier/Rauin 2006), aber im

1 In diesem Artikel werden nachfolgend die Begriffe Vergleichsarbeiten bzw. VERA verwendet, auch wenn in einzelnen Bundesländern die Begriffe für die jahrgangsbezogenen Tests in Jahrgangsstufe 3 und 8 abweichen.

Zeitverlauf gesunken ist, wobei Schulleitungen generell positivere Einschätzungen zeigen (vgl. Nachtigall/Hellrung 2013). Wichtige Einflussfaktoren auf die Nutzung sind etwa die Verständlichkeit der Rückmeldungen, die erlebte Nützlichkeit und Akzeptanz von Schulleistungsstudien sowie die Verankerung externer Evaluationen im Schulprogramm (vgl. Koch 2011) und die Kooperation im Kollegium (vgl. Asbrand/Heller/Zeitler 2012). Zudem zeigt sich, dass eine nachhaltige Veränderung der Unterrichtspraxis von einer Orientierung an der kriterialen Bezugsnorm beeinflusst wird, während soziale Vergleiche zwar positiv auf die Auseinandersetzung mit den Rückmeldungen wirken, nicht aber die Unterrichtsentwicklung anregen (vgl. Groß Ophoff 2013).

Seit dem Jahr 2009 tauchen vermehrt Forschungen mit qualitativen Designs auf. Im Mittelpunkt des Erkenntnisinteresses stehen die *Prozesse der Rezeption und Nutzung* bzw. die *Wahrnehmungen, Einschätzungen und Begründungen von Lehrkräften oder Schulleitungen*, die mit Bezug auf unterschiedliche Aspekte der Vergleichsarbeiten interpretiert und zum Teil auch typisiert werden. Bezugspunkt ist häufig das Rezeptionsmodell von Helmke und Hosenfeld (2005). Ob mit den Vergleichsarbeiten verbundene Steuerungsziele, zuvorderst die Impulsgebung zur Entwicklung eines kompetenzorientierten Unterrichtes, erreicht werden können, wird in einigen Arbeiten zu Wirkungen unter Berücksichtigung verschiedener Merkmale auf Individual- und Schulebene untersucht: Korngiebel (2014) konnte keine Nutzung der Rückmeldungen für Schulentwicklungsprojekte feststellen. Jäger-Gerstetten (2011) zeigte, dass Änderungen der Unterrichtspraxis nicht aufgrund von Vergleichsarbeiten vorgenommen wurden. Dies deckt sich mit Ergebnissen einer Interviewstudie von Maier (2008), wonach jeder bzw. jede zweite Befragte die Rückmeldungen nicht als Beitrag zur kritischen Reflexion des Unterrichtes betrachtete oder zur Ableitung von Entwicklungsmaßnahmen nutzte. Ob die Vergleichsuntersuchungen an den Schulen überhaupt als Element outputbezogener Steuerung im Sinne der mit ihnen verbundenen Intentionen wahrgenommen werden und wie aus den Rückmeldungen abgeleitete Maßnahmen für die Entwicklung des Unterrichtes aus Steuerungssicht bewertet werden können, ist Gegenstand von Arbeiten mit steuerungstheoretischem Bezug (vgl. Diemer 2013; Diemer/Kuper 2011; Kuper/Muslic 2012). Die Ergebnisse weisen darauf hin, dass die Tests unterschiedliche Steuerungswirkungen haben: Die rückgemeldeten Ergebnisse werden nicht nur als Outputinformationen wahrgenommen, sondern vor allem als Prozessinformationen, was Auswirkungen auf die Art der verwirklichten Maßnahmen im Unterricht hat. Dies mag auch daran liegen, dass bei vielen Lehrkräften Unklarheit über die Anlage von Vergleichsarbeiten und die fachdidaktischen Grundlagen eines kompetenzorientierten Unterrichtes herrscht (vgl. Maier/Ramsteck/Frühwacht 2013). Im KMK-Beschluss zur Weiterentwicklung der Vergleichsarbeiten (vgl. KMK 2012) werden solche Probleme bereits thematisiert.

In anderen Arbeiten wird die *Rezeption und Nutzung von Vergleichsarbeiten vor der Folie weiterer theoretischen Rahmungen* eingeordnet und analysiert: so bezo-

gen auf Professions- (vgl. Kuper/Hartung 2007) und Organisationstheorie (vgl. Hartung-Beck 2009), Wissenstheorie (vgl. van Ackeren et. al. 2013) oder bezogen auf das Vergleichen als Erkenntnismittel (vgl. Kuper/Diemer 2012). Daneben finden sich Fallstudien zu Steuerungskonzepten (vgl. Dederling 2008) oder zur Rolle der Schulleitungen im Rezeptionsprozess (vgl. Muslic/Ramsteck/Kuper 2013).

Graf, Emmrich, Harych und Brunner (2013) fanden auf Basis einer experimentellen Studie zu *Durchführungseffekten* heraus, dass Schülerinnen und Schüler in den Tests etwas besser abschneiden, wenn diese von Lehrkräften anstelle von externen Testleitern und -leiterinnen durchgeführt werden.

Ob mit den Vergleichsarbeiten auch *Kompetenzsteigerungen von Schülerinnen und Schülern* einhergehen, ist selten Gegenstand empirischer Forschung. Richter, Böhme, Becker, Pant und Stanat (2014) untersuchen den Zusammenhang zwischen der Nutzung von VERA zur kompetenzorientierten Unterrichtsentwicklung und dem Abschneiden der Schülerinnen und Schüler in den Tests und finden einen positiven Zusammenhang.

Wenige Arbeiten sind kritisch-reflektierender Art und setzen sich auf einer Meta-Ebene mit den *Funktionen der Vergleichsarbeiten und ihren Zielen auf Ebene der Konzepte* selbst auseinander, so z.B. Maier, Metz, Bohl, Kleinknecht und Schymalla (2012), die die Funktionsüberfrachtung von VERA oder das häufig genutzte Rezeptionsmodell von Helmke und Hosenfeld (2005) kritisch diskutieren, oder Bartinitzky (2006), dessen Kritik bildungstheoretisch ansetzt.

Insgesamt lässt sich resümieren, dass es sich bei Forschungen zu den länderübergreifenden Vergleichsarbeiten in ihrer weit überwiegenden Zahl der Anlage nach um Rezeptionsforschung handelt, welche aufgrund ihrer Orientierung an politischen Steuerungszielen als Wirksamkeitsforschung charakterisiert werden kann. Die Ergebnisse dieser Arbeiten tragen dazu bei, auf die intendierten und nichtintendierten Wirkungen von Vergleichsarbeiten hinzuweisen und Vorschläge zur Verbesserung abzuleiten. Viele Arbeiten knüpfen direkt an das Rezeptionsmodell von Helmke und Hosenfeld an (vgl. Helmke/Hosenfeld 2005) und bestätigen dabei die Modellannahmen. In den Arbeiten benannte Desiderata beziehen sich einerseits auf die (methodische) Anlage weiterer Untersuchungen (bspw. längsschnittliche Untersuchungen, repräsentative Stichproben, *Good-Practice*-Forschung, Interventionen), auf die Testinstrumente selbst (mehr Unterstützungsangebote, bessere Informationen zum Zweck der Vergleichsarbeiten) oder auf zu berücksichtigende Variablen.

Zusammenfassung und Fazit

Bislang zeigen die Auseinandersetzungen mit Bildungsmonitoringelementen seitens der Bildungsforschung, dass insbesondere für Schulinspektion und Vergleichsarbeiten eine Vielzahl an Forschungsbeiträgen zu verzeichnen ist, wohingegen Bildungsberichterstattung bislang kaum Gegenstand empirischer Untersuchungen war. Auffällig ist eine gewisse Kongruenz bildungspolitischer Zweck- und Zielsetzungen bezüglich der Instrumente sowie der konzeptionellen Beiträge und verwendeten Analysemodelle der Wissenschaft. Damit erfüllt die Bildungsforschung ihrerseits eine Aufklärungsfunktion, da den postulierten Annahmen und Erwartungen mit wissenschaftlichen Mitteln generierte „belastbare“ Erkenntnisse gegenübergestellt sowie Validität und Reliabilität der Ergebnisse ermittelt werden können. Auch werden in den Ausblicken zuweilen lohnenswerte Fortsetzungen der Arbeiten sowohl für die Forschung als auch für die bildungspolitische und verwaltungsseitige Praxis benannt. Diese Beobachtungen verweisen auf eine spezifische Verkoppelung von Bildungspolitik und -administration und Bildungsforschung (vgl. auch Lambrecht/Rürup 2012). Dafür spricht auch, dass genuin wissenschaftsimmanent hervorgebrachte Thematisierungen der Monitoringelemente, abseits von vorformulierten Wirksamkeitserwartungen, die Arbeiten eher selten fundieren. Ebenso sind grundsätzliche Kritiken der postulierten Annahmen und Mechanismen nur vereinzelt vorfindbar.

Des Weiteren ist auffällig, dass die faktischen Rollenverständnisse und Handlungsweisen von Politik und Administration in Bildungsmonitoringzusammenhängen in der Regel nicht Gegenstand von Thematisierungen der Bildungsforschung sind (vgl. z.B. Böttcher/Kotthoff 2007). Der Blick richtet sich, entsprechend der als Schaltstellen für Veränderungen identifizierten Ebenen (Schule, Unterricht), vornehmlich auf die Akteure pädagogischer Leitungsstrukturen. Gerade aber bei der Überprüfung von Rechenschaftslegungsfunktionen oder der Reaktionsweisen auf neue Wissensbestände, die auf Ergebnissen aus Monitoringverfahren gründen, nehmen die politisch-administrativen Akteure eine zentrale Rolle ein. Dadurch, dass die Studien aufgrund der Spezifität ihrer Fragestellungen häufig lediglich Ausschnitte des jeweiligen Verfahrens und der Akteurskonstellationen beleuchten, geraten Bedingungsgefüge, die der Gesamtarchitektur der Monitoringsysteme inhärent sind, kaum in den Blick. Diese Desiderata werden offenbar zumindest von Teilen der Bildungsforschung erkannt und womöglich zukünftig bearbeitet.

Literatur und Internetquellen

Abendroth, S./Sendzik, N./Järvinen, H. (2014): Nutzen Schulleitungen kommunale Bildungsberichte? Vortrag auf der KBBB-Tagung „Institutioneller Wandel im Bildungsbereich – Reform ohne Kritik?“. Jena. Unveröff. Manuskript.

- Ackeren, I. van/Binneweis, C./Clausen, M./Demski, D./Dormann, C./Koch, A.R./Laier, B./Preisendörfer, P./Preuß, D./Rosenbusch, C./Schmidt, U./Stump, M./Zlatkin-Troitschanskaia, O. (2013): Welche Wissensbestände nutzen Schulen im Kontext von Schulentwicklung? In: Ackeren, I. van/Heinrich, M./Thiel, F. (Hrsg.): Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBiS-Verbund. Die Deutsche Schule, 12. Beiheft, S. 51-73.
- Ackeren, I. van/Block, R. (2009): Schulsysteme in der Umstrukturierung. In: Sacher, W./Blömeke, S./Bohl, T./Haag, L./Lang-Wojtasik, H. (Hrsg.): Handbuch Schule. Bad Heilbrunn: Klinkhardt/UTB, S. 207-216.
- Altrichter, H./Ehren, M./McNamara, G./O'Hara, J. (2014): Wie will Schulinspektion wirken? Analyse von Annahmen über Wirkungsmechanismen von Schulinspektion in sechs europäischen Ländern. In: Pfeifer, M. (Hrsg.): Schulqualität und Schulentwicklung. Theorien, Analysen und Potenziale. Münster u.a.: Waxmann, S. 184-207.
- Altrichter, H./Maag Merki, K. (Hrsg.) (2010): Handbuch Neue Steuerung im Schulsystem. Wiesbaden: VS.
- Asbrand, B./Heller, N./Zeitler, S. (2012): Die Arbeit mit Bildungsstandards in Fachkonferenzen. In: Die Deutsche Schule 104, H. 1, S. 31-43.
- Barlösius, E. (2005): Die Macht der Repräsentation. Common Sense über soziale Ungleichheiten. Wiesbaden: VS.
- Bartinitzky, H. (2006): Wie VERA und Verwandtes die Bildungsqualität beschädigen. In: Die Deutsche Schule 98, H. 2, S. 201-213.
- Berkemeyer, N./Manitius, V. (2013): Gerechtigkeit als Kategorie der Analyse von Schulsystemen – das Beispiel Chancenspiegel. In: Schwippert, K./Bonsen, M./Berkemeyer, N. (Hrsg.): Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven. Münster u.a.: Waxmann, S. 223-240.
- Böhm-Kasper, O./Selders, O. (2013): „Schulinspektionen sollten regelmäßig durchgeführt werden“? Ländervergleichende Analyse der Wahrnehmung und Akzeptanz von Schulinspektionsverfahren. In: Ackeren, I. van/Heinrich, M./Thiel, F. (Hrsg.): Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBiS-Verbund. Die Deutsche Schule, 12. Beiheft, S. 121-153.
- Böttcher, W. (2013): Das Monitoring-Paradigma. Eine Kritik der deutschen Schulreform. In: Empirische Pädagogik 27, H. 4, S. 496-509.
- Böttcher, W./Keune, M. (2012): Externe Evaluation und die Steuerung der Einzelschule: Kontrolle oder Entwicklung? In: Ratermann, M./Stöbe-Blossey, S. (Hrsg.): Governance von Schul- und Elementarbildung. Vergleichende Betrachtungen und Ansätze der Vernetzung. Wiesbaden: VS, S. 63-80.
- Böttcher, W./Kotthoff, H.-G. (2007): Gelingensbedingungen einer qualitätsoptimierenden Schulinspektion. In: Böttcher, W./Kotthoff, H.-G. (Hrsg.): Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung. Münster u.a.: Waxmann, S. 223-229.
- Böttcher, W./Kotthoff, H.-G. (2010): Neue Formen der ‚Schulinspektion‘: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In: Altrichter, H./Maag Merki, K. (Hrsg.): Handbuch Neue Steuerung im Schulsystem. Wiesbaden: VS, S. 295-325.
- Bonsen, M./Büchter, A./Peek, R. (2006): Datengestützte Schul- und Unterrichtsentwicklung – Bewertungen der Lernstandserhebungen in NRW durch Lehrerinnen und Lehrer. In: Rolff, H.-G. (Hrsg.): Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven. Weinheim: Juventa, S. 125-148.
- Bos, W./Dederig, K./Holtappels, H.G./Müller, S./Rösner, E. (2007): Schulinspektionen in Deutschland – Eine kritische Bestandsaufnahme. In: Buer, J. van/Wagner, C. (Hrsg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt a.M.: Peter Lang, S. 241-257.

- Brüsemeister, T./Eubel, K.-D. (Hrsg.) (2003): Zur Modernisierung der Schule. Leitideen – Konzepte – Akteure. Ein Überblick. Bielefeld: transcript.
- Brüsemeister, T./Preuß, B./Wissinger, J. (2014): Schulentwicklung als Governance – Herausforderungen datenbasierter Schulentwicklung. In: Fickermann, D./Maritzen, N. (Hrsg.): Grundlagen für eine daten- und theoriegestützte Schulentwicklung. Konzeption und Anspruch des Hamburger Instituts für Bildungsmonitoring und Qualitätsentwicklung (IfBQ). Münster u.a.: Waxmann, S. 215-233.
- Dedering, K. (2008): Der Einfluss bildungspolitischer Maßnahmen auf die Steuerung des Schulsystems. Neue Erkenntnisse aus empirischen Fallstudien. In: Zeitschrift für Pädagogik 54, H. 6, S. 869-887.
- Dedering, K. (2012): Schulinspektion als wirksamer Weg der Systemsteuerung? In: Zeitschrift für Pädagogik 58, H. 1, S. 69-88.
- Dedering, K./Fritsch, N./Weyer, C. (2012): Die Ankündigung von Schulinspektionen und deren innerschulische Effekte – hektisches Treiben oder genügsame Gelassenheit? In: Hornberg, S./Parreira do Amaral, M. (Hrsg.): Deregulierung im Bildungswesen. Münster u.a.: Waxmann, S. 205-222.
- Diemer, T. (2013): Innerschulische Wirklichkeiten neuer Steuerung: Zur Nutzung zentraler Lernstandserhebungen. Wiesbaden: Springer.
- Diemer, T./Kuper, H. (2011): Formen innerschulischer Steuerung mittels zentraler Lernstandserhebungen. In: Zeitschrift für Pädagogik 57, H. 4, S. 554-571.
- Dietrich, F./Lambrecht, M. (2012): Menschen arbeiten mit Menschen. Schulinspektion und die Hoffnung auf den zwanglosen Zwang der „besseren Evidenz“. In: Die Deutsche Schule 104, H. 1, S. 58-71.
- Döbert, H. (2008): Die Bildungsberichterstattung in Deutschland. Oder: Wie können Indikatoren zu Innovationen im Bildungswesen beitragen? In: LISUM/bm:ukk/EDK (Hrsg.): Bildungsmonitoring, Vergleichsstudien und Innovationen. Von evidenzbasierter Steuerung zur Praxis. Berlin: WBV, S. 71-93.
- Döbert, H./Baethge, M./Hetmeier, H.-W./Seeber, S./Füssel, H.-P./Klieme, E./Rauschenbach, T./Rockmann, U./Wolter, A. (2009): Das Indikatorenkonzept der nationalen Bildungsberichterstattung in Deutschland. In: Tippelt, R. (Hrsg.): Steuerung durch Indikatoren. Methodologische und theoretische Reflektionen zur deutschen und internationalen Bildungsberichterstattung. Opladen: Barbara Budrich, S. 207-272.
- Döbert, H./Klieme, E. (2010): Indikatoren gestützte Bildungsberichterstattung. In: Tippelt, R./Schmidt, B. (Hrsg.): Handbuch Bildungsforschung. 3., durchgesehene Aufl. Wiesbaden: VS, S. 317-336.
- Gärtner, H. (2015): Von der Produktion regionaler Bildungsberichte und den Erwartungen an ihre Rezeption und Nutzung. Vortrag im Workshop „Kommunale Bildungsberichte zwischen Beobachtung und Steuerung, Gleichheit und Leistung“. Berlin. Unveröff. Manuskript.
- Gärtner, H./Hüseemann, D./Pant, H.A. (2009): Wirkungen von Schulinspektion aus Sicht betroffener Schulleitungen. Die Brandenburger Schulleiterbefragung. In: Empirische Pädagogik 23, H. 1, S. 1-18.
- Gärtner, H./Pant, H.A. (2011): Validierungsstrategien für Verfahren und Ergebnisse von Schulinspektion. In: Müller, S./Pietsch, M./Bos, W. (Hrsg.): Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht. Münster u.a.: Waxmann, S. 9-32.
- Graf, T./Emmrich, R./Harych, P./Brunner, M. (2013): Durchführungseffekte bei Vergleichsarbeiten in Jahrgangsstufe 8. In: Empirische Pädagogik 27, H. 4, S. 459-473.
- Groß Ophoff, J. (2013): Der Effekt der Bezugsnormorientierung auf die Reflexion und Nutzung von Rückmeldungen aus Vergleichsarbeiten. In: Empirische Pädagogik 27, H. 4, S. 442-458.

- Groß Ophoff, J./Hosenfeld, I./Koch, U. (2007): Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. In: *Empirische Pädagogik* 21, H. 4, S. 411-427.
- Hartung-Beck, V. (2009): *Schulische Organisationsentwicklung und Professionalisierung. Folgen von Lernstandserhebungen an Gesamtschulen*. Wiesbaden: VS.
- Heinrich, M. (2012): Zum Verhältnis von Schulinspektion und Schulaufsicht. Vortrag auf der DGBV-Tagung „Professionalisierung von Schulinspektion und Schulaufsicht“. Bad Kreuznach. Unveröff. Manuskript.
- Helmke, A./Hosenfeld, I. (2005): Standardbezogene Unterrichtsevaluation. In: Brägger, G./Bucher, B./Landwehr, N. (Hrsg.): *Schlüsselfragen zur externen Schulevaluation*. Bern: hep, S. 127-151.
- Hermstein, B./Manitius, V. (2015, im Erscheinen): Bildungsberichterstattung als diskursive Ordnung. Begründung der Annahme von sozialer Gerechtigkeit als Leitkategorie. In: Manitius, V./Hermstein, B./Berkemeyer, N./Bos, W. (Hrsg.): *Zur Gerechtigkeit von Schule*. Münster u.a.: Waxmann.
- Hüfner, A. (2006): Bildungsberichterstattung – Erwartungen aus Sicht der Politik. In: *Zeitschrift für Erziehungswissenschaft* 9, H. 6, S. 15-19.
- Husfeldt, V. (2011): Wirkungen und Wirksamkeit der externen Schulevaluation. In: *Zeitschrift für Erziehungswissenschaft* 14, H. 2, S. 259-282.
- Jäger-Gerstetten, S. (2011): Rezeption und Nutzung von Diagnose- und Vergleichsarbeiten an Schulen. Eine Interviewstudie mit baden-württembergischen Lehrkräften an Haupt-, Realschulen und Gymnasien. URL: <http://d-nb.info/1051226406/34>; Zugriffsdatum: 15.05.2015.
- Katenbrink, N./Schaffer, G. (2015, im Erscheinen): „Schlicht und ergreifend bleibt es eine Prüfung der Schule“. Rekonstruktion der „Inspektionsgeschichte“ einer Schule mit gravierenden Mängeln. In: Hermstein, B./Manitius, V./Berkemeyer, N. (Hrsg.): *Institutioneller Wandel im Bildungsbereich. Reform ohne Kritik?* Weinheim/Basel: Beltz Juventa.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2006): *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. URL: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_08_01-Gesamtstrategie-Bildungsmonitoring.pdf; Zugriffsdatum: 15.05.2015.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2010): *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. URL: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_00_00-Konzeption-Bildungsstandards.pdf; Zugriffsdatum: 15.05.2015.
- KMK (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2012): *Vereinbarung zur Weiterentwicklung von VERA. Beschluss der Kultusministerkonferenz vom 08.03.2012*. URL: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_03_08_Weiterentwicklung-VERA.pdf; Zugriffsdatum: 15.05.2015.
- Koch, U. (2011): Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten? Datenkompetenz von Lehrkräften und die Nutzung von Ergebnisrückmeldungen aus Vergleichsarbeiten. Münster u.a.: Waxmann.
- Konsortium Bildungsberichterstattung (2005): *Gesamtkonzeption der Bildungsberichterstattung*. URL: <http://www.bildungsbericht.de/daten/gesamtkonzeption.pdf>; Zugriffsdatum: 15.05.2015.
- Korngiebel, J. (2014): *Vergleichsarbeiten und ihr Potential für die Schul- und Unterrichtsentwicklung. Eine qualitative Untersuchung zur Nutzung der Lernstandserhebungen*

- an hessischen Gymnasien. URL: <http://archiv.ub.uni-marburg.de/diss/z2014/0221/pdf/djk.pdf>; Zugriffsdatum: 15.05.2015.
- Kühle, B./Peek, R. (2007): Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. In: *Empirische Pädagogik* 21, H. 4, S. 428-447.
- Kuper, H./Diemer, T. (2012): Vergleichsarbeiten: Theoretische und empirische Betrachtungen zum Nutzen des Vergleichens. In: Wacker, A./Maier, U./Wissinger, J. (Hrsg.): *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikationen*. Wiesbaden: Springer VS, S. 225-245.
- Kuper, H./Hartung, V. (2007): Überzeugungen zur Verwendung des Wissens aus Lernstandserhebungen. In: *Zeitschrift für Erziehungswissenschaft* 10, H. 2, S. 214-229.
- Kuper, H./Muslic, B. (2012): Evidenzbasierte Steuerung. Der Umgang mit Rückmeldungen aus Lernstandserhebungen als Instrument der evidenzbasierten Steuerung. In: Hornberg, S./Parreira do Amaral, M. (Hrsg.): *Deregulierung im Bildungswesen*. Münster u.a.: Waxmann, S. 151-164.
- Lambrecht, M./Rürup, M. (2012): Bildungsforschung im Rahmen einer evidence based policy. Das Beispiel „Schulinspektion“. In: Wacker, A./Maier, U./Wissinger, J. (Hrsg.): *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikationen*. Wiesbaden: Springer VS, S. 57-77.
- Landwehr, N. (2011): Thesen zur Wirkung und Wirksamkeit der externen Schulevaluation. In: Quesel, C./Husfeldt, V./Landwehr, N./Steiner, P. (Hrsg.): *Wirkungen und Wirksamkeit der externen Schulevaluation*. Bern: hep, S. 35-71.
- Lüders, C. (2006): Was leistet wissenschaftliche Sozialberichterstattung für Fachpraxis und Politik? In: *Zeitschrift für Erziehungswissenschaft*, Beiheft 6, S. 27-41.
- Maier, U. (2008): Was lernen Schulen aus zentralen Tests? Was sollten Bildungspolitiker lernen, wenn sie testen lassen? In: *Die Deutsche Schule* 100, H. 1, S. 66-72.
- Maier, U./Kuper, H. (2012): Vergleichsarbeiten als Instrumente der Qualitätsentwicklung an Schulen. Überblick zum Forschungsstand. In: *Die Deutsche Schule* 104, H. 1, S. 88-99.
- Maier, U./Metz, K./Bohl, T./Kleinknecht, M./Schymalla, M. (2012): Vergleichsarbeiten als Instrument der datenbasierten Schul- und Unterrichtsentwicklung in Gymnasien. In: Wacker, A./Maier, U./Wissinger, J. (Hrsg.): *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikationen*. Wiesbaden: Springer VS, S. 197-224.
- Maier, U./Ramsteck, C./Frühwacht, A. (2013): Lehr- und lerntheoretische Argumentationsmuster bei der Interpretation und Nutzung von Vergleichsarbeitsrückmeldungen durch Gymnasiallehrkräfte. In: Ackeren, I. van/Heinrich, M./Thiel, F. (Hrsg.): *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBiS-Verbund*. *Die Deutsche Schule*, 12. Beiheft, S. 74-96.
- Maier, U./Rauin, U. (2006): Vergleichsarbeiten. Hilfe zur Unterrichtsentwicklung? Zentrale Lernstandserhebungen aus Sicht baden-württembergischer Lehrkräfte. In: *Die Deutsche Schule* 98, H. 4, S. 403-421.
- Maritzen, N. (2006): Schulinspektionen zwischen Aufsicht und Draufsicht – Eine Trendanalyse. In: Buchen, H./Horster, L./Rolff, H.-G. (Hrsg.): *Schulinspektion und Schulleitung*. Stuttgart: Raabe Verlag für Bildungsmanagement, S. 7-26.
- Müller, S./Pietsch, M. (2011): Was wir messen, wenn wir Unterrichtsqualität messen. Inter-Beurteilerübereinstimmungen und -Reliabilität bei Unterrichtsbeobachtungen im Rahmen von Schulqualität. In: Müller, S./Pietsch, M./Bos, W. (Hrsg.): *Schulinspektionen in Deutschland. Eine Zwischenbilanz aus empirischer Sicht*. Münster u.a.: Waxmann, S. 33-56.
- Muslic, B./Ramsteck, C./Kuper, H. (2013): Das Verhältnis von Schulleitung und Schulaufsicht im Kontext testbasierter Schulreform. Kontrastive Fallstudien zur

- Rezeption von Lernstandsergebnissen im Mehrebenensystem der Schule. In: Ackeren, I. van/Heinrich, M./Thiel, F. (Hrsg.): Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBiS-Verbund. Die Deutsche Schule, 12. Beiheft, S. 97-120.
- Nachtigall, C./Hellrung, K. (2013): Zur zeitlichen Entwicklung der Rezeption von Vergleichsarbeiten. In: Empirische Pädagogik 27, H. 4, S. 423-441.
- Niedlich, S./Brüsemeister, T. (2012): Bildungsmonitoring zwischen Berichterstattung und Steuerungsanspruch – Entwicklungslinien und akteurtheoretische Implikationen. In: Wacker, A./Maier, U./Wissinger, J. (Hrsg.): Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsseitige Implikationen. Wiesbaden: VS, S. 131-152.
- Perels, F./Zahn, A. (2013): Auswirkungen der Variation der Dauer von Unterrichtsbeobachtungen im Rahmen der Hessischen Schulinspektion – Vergleich von 20- und 45-minütigen Unterrichtsbeobachtungen. In: Unterrichtswissenschaft 41, S. 235-251.
- Pietsch, M./Janke, N./Mohr, I. (2014): Führt Schulinspektion zu besseren Schülerleistungen? Difference-in-Difference-Studien zu Effekten der Schulinspektion Hamburg auf Lernzuwächse und Leistungstrends. In: Zeitschrift für Pädagogik 60, H. 3, S. 446-470.
- Pietsch, M./Tosana, S. (2008): Beurteilereffekte bei der Messung von Unterrichtsqualität: Das Multifacetten-Rasch-Modell und die Generalisierbarkeitstheorie als Methoden der Qualitätssicherung in der externen Evaluation von Schulen. In: Zeitschrift für Erziehungswissenschaft 11, H. 3, S. 430-452.
- Richter, D./Böhme, K./Becker, M./Pant, H.A./Stanat, P. (2014): Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten. Zusammenhänge zu Veränderungen im Unterricht und den Kompetenzen von Schülerinnen und Schülern. In: Zeitschrift für Pädagogik 60, H. 2, S. 225-244.
- Rürup, M. (2008): Typen der Schulinspektion in den deutschen Bundesländern. In: Die Deutsche Schule 100, H. 4, S. 467-477.
- Rürup, M./Fuchs, H.-W./Weishaupt, H. (2010): Bildungsberichterstattung – Bildungsmonitoring. In: Altrichter H./Maag Merki, K. (Hrsg.): Handbuch Neue Steuerung. Wiesbaden: VS, S. 377-401.
- Scheerens, J./Hendriks, M. (2004): Benchmarking the Quality of Education. In: European Educational Research Journal 3, H. 1, S. 101-114.
- Schwank, E./Sommer, N. (2012): Wirkungen der Schulinspektion anhand der Wahrnehmung der Lehrkräfte. Ergebnisse aus einer Befragung im Rahmen der Inspektions-evaluation. In: Schulverwaltung Niedersachsen 23, H. 4, S. 106-110.
- Siepeke, T./Tegge, D./Egger, M. (2014): Kommunale Bildungsberichterstattung – Standards und Varianzen. In: Drossel, K./Strietholt, R./Bos, W. (Hrsg.): Empirische Bildungsforschung und evidenzbasierte Reformen im Bildungswesen. Münster u.a.: Waxmann, S. 155-177.
- Sowada, M.G./Dedering, K. (2014): Ermessensspielräume in der Bewertungsarbeit von Schulinspektor/innen. In: Zeitschrift für Bildungsforschung 4, H. 2, S. 119-135.
- Wacker, A./Maier, U./Wissinger, J. (Hrsg.) (2012): Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung – Empirische Befunde und forschungsmethodische Implikationen. Wiesbaden: VS.
- Weishaupt, H. (2006): Der Beitrag von Wissenschaft und Forschung zur Bildungs- und Sozialberichterstattung. In: Zeitschrift für Erziehungswissenschaft, Beiheft 6, S. 42-52.
- Weishaupt, H. (2009): Indikatoren für die regionale Bildungsberichterstattung. In: Tippelt, R. (Hrsg.): Steuerung durch Indikatoren. Methodologische und theoretische Reflektionen zur deutschen und internationalen Bildungsberichterstattung. Opladen: Barbara Budrich, S. 189-205.
- Weishaupt, H./Zimmer, K. (2013): Indikatoren kultureller Bildung. In: Zeitschrift für Erziehungswissenschaft 16, H. 3, S. 83-98.

Wurster, S./Gärtner, H. (2013): Erfassung von Bildungsprozessen im Rahmen von Schulinspektion und deren potenzieller Nutzen für die empirische Bildungsforschung. In: Unterrichtswissenschaft 41, H. 3, S. 216-234.

Björn Hermstein, M.A., geb. 1984, wissenschaftlicher Mitarbeiter am Institut für Erziehungswissenschaft der Friedrich-Schiller-Universität Jena und am Institut für Schulentwicklungsforschung (IFS) der TU Dortmund.

E-Mail: bjoern.hermstein@uni-jena.de; hermstein@ifs.tu-dortmund.de

Ina Semper, M.A., geb. 1979, wissenschaftliche Mitarbeiterin am Institut für Erziehungswissenschaft der Friedrich-Schiller-Universität Jena.

E-Mail: ina.semper@uni-jena.de

Nils Berkemeyer, Prof. Dr., geb. 1975, Inhaber des Lehrstuhls für Schulpädagogik und Schulentwicklung am Institut für Erziehungswissenschaft der Friedrich-Schiller-Universität Jena.

E-Mail: nils.berkemeyer@uni-jena.de

Lisa Mende, 1. Staatsexamen für das Lehramt Gymnasium, geb. 1990, wissenschaftliche Mitarbeiterin am Institut für Erziehungswissenschaft der Friedrich-Schiller-Universität Jena.

E-Mail: lisa.mende@uni-jena.de

Anschrift: Friedrich-Schiller-Universität Jena, Institut für Erziehungswissenschaft, Am Planetarium 4, 07743 Jena